Max-Planck-Institut
für Mathematik
in den Naturwissenschaften
Leipzig

The homological nature of entropy

(revised version: May 2015)

by

Pierre Baudot and Daniel Bennequin

THE HOMOLOGICAL NATURE OF ENTROPY

PIERRE BAUDOT AND DANIEL BENNEQUIN

ABSTRACT. We propose that entropy is a universal co-homological class in a theory associated to a family of observable quantities and a family of probability distributions. Three cases are presented:
(1) classical probabilities and random variables;
(2) quantum probabilities and observable operators;
(3) dynamic probabilities and observation trees.
This gives rise to a new kind of topology for information processes, that accounts for the main information functions: entropy, mutual-informations at all orders, and Kullback-Leibler divergence and generalizes them in several ways. The article is divided into two parts, that can be read independently. In the first part, the introduction, we provide an overview of the results, some open questions, future results and lines of research, and discuss briefly the application to complex data. In the second part we give the complete definitions and proofs of the theorems A, C and E in the introduction, which show why entropy is the first homological invariant of a structure of information in four contexts: static classical or quantum probability, dynamics of classical or quantum strategies of observation of a finite system.

CONTENTS

1

1. Introduction

1.1. **What is Information ?**

"What is information ?" is a question that has received several answers according to the different problems investigated. The best known definition was given by Shannon [49], using random variables and a probability law, for the problem of optimal message compression. However, the first definition was given by Fisher, as a metric associated to a smooth family of probability distributions, for optimal discrimination by statistical tests; it is a limit of the Kullback-Leibler divergence, which was introduced to estimate the accuracy of a statistical model of empirical data, and which can be also viewed as a quantity of information. More generally Kolmogorov considered that the concept of information must precede probability theory (*cf.* [33]). However, Evariste Galois saw the application of group theory for discriminating solutions of an algebraic equation as a first step toward a general theory of ambiguity, that was developed further by Riemann, Picard, Vessiot, Lie, Poincare and Cartan, for systems of differential equations; it is also a theory of information. In another direction Rene Thom claimed that information must have a topological content (see [51]); he gave the example of the unfolding of the coupling of two dynamical systems, but he had in mind the whole domain of algebraic or differential topology.

All these approaches have in common the definition of secondary objects, either functions, groups or homology cycles, for measuring in what sense a pair of objects departs from independency. For instance, in the case of Shannon, the mutual information is $I(X;Y) = H(X) + H(Y) - H(X,Y)$, where H denotes the usual Gibbs entropy $(H(X) = -\sum_x P(X = x) \ln_2 P(X = x))$, and for Galois it is the quotient set $IGal(L_1; L_2|K) = (Gal(L_1|K) \times Gal(L_2|K))/Gal(L|K)$, where L_1, L_2 are two fields containing a field K in an algebraic closure Ω of K, where L is the field generated by L_1 and L_2 in Ω, and where $Gal(L_i|K)$ (for $i = \emptyset, 1, 2$) denotes the group introduced by Galois, made by the field automorphisms of L_i fixing the elements of K.

We suggest that all information quantities are of co-homological nature, in a setting which depends on a pair of categories (*cf.* [39, 38]); one for the data on a system, like random variables or functions of solutions of an equation, and one for the parameters of this system, like probability laws or coefficients of equations; the first category generates an algebraic structure like a monoid, or more generally a monad (*cf.* [39]), and the second category generates a representation of this structure, as do for instance conditioning, or adding new numbers; then information

quantities are co-cycles associated with this module.

We will see that, given a set of random variables on a finite set Ω and a simplicial subset of probabilities on Ω, the *entropy* appears as the only one universal co-homology class of degree one. The higher mutual information functions that were defined by Shannon are co-cycles (or twisted co-cycles for even orders), and they correspond to higher homotopical constructions. In fact this description is equivalent to the theorem of Hu Kuo Ting [27], that gave a set theoretical interpretation of the mutual information decomposition of the total entropy of a system. Then we can use information co-cycles to describe forms of the information distribution between a set of random data; figures like ordinary links, or chains or Borromean links appear in this context, giving rise to a new kind of topology.

1.2. **Information Homology.**

Here we call random variables (r.v) on a finite set Ω *congruent* when they define the same partition (remind that a partition of Ω is a family of disjoint non-empty subsets covering Ω and that the partition associated to a r.v X is the family of subsets Ω_x of Ω defined by the equations $X(\omega) = x$; the join r.v YZ, also denoted by (Y, Z), corresponds to the less fine partition that is finer than Y and Z. This defines a monoid structure on the set $\Pi(\Omega)$ of partitions of Ω, with 1 as a unit, and where each element is idempotent, *i.e.*, $\forall X, XX = X$. An *information category* is a set $\mathcal{S}$ of r.v such that, for any $Y, Z \in \mathcal{S}$ less fine than $U \in \mathcal{S}$, the join YZ belongs to $\mathcal{S}$, *cf.* [5]. An ordering on $\mathcal{S}$ is given by $Y \leq Z$ when Z refines Y, which also defines the morphisms $Z \to Y$ in the category $\mathcal{S}$. In what follows we always assume that 1 belongs to $\mathcal{S}$. The simplex $\Delta(\Omega)$ is defined as the set of families of numbers $\{p_\omega; \omega \in \Omega\}$, such that $\forall \omega, 0 \leq p_\omega \leq 1$ and $\sum_\omega p_\omega = 1$; it parameterizes all probability laws on Ω. We choose a simplicial sub-complex $\mathcal{P}$ in $\Delta(\Omega)$, which is stable by all the conditioning operations by elements of $\mathcal{S}$. By definition, for $N \in \mathbb{N}$, an information N-cochain is a family of measurable functions of $P \in \mathcal{P}$, with values in $\mathbb{R}$ or $\mathbb{C}$, indexed by the sequences $(S_1; ...; S_N)$ in $\mathcal{S}$ majored by an element of $\mathcal{S}$, whose values depend only of the image law $(S_1, ..., S_N)_* P$. This condition is natural from a topos point of view, *cf.* [39]; we interpret it as a "locality" condition. Note that we write $(S_1; ...; S_N)$ for a sequence, because $(S_1, ..., S_N)$ designates the joint variable. For $N = 0$ this gives only the constants. We denote by $\mathcal{C}^N$ the vector space of N-cochains of information. The following formula corresponds to the *averaged conditioning* of Shannon [49]:

$$(1) \qquad S_0.F(S_1; ...; S_N; \mathbb{P}) = \sum \mathbb{P}(S_0 = v_j) F(S_1; ...; S_N; \mathbb{P}|S_0 = v_j),$$

where the sum is taken over all values of S_0, and the vertical bar is ordinary conditioning. It satisfies the associativity condition $(S_0'S_0).F = S_0'.(S_0.F)$. The coboundary operator δ is defined by

$$(2) \qquad \delta F(S_0; ...; S_N; \mathbb{P})$$

$$= S_0.F(S_1; ...; S_N; \mathbb{P}) + \sum_0^{N-1} (-1)^{i+1} F(...; (S_i, S_{i+1}); ...; S_N; \mathbb{P}) + (-1)^{N+1} F(S_0; ...; S_{N-1}; \mathbb{P}),$$

It corresponds to a standard non-homogeneous bar complex (*cf.* [38]). Another co-boundary operator on $\mathcal{C}^N$ is δ_t (t for twisted or trivial action or topological complex), that is defined by the above formula with the first term $S_0.F(S_1; ...; S_N; \mathbb{P})$

replaced by $F(S_1; ...; S_N; \mathbb{P})$. The corresponding co-cycles are defined by the equations $\delta F = 0$ or $\delta_t F = 0$, respectively. We easily verify that $\delta \circ \delta = 0$ and $\delta_t \circ \delta_t = 0$; then co-homology $H^*(\mathcal{S}; \mathcal{P})$ resp. $H_t^*(\mathcal{S}; \mathcal{P})$ is defined by taking co-cycles modulo the elements of the image of δ resp. δ_t, called co-boundaries. The fact that classical entropy $H(X; \mathbb{P}) = -\sum_i p_i \log_2 p_i$ is a 1-co-cycle is the fundamental equation $H(X, Y) = H(X) + X.H(Y)$.

Theorem A. (*cf.* Theorem 1 section 2.3, [5]): For the full simplex $\Delta(\Omega)$, and if $\mathcal{S}$ is the monoid generated by a set of at least two variables, such that each pair takes at least four values, then the information co-homology space of degree one is one-dimensional and generated by the classical entropy.

Problem 1. Compute the homology of higher degrees.

We conjecture that for binary variables it is zero, but that in general non-trivial classes appear, deduced from polylogarithms. This could require us to connect with the works of Dupont, Bloch, Goncharov, Elbaz-Vincent, Gangl *et al.* on motives (*cf.* [19]), which started from the discovery of Cathelineau (1988) that entropy appears in the computation of the degree one homology of the discrete group SL_2 over $\mathbb{C}$ with coefficients in the adjoint action (*cf.* [14]).

Suppose $\mathcal{S}$ is the monoid generated by a finite family of partitions. The higher mutual informations were defined by Shannon as alternating sums:

$$(3) \qquad I_N(S_1; ...; S_N; \mathbb{P}) = \sum_{k=1}^{k=N} (-1)^{k-1} \sum_{I \subset [N]; card(I)=k} H(S_I; \mathbb{P}),$$

where S_I denotes the join of the S_i such that $i \in I$. We have $I_1 = H$ and $I_2 = I$ is the usual mutual information: $I(S; T) = H(S) + H(T) - H(S, T)$.

Theorem B. (*cf.* section 3, [5]): $\quad I_{2m} = \delta_t \delta \delta_t ... \delta \delta_t H, \quad I_{2m+1} = -\delta \delta_t \delta \delta_t ... \delta \delta_t H,$ where there are $m-1$ δ and m δ_t factors for I_{2m} and m δ and m δ_t factors for I_{2m+1}.

Thus odd information quantities are information co-cycles, because they are in the image of δ, and even information quantities are twisted (or topological) co-cycles, because they are in the image of δ_t.

In [5] we show that this description is equivalent to the theorem of Hu Kuo Ting (1962) [27], giving a set theoretical interpretation of the mutual information decomposition of the total entropy of a system: mutual information, join and averaged conditioning correspond respectively to intersection, union and difference $A \backslash B = A \cap B^c$. In special cases we can interpret I_N as homotopical algebraic invariants. For instance for $N = 3$, suppose that $I(X; Y) = I(Y; Z) = I(Z; X) = 0$, then $I_3(X; Y; Z) = -I((X, Y); Z)$ can be defined as a Milnor invariant for links, generalized by Massey, as they are presented in [37] (*cf.* page 284), through the 3-ary obstruction to associativity of products in a subcomplex of a differential algebra, *cf.* [5]. The absolute minima of I_3 correspond to Borromean links, interpreted as synergy, *cf.* [41, 12].

1.3. **Extension to Quantum Information.**

Positive hermitian $n \times n$-matrices ρ, normalized by $Tr(\rho) = 1$, are called density of states (or density operators) and are considered as quantum probabilities on $E = \mathbb{C}^n$. Real quantum observables are $n \times n$ hermitian matrices, and, by definition, the

amplitude, or expectation, of the observable Z in the state ρ is given by the formula $\mathbb{E}(Z) = Tr(Z\rho)$ (see e.g., [46]). Two real observables Y, Z are said *congruent* if their eigenspaces are the same, thus orthogonal decomposition of E are the quantum analogs of partitions. The join is well defined for commuting observables. An information structure $\mathbf{S}$ is given by a subset of observables, such that, if Y, Z have common refined eigenspaces decomposition in $\mathbf{S}$, their join (Y, Z) belongs to $\mathbf{S}$. We assume that $\{E\}$ belongs to $\mathbf{S}$. What plays the role of a probability functor is a map $\mathbf{Q}$ from $\mathbf{S}$ to sets of positive hermitian forms on E, which behaves naturally with respect to the quantum direct image, thus $\mathbf{Q}$ is a covariant functor. We define information N-cochains as for the classical case, starting with the numerical functions on the sets $\mathbf{Q}_X; X \in \mathbf{S}$, which behave naturally under direct images.

The restriction of a density ρ by an observable Y is $\rho_Y = \sum_A E_A^* \rho E_A$, where the E_A's are the spectral projectors of the observable Y. The functor $\mathbf{Q}$ is said to match $\mathbf{S}$ (or to be complete and minimal with respect to $\mathbf{S}$) if, for each $X \in \mathbf{S}$, the set $\mathbf{Q}_X$ is the set of all possible densities of the form ρ_X.

The action of a variable on the cochains space $\mathcal{C}_Q^*$ is given by the *quantum averaged conditioning*:

$$(4) \qquad Y.F(Y_0; ...; Y_m; \rho) = \sum_A Tr(E_A^* \rho E_A) F(Y_0; ...; Y_m; E_A^* \rho E_A)$$

From here we define coboundary operators δ_Q and δ_{Qt} by the formula (22), then the notions of co-cycles, co-boundaries and co-homology classes follow. We have $\delta_Q \circ \delta_Q = 0$ and $\delta_{Qt} \circ \delta_{Qt} = 0$; *cf.* [5].

When the unitary group U_n acts transitively on $\mathbf{S}$ and $\mathbf{Q}$, there is a notion of invariant cochains, forming a subcomplex of information cochains, and giving a more computable co-homology than the brut information co-homology. We call it the invariant information co-homology and denote it by $H_U^*(\mathbf{S}; \mathbf{Q})$.

The *Von-Neumann entropy* of ρ is $S(\rho) = \mathbb{E}_\rho(-\log_2(\rho)) = -Tr(\rho \log_2(\rho))$; it defines a 0-cochain S_Y by restricting S to the sets $\mathbf{Q}_X$. The classical entropy is $H(Y; \rho) = -\sum_A Tr(E_A^* \rho E_A) \log_2(Tr(E_A^* \rho E_A))$. Both these co-chains are invariant. It is well known that $S_{(X,Y)}(\rho) = H(X; \rho) + X.S_Y(\rho)$ when X, Y commute, *cf.* [46]. In particular, by taking $Y = 1_E$ we see that classical entropy measures the default of equivariance of the quantum entropy, *i.e.*, $H(X; \rho) = S_X(\rho) - (X.S)(\rho)$. But using the case where X refines Y, we obtain that the entropy of Shannon is the co-boundary of (minus) the Von Neumann entropy.

Theorem C. (*cf.* Theorem 3 section 4.3): For $n \geq 4$ and when $\mathbf{S}$ is generated by at least two decompositions such that each pair has at least four subspaces, and when $\mathbf{Q}$ is matching $\mathbf{S}$, the invariant co-homology H_U^1 of δ_Q in degree one is zero, and the space H_U^0 is of dimension one. In particular, the only invariant 0-cochain such that $\delta S = -H$ is the Von Neumann entropy.

(This statement, which will be proved below, corrects a similar statement which was made in the announcement [8].)

1.4. **Concavity and Convexity Properties of Information Quantities.**

The simplest classical information structure $\mathcal{S}$ is the monoid generated by a family of "elementary" *binary* variables $S_1, ..., S_n$. It is remarkable that in this case, the information functions $I_{N,J} = I_N(S_{j_1}; ...S_{j_N})$ over all the subsets $J = \{j_1, ..., j_N\}$ of $[n] = \{1, ..., n\}$, different from $[n]$ itself, give algebraically independent functions on

the probability simplex $\Delta(\Omega)$ of dimension $2^n - 1$. They form coordinates on the quotient of $\Delta(\Omega)$ by a finite group.

Let $\mathcal{L}_d$ denotes the Lie derivative with respect to $d = (1, ..., 1)$ in the vector space $\mathbb{R}^{2^n}$, and $\triangle$ the Euclidian Laplace operator on $\mathbb{R}^{2^n}$, then $\Delta = \triangle - 2^{-n} \mathcal{L}_d \circ \mathcal{L}_d$ is the Laplace operator on the simplex $\Delta(\Omega)$ defined by equating the sum of coordinates to 1.

Theorem D. (*cf.* [6]): On the affine simplex $\Delta(\Omega)$ the functions $I_{N,J}$ with N odd (resp. even) satisfies the inequality $\Delta I_N \geq 0$ (resp. $\Delta I_N \leq 0$).

In other terms, for N odd the $I_{N,J}$ are super-harmonic which is a kind of weak concavity and for N even they are sub-harmonic which is a kind of weak convexity. In particular, when N is even (resp. odd) $I_{N,J}$ has no local maximum (resp. minimum) in the interior of $\Delta(\Omega)$.

Problem 2. What can be said of the other critical points of $I_{N,J}$? What can be said of the restriction of one information function on the intersection of levels of other information functions? Information topology depends on the shape of these intersections and on the Morse theory for them.

1.5. **Monadic Cohomology of Information.**

Now we consider the category $\mathcal{S}^*$ of *generalized ordered partitions* of Ω over $\mathcal{S}$: they are sequences $S = (E_1, ..., E_m)$ of subsets of Ω such that $\cup_j E_j = \Omega$ and $E_i \cap E_j = \emptyset$ as soon as $i \neq j$. The number m is named the degree of S. Note the important technical point that some of the sets E_j can be the empty set. In the same spirit we introduce *generalized ordered orthogonal decompositions* of E for the quantum case; but in this summary, for simplicity we restrict ourselves to the classical case. Also we forget to add generalized to ordered up to now in this summary. A *rooted tree decorated* by $\mathcal{S}^*$ is an oriented finite planar tree Γ, with a marked initial vertex s_0, named the *root* of Γ, where each vertex s is equipped with an element F_s of $\mathcal{S}^*$, such that edges issued from s correspond to the values of F_s. When we want to mention that we restrict to partitions less fine than a partition X we put an index X, like in $\mathcal{S}_X^*$.

The notation $\mu(m; n_1, ..., n_m)$ denotes the operation which associates to an ordered partition S of degree m and to m ordered partitions S_i of respective degrees n_i, the ordered partition that is obtained by cutting the pieces of S using the pieces of S_i and respecting the order. An evident unit element for this operation is the unique partition π_0 of degree 1. The symbol μ_m denotes the collection of those operations for m fixed. The introduction of empty subsets in ordered partitions insures that the result of $\mu(m; n_1, ..., n_m)$ is a partition of length $n_1 + ... + n_m$, thus the μ_m do define what is named an *operad*; *cf.* [37, 20]. The axioms of unity, associativity and covariance for permutations are satisfied. See [42, 43, 20, 37] for the definition of operads.

The most important algebraic object which is associated to an operad is a *monad* (*cf.* [39, 20]), *i.e.*, a functor $\mathcal{V}$ from a category $\mathcal{A}$ to itself, equipped with two natural transformations $\mu : \mathcal{V} \circ \mathcal{V} \to \mathcal{V}$ and $\eta : \mathbb{R} \to \mathcal{V}$, which satisfy to the following axioms:

$$(5) \qquad \mu \circ (\mathcal{V}\mu) = \mu \circ (\mu\mathcal{V}), \quad \mu \circ (\mathcal{V}\eta) = Id = \mu \circ (\eta\mathcal{V})$$

In our situation, we can apply the Schur construction (*cf.* [20]) to the μ_m to get a monad: take for V the real vector space freely generated by $\mathcal{S}^*$; it is naturally graded, so it is the direct sum of spaces $V(m); m \geq 1$ where the symmetric group $\mathfrak{S}_m$ acts naturally to the right, then introduce, for any real vector space W the real vector space $\mathcal{V}(W) = \bigoplus_{m \geq 0} V(m) \otimes_{\mathfrak{S}_m} W^{\otimes m}$; the Schur composition is defined by $\mathcal{V} \circ \mathcal{V} = \bigoplus_{m \geq 0} V(m) \otimes_{\mathfrak{S}_m} \mathcal{V}^{\otimes m}$. It is easy to verify that the collection $(\mu_m; m \in \mathbb{N})$ defines a natural transformation $\mu : \mathcal{V} \circ \mathcal{V} \to \mathcal{V}$, and the trivial partition π_0 defines a natural transformation $\eta : \mathbb{R} \to \mathcal{V}$, that satisfied to the axioms of a monad.

Also we fix a functor of probability laws Q_X over the category $\mathcal{S}$. Let $\mathcal{M}_X(m)$ be the vector space freely generated over $\mathbb{R}$ by the symbols (P, i, m) where P belongs to Q_X, and $1 \leq i \leq m$. In the last section of the second part we show how this space arises from the consideration of divided probabilities. This is apparent on the following definition of the right action of the operad $\mathcal{V}$ on the family $\mathcal{M}_X(m); m \in \mathbb{N}^*$: a sequence $S_1, ..., S_m$ or ordered partitions in $\mathcal{S}_X^*$ acts to a generator (P, i, m) by giving the vector $\sum_j p_j (P_j, (i, j), n)$ where p_j is the probability $P(S_i = j)$ and P_j is the conditioned probability $P|(S_i = j)$. We denote by $\theta_m((P, i, m), (S_1, ..., S_m))$ this vector.

Now we consider the Schur functor $\mathcal{M}_X(W) = \bigoplus_m \mathcal{M}_X(m) \otimes_{\mathfrak{S}_m} W^{\otimes m}$; the operations θ_m define a natural transformation $\theta : \mathcal{M} \circ \mathcal{V} \to \mathcal{M}$, which is an action

to the right in the sense of monads, *i.e.*, $\theta \circ (\mathcal{F}\mu) = \theta \circ (\theta\mathcal{V})$; $\theta \circ (\mathcal{F}\eta) = Id$. (We forgot the index X for simplicity.)

Now we consider the bar resolution of $\mathcal{M}$: $.... \to \mathcal{M} \circ \mathcal{V}^{\circ(k+1)} \to \mathcal{M} \circ \mathcal{V}^{\circ k} \to ...$, as in Beck (triples, ...) [9], and Fresse [20], with its simplicial structure deduced from θ and μ, and the complex of natural transformations of $\mathcal{V}$-right modules $C^*(\mathcal{M}) = Hom_V(\mathcal{M} \circ \mathcal{V}^{\circ *}, \mathcal{R})$, where $\mathcal{R}$ is the trivial right module given by $\mathcal{R}(m) = \mathbb{R}$. As in the classical case, we restrict us to co-chains that are measurable in the probability (P, i, m).

The co-boundary is defined by the Hochschild formula, extended by MacLane and Beck to monads (see Beck [9]):

$$(6) \quad \delta F = F \circ (\theta\mathcal{V}^{\circ k}) - \sum_{i=0,...,k-1} (-1)^i F \circ \mathcal{M}\mathcal{V}^{\circ i}\mu\mathcal{V}^{\circ k-i-1} - (-1)^k F \circ \mathcal{M}\mathcal{V}^{\circ k}\epsilon.$$

The cochains are described by families of scalar measurable functions $F_X(S_1; ...; S_k; (P, i, m))$, where $S_1; ...; S_k$ is a forest of m trees of level k labelled by $\mathcal{S}_X^*$, and where the value on (P, i, m) depends only on the tree $S_1^i; S_2^i; ...; S_k^i$.

We impose now the condition, named *regularity*, that $F_X(S_1, ..., S_k; (P, i, m)) = F_X(S_1^i; S_2^i; ...; S_k^i; P)$. The regular co-chains form a sub-complex $C_r^*(\mathcal{M})$; by definition, its homology is the *arborescent information co-homology*.

The regular cochains of degree k are determined by their values for $m = 1$ and decorated trees of level k, where the co-boundary takes the form:

$$\delta F(S; S_1; ...; S_k; \mathbb{P})$$

$$= \left(\sum_i\right) \mathbb{P}(S = i) F(S_1^i; ...; S_k^i; \mathbb{P}|(S = i)) + \sum_{i=1}^{i=k} (-1)^i F(S; ...; \mu(S_{i-1} \circ S_i); S_{i+1}; ...; S_k; \mathbb{P})$$

$$+ (-1)^{k+1} F(S; ...; S_{k-1}; \mathbb{P})$$

This gives co-homology groups $H_\tau^*(\mathcal{S}, \mathcal{P})$, τ for tree. The fact that entropy $H(S_*\mathbb{P}) = H(S; \mathbb{P})$ defines a 1-cocycle is a result of an equation of Fadeev, generalized by Baez, Fritz and Leinster [3], who gave another interpretation, based on the operad structure of the set of all finite probability laws. See also Marcolli and Thorngren [40].

Theorem E. (*cf.* Theorem 4 section 6.3, [7]): If Ω has more than four points, $H_\tau^1(\Pi(\Omega), \Delta(\Omega))$ is the one dimensional vector space generated by the entropy.

Another co-boundary δ_t on $\mathcal{C}_r^*(\mathcal{M})$ corresponds to another right action of the monad $\mathcal{V}_X$, which is deduced from the maps θ_t that send $(P, i, m) \otimes S_1 \otimes ... \otimes S_m)$ to the sum of the vectors $(P, (i, j), n)$ for $j = 1, ..., n_i$ that are associated to the end branches of S_i. It gives a twisted version of information co-homology as we have done in the first paragraph. This allows us to define higher information quantities for strategies: for $N = 2M + 1$ odd, $I_{\tau,N} = -(\delta\delta_t)^M H$, and for $N = 2M + 2$ even, $I_{\tau,N} = \delta_t(\delta\delta_t)^M H$.

This gives for $N = 2$, a notion of mutual information between a variable S of length m and a collection T of m variables $T_1, ..., T_m$:

$$(8) \qquad I_\tau(S; T; \mathbb{P}) = \sum_{i=1}^{i=m} (H(T_i; \mathbb{P}) - \mathbb{P}(S = i)H(T_i; \mathbb{P}|S = i)).$$

When all the T_i are equals we recover the ordinary mutual information of Shannon plus a multiple of the entropy of T_i.

1.6. **The Forms of Information Strategies.**

A rooted tree Γ decorated by $\mathcal{S}_*$ can be seen as a strategy to discriminate between points in Ω. For each vertex s there is a minimal set of chained edges $\alpha_1, ..., \alpha_k$ connecting s_0 to s; the cardinal k is named the *level* of s; this chain defines a sequence $(F_0, v_0; F_1, v_1; ...; F_{k-1}, v_{k-1})$ of observables and values of them; then we can associate to s the subset Ω_s of Ω where each F_j takes the value v_j. At a given level k the sets Ω_s form a partition π_k of Ω; the first one π_0 is the unit partition of length 1, and π_l is finer than π_{l-1} for any l. By recurrence over k it is easy to deduce from the orderings of the values of F_s an embedding in the Euclidian plane of the subtrees $\Gamma(k)$ at level k such that the values of the variables issued from each vertex are oriented in the direct trigonometric sense, thus π_k has a canonical ordering ω_k. Remark that many branches of the tree gives the empty set for Ω_s after some level; we name them dead branches. It is easy to prove that the set $\Pi(\mathcal{S})_*$ of ordered partitions that can be obtained as a (π_k, ω_k) for some tree Γ and some level k is closed by the natural ordered join operation, and, as $\Pi(\mathcal{S})_*$ contains π_0, it forms a monoid, which contains the monoid $M(\mathcal{S}_*)$ generated by $\mathcal{S}_*$.

Complete discrimination of Ω by $\mathcal{S}_*$ exists when the final partition of Ω by singletons is attainable as a π_k; optimal discrimination correspond to minimal level k. When the set Ω is a subset of the set of words $x_1, ..., x_N$ with letters x_i belonging to given sets M_i of respective cardinalities m_i, the problem of optimal discrimination by observation strategies Γ decorated by $\mathcal{S}_*$ is equivalent to a problem of minimal rewriting by words of type $(F_0, v_0), (F_1, v_1), ..., (F_k, v_k)$; it is a variant of optimal coding, where the alphabet is given. The topology of the poset of discriminating strategies can be computed in terms of the free Lie algebra on Ω, *cf.* [20].

Probabilities $\mathbb{P}$ in $\mathcal{P}$ correspond to *a priori* knowledge on Ω. In many problems $\mathcal{P}$ is reduced to one element, that is the uniform law. Let s be a vertex in a strategic tree Γ, and let $\mathcal{P}_s$ be the set of probability laws that are obtained by conditioning through the equations $F_i = v_i; i = 0, ..., k-1$ for a minimal chain leading from s_0 to s. We can consider that the sets $\mathcal{P}_s$ for different s along a branch measure the evolution of knowledge when applying the strategy. The entropy $H(F; \mathbb{P}_s)$ for F in $\mathcal{S}_*$ and $\mathbb{P}_s$ in $\mathcal{P}_s$ gives a measure of information we hope to obtain when applying F at s in the state $\mathbb{P}_s$. The maximum entropy algorithm consists in choosing at each vertex s a variable that has the maximal conditioned entropy $H(F; \mathbb{P}_s)$.

Theorem F. (*cf.* [7]): To find one false piece of different weight among N pieces for $N \geq 3$, when knowing the false piece is unique, by the minimal numbers of weighing, one can use the maximal entropy algorithm.

However we have another measure of information of the resting ambiguity at s, by taking for the *Galois group* G_s the set of permutations of Ω_s which respects globally the set $\mathcal{P}_s$ and the set of restrictions of elements of $\mathcal{S}_*$ to Ω_s, and which preserves one by one the equations $F_i = v_i$. Along branches of Γ this gives a decreasing sequence of groups, whose successive quotients measure the evolution of acquired information in an algebraic sense.

Problem 3. Generalize Theorem F. Can we use algorithms based on the Galoisian measure of information? Can we use higher information quantities associated to trees for optimal discrimination?

1.7. **Conclusion and Perspective.**

Concepts of Algebraic topology were recently applied to Information theory by several researchers. In particular notions coming from category theory, homological algebra and differential geometry were used for revisiting the nature and scope of entropy, *cf.* for instance Baez *et al.* [3], Marcolli and Thorngren [40] and Gromov [24]. In the present note we interpreted entropy and Shannon information functions as co-cycles in a natural co-homology theory of information, based on categories of observable and complexes of probability. This allowed us to associate topological figures, like Borromean links, with particular configuration of mutual dependency of several observable quantities. Moreover we extended these results to a dynamical setting of system observation, and we connected probability evolutions with the measures of ambiguity given by Galois groups. All those results provide only the first steps toward a developed Information Topology. However, even at this preliminary stage, this theory can be applied to the study of distribution and evolution of Information in concrete physical and biological systems. This kind of approach already proved its efficiency for detecting collective synergic dynamic in neural coding [12], in genetic expression [54], in cancer signature [32], or in signaling pathways [53]. In particular, information topology could provide the principles accounting for the structure of information flows in biological systems and notably in the central nervous system of animals.

2. CLASSICAL INFORMATION TOPOS. THEOREM ONE

2.1. **Information Structures and Probability Families.**

Let Ω be a finite set, the set $\Pi(\Omega)$ of all partitions of Ω constitutes a category with one arrow $Y \to Z$ from Y to Z when Y is more fine than Z, we also say in this case that Y divides Z. In $\Pi(\Omega)$ we have an initial element, which is the partition by points, denoted ω and a final element, which is Ω itself and is denoted by 1. The joint partition YZ or (Y, Z), of two partitions Y, Z of Ω is the less fine partition that divides Y and Z, *i.e.*, their gcd. For any X we get $XX = X$, $\omega X = \omega$ and $1.X = X$.

By definition an *information structure* $\mathcal{S}$ on Ω is a subset of $\Pi(\Omega)$, such that for any element X of $\mathcal{S}$, and any pair of elements Y, Z in $\mathcal{S}$ that X refines, the joint partition YZ also belongs to $\mathcal{S}$.
In addition we will always assume that the final partition 1 belongs to $\mathcal{S}$. In terms of observations, it means that at least something is a certitude.

Examples: start with a set $\Sigma = \{S_i; 1 \leq i \leq n\}$ of partitions of Ω. For any subset $I = \{i_1, ..., i_k\}$ of $[n] = \{1, ..., n\}$, the joint $(S_{i_1}, ..., S_{i_k})$, also denoted S_I, divides each S_{i_j}. The set $W = W(\Sigma)$ of all the S_I, when I describes the subsets of $[n]$ is an information struture. It is even a commutative monoid, because any product of elements of W belongs to W, and the partition associated to Ω itself gives the identity element of W. The product $S_{[n]}$ of all the S_i is maximal; it divides all the other elements. As $\Pi(\Omega)$ the monoid $W(\Sigma)$ is idempotent, *i.e.*, for any X we have $XX = X$.

By definition, the faces of the abstract simplex $\Delta([n])$ are the subsets of $[n]$; its vertices are the singletons. Thus the monoid $W(\Sigma)$ can be identified with the first barycentric subdivision of the simplex $\Delta([n])$.

Remind that a simplicial subcomplex of $\Delta([n])$ is a subset of faces that contains all faces of any of its elements. Then any simplicial sub-complex K of $\Delta([n])$ gives

a simplicial information structure $\mathcal{S}(K)$, embedded in $W(\Sigma)$. In fact, if Y and Z are faces of a simplex X belonging to K, YZ is also a face in X, thus it belongs to K. The maximal faces $\Sigma_a; a \in A$ of K correspond to the finest elements in $\mathcal{S}(K)$; the vertices of a face Σ_a gives a family of partitions, which generates a sub-monoid $W_a = W(\Sigma_a)$ of W; it is a sub-information structures (full sub-category) of $\mathcal{S}(K)$, having the same unit, but having its own initial element ω_a. These examples arise naturally when formalizing measurements if some obstructions or *a priori* decisions forbid a set of joint measurements.

This kind of examples were considered by Han [26] see also McGill [44].

Example 1. Ω has four elements $(00), (01), (10), (11)$; the variable S_1 (resp. S_2) is the projection pr_1 (resp. pr_2), on $E_1 = E_2 = \{0, 1\}$; Σ is the set $\{S_1, S_2\}$. The monoid $W(\Sigma)$ has four elements $1, S_1, S_2, S_1 S_2$. The partition $S_1 S_2 = S_2 S_1$ corresponds to the variable $Id : \Omega \to \Omega$.

Example 2. Same Ω as before, with the same names for the elements, but we take all the partitions of Ω in $\mathcal{S}$. In addition to $1, S_1, S_2$ and $S = S_1 S_2$, there is S_3, the last partition in two subsets of cardinal two, which can be represented by the sum of the indices: $S_3(00) = 0, S_3(11) = 0, S_3(01) = 1, S_3(10) = 1$, the four partitions Y_ω, for $\omega \in \Omega$, formed by a singleton $\{\omega\}$ and its complementary, and finally the six partitions $X_{\mu,\nu} = Y_\mu Y_\nu$, indexed by pairs of points in Ω satisfying $\mu < \nu$ in the lexical order. The product of two distinct Y is a X, the product of two distinct X or two distinct S_i is S, the product of one Y and a S_i is a X, of one Y and a X is this X or S, of one S and a X is this X or S. In particular the monoid W is also generated by the three S_i and the four Y_ω; it is called the monoid of partitions of Ω, and the associative algebra $\Lambda(\mathcal{S})$ of this monoid is called the partition algebra of Ω.

Example 3. Same Ω as before, that is $\Omega = \Delta(4)$, with the notations of example 2 for the partitions; but we choose as generating family the set Υ of the four partitions $Y_\mu; \mu \in \Omega$; the joint product of two such partitions is either a Y_μ (when they coincide) or a $X_{\mu\nu}$ (when they are different). The monoid $W(\Upsilon)$ has twelve elements.

Example 4. Ω has 8 elements, noted $(000), ..., (111)$, and we consider the family Σ of the three binary variables S_1, S_2, S_3 given by the three projections. If we take all the joints, we have a monoid of eight elements. However, if we forbid the maximal face (S_1, S_2, S_3), we have a structure $\mathcal{S}$ which is not a monoid; it is the set formed by $1, S_1, S_2, S_3$ and the three joint pairs $(S_1, S_2), (S_1, S_3), (S_2, S_3)$.

On the side of probabilities, we choose a Boolean algebra $\mathcal{B}$ of sets in Ω, *i.e.*, a subset $\mathcal{B}$ of the set $\mathcal{P}(\Omega)$ of subsets of Ω that contains the empty set $\emptyset$ and the full set Ω, and is closed by union and intersection. In this finite context, it is easy to prove that $\mathcal{B}$ is constituted by all the unions of its minimal elements (called atoms). Associated to this case, we will consider only information structures that are made by partitions whose each element belongs to $\mathcal{B}$. Consequently we could replace everywhere Ω by the finite set $\Omega_\mathcal{B}$ of the atoms of $\mathcal{B}$, but we will see that several Boolean sub-algebras appear naturally in the process of observation, thus we prefer to mention the choice of $\mathcal{B}$ at the beginning of observations.

Then we consider the set $\Delta(\Omega_\mathcal{B})$, or $\Delta(\mathcal{B})$, of all probability laws on $(\Omega, \mathcal{B})$, *i.e.*, all real functions p_x of the atoms x of $\mathcal{B}$ (the points of $\Omega_\mathcal{B}$), satisfying $p_x \geq 0$ and

$\sum_x p_x = 1$. We see that this set of probabilities is also a simplex $\Delta([N])$, where N is the cardinality of $\Omega_{\mathcal{B}}$.

As on the side of partitions, we will consider more generally any simplicial sub-complex $\mathcal{Q}$ of $\Delta(\mathcal{B})$, and call it a probability complex. In the appendix, we show that this kind of examples correspond to natural forbidding rules, that can express physical constraints on the observed system.

A partition Y which is measurable with respect to $\mathcal{B}$ is made by elements Y_i for $i = 1, ..., m$, belonging to $\mathcal{B}$. Let P be an element of $\Delta(\mathcal{B})$; the conditioning of P by the element Y_i is defined only if $P(Y_i) \neq 0$, and given by the formula $P(B|Y = y_i) = P(B \cap Y_i)/P(Y_i)$. We will consider it as a probability on Ω equipped with $\mathcal{B}$, not as a probability on Y_i. Remark that if P belongs to a simplicial family $\mathcal{Q}$, the probability $P(B|Y = y_i)$ is also contained in $\mathcal{Q}$. In fact, if the smallest face of $\mathcal{Q}$ which contains P is the simplex σ on the vertices $x_1, ..., x_k$, then the conditioning of P by Y_i, being equal to 0 for the other atoms x, belongs to a face of σ, which is in $\mathcal{Q}$, because $\mathcal{Q}$ is a complex.

For a probability family $\mathcal{Q}$, *i.e.*, a set of probabilities on Ω, and a set of partitions $\mathcal{S}$, we say that $\mathcal{Q}$ and $\mathcal{S}$ are adapted one to each other if the conditioning of every element of $\mathcal{Q}$ by every element of S belongs to $\mathcal{Q}$.

By definition, the algebra $\mathcal{B}_Y$ is the set of unions of elements of the partition Y. We can consider it as a Boolean algebra on Ω contained in $\mathcal{B}$ or as Boolean algebra on the quotient set Ω/Y.

The image $Y_* Q$ of a probability Q for $\mathcal{B}$ by the partition Y is the probability on Ω for the sub-algebra $\mathcal{B}_Y$, that is given by $Y * Q(t) = Q(t)$ for $t \in \mathcal{B}_Y$. It is the forgetting operation, also frequently named marginalization by Y.

By definition, the set $\mathcal{Q}_Y$ is the image of Y_*. Let us prove that it is a simplicial sub-complex of $\Delta(\mathcal{B}_Y)$: take a simplex σ of $\mathcal{Q}$, denote its vertices by $x_1, ..., x_k$, note δ_j the Dirac mass of x_j, and look at the partition $\sigma_i = Y_i \cap \sigma$ of σ induced by Y, then for all the $x_j \in \sigma_i$ the images $Y_* \delta_j$ coincide. Let us denote this image by $\delta(Y, \sigma_i)$; it is an element of $\mathcal{Q}_Y$. For every law Q in σ, the image $Y_* Q$ belongs to the simplex on the laws $\delta(Y, \sigma_i)$, and any point in this simplex belongs to $\mathcal{Q}_Y$. Q.E.D.

If $X \to Y$ is an arrow in $\Pi(\Omega_{\mathcal{B}})$, the above argument shows that the map $\mathcal{Q}_X \to \mathcal{Q}_Y$ is a simplicial mapping.

Conditioning by Y and marginalization by Y_* are related by the barycentric law (or theorem of total probability, Kolmogorov 1933 [34]): for any measurable set A in $\mathcal{B}$ we have

$$(9) \qquad P(A) = P(Y = y_1)P|(Y = y_1)(A) + ... + P(Y = y_m)P|(Y = y_m)(A).$$

Remark that the notions of information structures and probability complexes extend to infinite sets; this is developed in paper [5].

In this context, we have a formula for any integrable function φ on Ω with respect to P:

$$(10) \qquad \int_{\Omega} \varphi(\omega)dP(\omega) = \int_{\Omega/Y} d(Y_* P)(\omega') \int_{\Omega} \varphi(\omega)d(P|(Y = \omega'))(\omega).$$

Consider a finite set Ω, equipped with a Boolean algebra $\mathcal{B}$, a probability family $\mathcal{Q}$ for it and an information structure $\mathcal{S}$ adapted to $\mathcal{B}$.

For each object X in $\mathcal{S}$, the set $\mathcal{S}_X$ made by the partitions Y that are divided by X is a closed sub-category, possessing an internal law of monoid. The object X

is initial. To any arrow $X \to Y$ is associated the inclusion $\mathcal{S}_Y \to \mathcal{S}_X$, thus we get a contra-variant functor from $\mathcal{S}$ to the category of monoids.

On the other side we have a natural co-variant functor of $\mathcal{S}$ to the category of sets, which associates to each partition $X \in \mathcal{S}$ the set $\mathcal{Q}_X$ of probability laws in the image of $\mathcal{Q}$ on the quotient set Ω/X, and which associates to each arrow $X \to Y$ the surjection $\mathcal{Q}_X \to \mathcal{Q}_Y$ which is given by direct image $P_X \mapsto Y_* P_X$. If $\mathcal{Q}$ is simplicial the functor goes to the category of simplicial complexes.

Definition 1. For $X \in \mathcal{S}$, the *functional module* $\mathcal{F}_X(\mathcal{Q})$ is the real vector space of measurable functions on the space $\mathcal{Q}_X$; for each arrow of divisibility $X \to Y$, we have an injective linear map $f \mapsto f^{Y|X}$ from $\mathcal{F}_Y$ to $\mathcal{F}_X$, given by

$$(11) \qquad f^{Y|X}(P_X) = f(Y_* P_X).$$

In this manner, we obtain a contra-variant functor $\mathcal{F}$ from the category $\mathcal{S}$ to the category of real vector spaces.

If $\mathcal{Q}$ and $\mathcal{S}$ are adapted one to each other, the functor $\mathcal{F}$ admits a canonical action of the monoid functor $X \mapsto \mathcal{S}_X$, given by the average formula

$$(12) \qquad (Y.f)(P) = \int dY_* P(y) f(P|(Y = y)).$$

To verify this is an action of monoid, we must verify that for any Z which divides Y, and any $f \in \mathcal{F}_Y$, we have, in $\mathcal{F}_X$ the identity

$$(13) \qquad (Z.f)^{Y|X} = Z.(f^{Y|X});$$

that means, for any $P \in \mathcal{Q}_X$:

$$(14) \qquad \int_{E_Z} dZ_* P(z) f^{Y|X}(P|(Z = z)) = \int_{E_Z} dZ_* P(z) f((Y_* P)|(Z = z)).$$

But this results from the identity $Y_*(P|(Z = z)) = (Y_* P)|(Z = z)$ due to $Y_* P(Z = z) = P(Z = z)$.

The arrows of direct images and the action of averaged conditioning satisfy the axiom of distributivity: if Y and Z divide X, but not necessarily Z divides Y, we have

$$(15) \qquad Z.(f^Y)(P, X) = (Z.f)((Z, Y)_* P, (Y, Z)) = (Z.f)^{(Z,Y)}(P, X).$$

Proof. The first identity comes from the fact that $(Z, Y)_*(P|(Z = z)) = Y_*(P|(Z = z))$; the second one follows from the fact that we have an action of the monoid $\mathcal{S}_X$.

As the formula (12) is central in our work, we insist a bit on it, and comment its meaning, at least in this finite setting:

Let $P \mapsto f(P)$ be an element of $\mathcal{F}_X$, and Y be the goal of an arrow $X \to Y$, we have

$$(16) \qquad Y.f(P) = \sum_j \mathbb{P}(Y = y_j) f(\mathbb{P}|Y = j).$$

where j describes the indices of the partition Y.

We will see when discussing functions of several partitions that this formula is due to Shannon and correspond to conditional information.

Lemma 1. for any pair (Y, Z) of variables in $\mathcal{S}_X$, and any F for which the integrals converge, we have $(Y, Z).F = Y.(Z.F)$.

Proof. We note p_i the probability that $Y = y_i$, π_{ij} the joint probability of $(Y = y_i, Z = z_j)$, and q_{ij} the conditional probability of $Z = z_j$ knowing that $Y = y_i$, then

$$
\begin{aligned}
(Y, Z).F(\mathbb{P}) &= \sum_i \sum_j \pi_{ij} F(\mathbb{P}|(Y = y_i, Z = z_j)) \\
&= \sum_i p_i (\sum_j q_{ij} F(\mathbb{P}|(Y = y_i, Z = z_j)) \\
&= \sum_i p_i (\sum_j q_{ij} F(\mathbb{P}|(Y = y_i))|(Z = z_j)) \\
&= \sum_i p_i (Z.F)(\mathbb{P}|(Y = y_i)) \\
&= Y.(Z.F)(\mathbb{P}).
\end{aligned}
$$

Remark 1. In the general case, where Ω is not necessarily finite and $\mathcal{B}$ is any sigma-algebra, the Lemma 1 is a version of the Fubini theorem.

Let us consider the category $\mathcal{S}$ equipped with the discrete topology, to get a site (*cf.* SGA [1]). Over a discrete site every presheaf is a sheaf. The contravariant functor $X \mapsto \mathcal{S}_X$ gives a structural sheaf of monoids, and by passing to the algebras $\mathcal{A}_X$ over $\mathbb{R}$ which are generated by the (finite) monoids, we get a sheaf in rings, thus S becomes a ringed site. Moreover, by considering all contra-variant functors $X \mapsto \mathcal{N}_X$ from $\mathcal{S}$ to modules over the algebra functor $\mathcal{A}$, we obtain a ringed topos, that we name the *information topos* associated to $\Omega, \mathcal{B}, \mathcal{S}$. This ringed topos concerns only the observables given by partitioning.

Take now in account a probability family $\mathcal{Q}$ which is adapted to $\mathcal{S}$, for instance a simplicial family; we obtain a functor $X \mapsto Q_X$ translating the marginalization by the partitions, considered as observable quantities, and the conditioning by observables is translated by a special element $X \mapsto \mathcal{F}_X$ of the information topos.

In this way it is natural to expect that topos co-homology, as introduced by Grothendieck, Verdier and their collaborators (see SGA 4 [1]), captures the invariant structure of observation, and defines in this context what information is. This is the main outcome of our work.

As a consequence of Grothendieck's article (Tohoku, 1957 [25]), a ringed topos possesses enough injective objects, *i.e.*, any object is the sub-object of an injective object, moreover, up to isomorphism, there is a unique minimal injective object containing a given object, called its injective envelope (*cf.* Gabriel, seminaire Dubreil, exp. 17 [21]). Thus each object in the category $\mathcal{D}_{\mathcal{S}}$ of modules over a ringed site $\mathcal{S}$ possesses a canonical injective resolution $I_*(N)$; then the group $Ext^n_{\mathcal{D}}(M, N)$ can be defined as the homology of the complex $Hom_{\mathcal{D}}(M, I_n(N))$. Those groups are denoted by $H^n(M; N)$.

The "comparison theorem" (*cf.* Bourbaki, Alg.X Th1, p.100 [11], or MacLane 1975, p. 261 [38]) asserts that, for any projective (resp. injective) resolution of M (resp. N) there exists a natural map of complexes between the resulting complex of homomorphisms and the above canonical complex, and that this map induces an isomorphism in co-homology.

In our context, we take for M the trivial constant module $\mathbb{R}_{\mathcal{S}}$ over $\mathcal{S}$, and we take for N the functional module $\mathcal{F}(\mathcal{Q})$.

The existence of free resolutions of $\mathbb{R}_{\mathcal{S}}$ makes things easier to handle.

Hence we propose that the natural information quantities are classes in the co-homology groups $H^*(\mathbb{R}_{\mathcal{S}}, \mathcal{F}(\mathcal{Q}))$.

This is reminiscent of Galois co-homology see SGA [1], where M is also taken as the constant sheaf over the category of G-objects seen as a site.

In [5] we develop further this more geometric approach, by considering several resolutions. But in this paper, in order to be concrete, we will only focus on a more elementary approach, associated to a special resolution, called the non-homogeneous bar-resolution, which also leads to the general result. This is the object of the next section.

2.2. Non-Homogeneous Information Co-Homology.

For each relative integer $m \geq 0$, and each object $X \in \mathcal{S}$, we consider the real vector space $S_m(X)$, freely generated by the m-uples of elements of the monoid $\mathcal{S}_X$, and we define $C^m(X)$ as the real vector space of linear functions from $S_m(X)$ to the space $\mathcal{F}_X$ of measurable functions from $\mathcal{Q}_X$ to $\mathbb{R}$.

Then we define the set $\mathcal{C}^m$ of m-cochains as the set of collections $F_X \in C^m(X)$ satisfying the following condition, named *joint locality*:

For each Y divided by X, when each variable X_j is divided by Y, we must have

$$(17) \qquad F_Y(X_1; ...; X_m; Y_*\mathbb{P}) = F_X(X_1; ...; X_m; \mathbb{P}).$$

Thus a co-chain F is a natural transformation from the functor $S_m(X)$ from $\mathcal{S}$ to the category of real vector spaces to the functor $\mathcal{F}$ of measurable functions on $\mathcal{Q}_X$. Hence, F is not an ordinary numerical function of probability laws $\mathbb{P}$ and a set $(X_1, ..., X_m)$ of m random variables, but we can speak of its value $F_X(X_1; ...; X_m; \mathbb{P})$ for each X in $\mathcal{S}$. For X given the co-chains form a sub-vector space $\mathcal{C}^m(X)$ of $C^m(X)$.

If we apply the condition to $Y = (X_1, ..., X_m)$ we find that $F(X_1; ...; X_m; \mathbb{P})$ depends only on the direct image of $\mathbb{P}$ by the joint variable of the X_i's. This implies that, if F belongs to $\mathcal{C}^m(X)$, we have

$$(18) \qquad F(X_1; ...; X_m; \mathbb{P}) = F(X_1; ...; X_m; (X_1...X_n)_*\mathbb{P}),$$

Conversely, suppose that F satisfies the conditions (18) and consider X, Y two variables such that X divides Y, and that Y divides each X_j, and let P be a probability in $\mathcal{Q}_X$; then the joint variable $Z = (X_1, ..., X_m)$ divides Y and X, thus we have $Z_*P = Z_*(X_*P) = Z_*(Y_*P)$, and

$$(19) \qquad F(X_1; ...; X_m; Y_*P) = F(X_1; ...; X_m; Z_*P) = F(X_1; ...; X_m; X_*P).$$

Which proves that F belongs to $\mathcal{C}^m(X)$.

Let F be an element of $\mathcal{C}^m(X)$, and Y an element of $\mathcal{S}_X$; then we define

$$(20) \qquad Y.F(X_1; ...; X_m; \mathbb{P}) = \sum \mathbb{P}(Y = y_j) F(X_1; ...; X_m; \mathbb{P}|Y = y_j).$$

It follows from the equivalent condition (18) that $Y.F$ also belongs to $\mathcal{C}^m(X)$.

Moreover, the proof of Lemma 1 applies and give that, for any pair (Y, Z) of variables in $\mathcal{S}_X$, and any F in $\mathcal{C}^m(X)$, we have $(Y, Z).F = Y.(Z.F)$.

Thus (1) defines an action of the semigroup $\mathcal{S}_X$ on the vector spaces $\mathcal{C}^m(X)$.

Remark 2. The operation of $\mathcal{S}_X$ can be rewritten more compactly by using integrals:

$$(21) \qquad Y.F(X_1; ...; X_m; \mathbb{P}) = \int_\Omega F(X_1; ...; X_m; \mathbb{P}|(Y = Y(\omega)))dP(\omega).$$

The differential δ for computing co-homology is given by the Eilenberg-MacLane formula (1943):

(22)
$$\delta^m F(Y_1; ...; Y_{m+1}; P)$$

$$= Y_1.F(Y_2; ...; Y_{m+1}; P) + \sum_{1}^{m} (-1)^i F(...; (Y_i, Y_{i+1}); ...; Y_{m+1}; P) + (-1)^{m+1} F(Y_1; ...; Y_m; P).$$

Since this formula corresponds to the standard inhomogeneous bar-resolution in the case of semi-groups and algebras (Cf. MacLane p.115 [39] and Cartan-Eilenberg pp.174–175. [13]), we name δ the Hochschild co-boundary, as in the case of semi-groups, and algebras.

Remark that a function F satisfying the joint locality condition, (*i.e.*, the hypothesis that $F(Y_1; ...; Y_m; P)$ depends only on $(Y_1, ..., Y_m)_* P$), has a co-boundary which is also jointly local, because the variables appearing in the definition are all joint variables of the Y_j. (This this would not have been true for the stronger locality hypothesis asking that F depends only on the collection $(Y_j)_* P; j = 1, ..., m$.)

It is easy to verify that $\delta^m \circ \delta^{m-1} = 0$. We denote by Z^m the kernel of δ^m and by B^m the image of δ^{m-1}. The elements of Z^m are named m-cocycles, we consider them as *information quantities*, and the elements of B^m are m-coboundaries.

Definition 2. For $m \geq 0$, the quotient

(23)
$$H^m(\mathcal{C}^*) = Z^m / B^m$$

is the m-th *cohomology group of information* of the information structure $\mathcal{S}$ on the simplicial family of probabilities $\mathcal{Q}$. We denote it by $H^m(\mathcal{S}; \mathcal{Q})$.

The information co-homology satisfies functoriality properties:

Consider two pairs of information structures and probability families, $(\mathcal{S}, \mathcal{Q})$ and $(\mathcal{S}', \mathcal{Q}')$ on two sets Ω, Ω' equipped with the σ-algebras $\mathcal{B}, \mathcal{B}'$ respectively, and φ a *surjective* measurable map from $(\Omega, \mathcal{B})$ to $(\Omega', \mathcal{B}')$, such that $\varphi_*(\mathcal{Q}) \subseteq \mathcal{Q}'$ (*i.e.*, $\varphi_*(Q) \in \mathcal{Q}'$ for every $Q \in \mathcal{Q}$), and such that $\mathcal{S} \subseteq \varphi^* \mathcal{S}'$ (*i.e.*, $\forall X \in \mathcal{S}, \exists X' \in \mathcal{S}', X = X' \circ \varphi$); then we have the following construction:

Proposition 1. For each integer $m \geq 0$, a natural linear map

(24)
$$\varphi^* : H^m(\mathcal{Q}'; \mathcal{S}') \to H^m(\mathcal{Q}; \mathcal{S}),$$

is defined by the following application at the level of local co-chains:

(25)
$$\varphi^*(F')(X_1; ...; X_m; P) = F'(X_1'; ...; X_m'; \varphi_*(P)),$$

for a collection of variables $X_j'; j = 1, ..., m$ satisfying $X_j = X_j' \circ \varphi$ for each j.

Proof. First, remark that $X_j = X"_j \circ \varphi$ implies $X_j' = X"_j$ because φ is surjective. As F' is (jointly) local, the co-chain $F = \varphi^*(F')$ is also (jointly) local. Finally, it is evident that the map $F' \mapsto F$ commutes with the co-boundary operator. Therefore the proposition follows.

Another co-homological construction works in the reversed direction:

Consider two information structures $(\mathcal{S}, \mathcal{Q})$ and $(\mathcal{S}', \mathcal{Q}')$ on two sets Ω, Ω' equipped with σ-algebras $\mathcal{B}, \mathcal{B}'$ respectively, and φ a measurable map from $(\Omega, \mathcal{B})$ to $(\Omega', \mathcal{B}')$,

such that $\mathcal{Q}' \subseteq \varphi_*(\mathcal{Q})$ (*i.e.*, $\forall Q' \in \mathcal{Q}', \exists Q \in \mathcal{Q}, Q' = \varphi_*(Q)$), and such that $\varphi^*\mathcal{S}' \subseteq \mathcal{S}$ (*i.e.*, $\forall X' \in \mathcal{S}', X = X' \circ \varphi \in \mathcal{S}$); then the following result is true:

Proposition 2. For each integer $m \geq 0$, a natural linear map

(26) $$\varphi_* : H^m(\mathcal{Q}'; \mathcal{S}') \to H^m(\mathcal{Q}; \mathcal{S}),$$

is defined by the following application at the level of co-chains:

(27) $$\varphi_*(F)(X_1'; ...; X_m'; P') = F(X_1' \circ \varphi; ...; X_m' \circ \varphi; P),$$

for a probability law $P \in \mathcal{Q}$ and its image $P' = \varphi_*(P)$.

Proof. First, remark that, if Q also satisfies $P' = \varphi_*(Q)$, we have $F(X_1' \circ \varphi; ...; X_m' \circ \varphi; P) = F(X_1' \circ \varphi; ...; X_m' \circ \varphi; Q)$. To establish that point, let us denote $X_j = X_j' \circ \varphi; j = 0, ..., m$, and $X' = (X_1', ..., X_m')$, $X = (X_1, ..., X_m)$ the joint variables; the quantity $F(X_1' \circ \varphi; ...; X_m' \circ \varphi; P)$ depends only on X_*P, but this law can be rewritten $X_*'P'$, which is also equal to X_*Q. In particular, if F is local, then $F' = \varphi_*F$ is local.

As it is evident that the map $F \mapsto F'$ commutes with the co-boundary operator, the proposition follows.

Remark this way of functoriality uses the locality of co-cycles.

Corollary 1. In the case where $\mathcal{Q}' = \varphi_*(\mathcal{Q})$ and $\mathcal{S} = \varphi^*\mathcal{S}'$, the maps φ^* and φ_* in information co-homology are inverse one of each other.

This is our formulation of the invariance of the information co-homology for equivalent information structures.

When $m = 0$, co-cochains are functions f of P_X in $\mathcal{Q}_X$ such that $f(Y_*P_X) = f(P_X)$ for any Y multiple of X (*i.e.*, coarser than X). As we assume 1 belongs to $\mathcal{S}$, and the set Q_1 has only one element, f must be a constant. And every constant is a co-cycle, because

(28)
$$\delta.f(X_0; P) = X_0.f(P) - f(P) = \sum_j P(X_0 = x_j)f(P|X_0 = x_j) - f(P) = f(1)(1-1) = 0.$$

Consequently H^0 is $\mathbb{R}$. This corresponds to the hypothesis $1 \in \mathcal{S}$, meaning connexity of the category. If m components exist, we recover them in the same way and H^0 is isomorphic to $\mathbb{R}^m$.

We now consider the case $m = 1$. From what precedes we know that there is no non-trivial co-boundary.

Non-homogeneous 1-cocycles of information are families of functions $f_X(Y; P_X)$, measurable in the variable P in $\mathcal{Q}$, labelled by elements $Y \in \mathcal{S}_X$, which satisfies the locality condition, stating that each time we have $Z \to X \to Y$ in $\mathcal{S}$, we have

(29) $$f_X(Y; X_*P_Z) = f_Z(Y; P_Z)$$

and the co-cycle equation, stating that for two elements Y, Y' of $\mathcal{S}_X$, we have

(30) $$f((Y, Y'); P) = f(Y; P) + Y.f(Y'; P).$$

Remark that locality implies that it is sufficient to know the $f_Y(Y; Y_*P)$ to recover $f_X(Y; P)$ for all partition X in $\mathcal{S}$ that divides Y.

It is in this sense that we frequently omit the index X in f_X.

Remark also that for any 1-cocycle f we have $f(1; P) = 0$.

In fact, the co-cycle equation tells that

(31) $$f((1,1);P) = f(1;P) + 1.f(1;P).$$

but

(32) $$1.f(1;P) = f(1;P|1=1) = f(1;P),$$

and $(1,1) = 1$, thus $f(1;P) = 0$.

More generally, for any X, and any value x_i of X, we have

(33) $$f(X;P|(X=x_i)) = 0,$$

In fact a special case of Equation (30) is

(34) $$f((X,X);P) = f(X;P) + X.f(X;P).$$

which implies $X.f(X;P) = 0$; however, by definition,

(35) $$X.f(X;P) = \sum_i P(X=x_i)f(X;P|(X=x_i)),$$

thus for every i we must have $f(X;P|(X=x_i)) = 0$, due to $P \geq 0$. This generalizes $f(1;P) = 0$ for any P, because, for a probability conditioned by $X = x_i$, the partition X appears the same as 1, that is a certitude.

Remark also that for each pair of variables (X,Y), a 1-cocycle must satisfy the following symmetric relation:

(36) $$f(Y;\mathbb{P}) - Z.f(Y;\mathbb{P}) = f(Z;\mathbb{P}) - Y.f(Z;\mathbb{P}).$$

2.3. Entropy.

Any multiple of the Shannon entropy is a non-homogeneous information co-cycle. Remind that entropy H is defined for one partition X by the formula

(37) $$H(X;\mathbb{P}) = -\sum_i p_i \log p_i,$$

where the p_i denotes the values of $\mathbb{P}$ on the elements of the partition X. In particular the function H depends only on $X_*(\mathbb{P})$, which is locality. The co-cycle equation expresses the fundamental property for an information quantity, writen by Shannon:

(38) $$H(X,Y) = H(X) + H_X(Y)$$

Thus every constant multiple $f = \lambda H$ of H defines a co-cycle.
Remark that the corresponding "homogeneous 1-cocycle" is the *entropy variation*:

(39) $$F(X;Y;\mathbb{P}) = H(X;\mathbb{P}) - H(Y;\mathbb{P}).$$

This means that it satisfies the "invariance property":

$$\begin{aligned}
F((Z,X);(Z,Y)) &= H(Z,X) - H(Z,Y) \\
&= H(Z) + H_Z(X) - H(Z) - H_Z(Y) \\
&= Z.F(X;Y),
\end{aligned}$$

and the "simplicial equation":

(40) $$F(Y;Z) - F(X;Z) + F(X;Y) = 0$$

Note that the entropy variation $H(X;P) - H(Y;P)$ exists in a wider range of condition, *i.e.*, when Ω is infinite, if the laws of X and Y are absolutely continuous with respect to a same probability law $\mathbb{P}_0$: we only have to replace the finite sum by the integral of the function $-\varphi \log \varphi$ where φ denotes the density with respect

to $\mathbb{P}_0$. Changing the reference law $\mathbb{P}_0$ changes the quantities $H(X)$ and $H(Y)$ by the same constant, thus does not change the variation $H(X; P) - H(Y; P)$.

We will prove now that, for many *simplicial* structures $\mathcal{S}$, and sufficiently large adapted probability complexes $\mathcal{Q}$, any information co-homology class of degree one is a multiple of the entropy class.

In particular this would be true for $\mathcal{S} = W(\Sigma)$ and $\mathcal{Q} = \Delta(\Omega)$, when Σ has more than two elements and Ω more than four elements, but this is also true in more refined situation, as we will see.

We assume that the functor of probabilities $\mathcal{Q}_X$ contains all the laws on Ω/X, when X belongs to $\mathcal{S}$. In such a case, by definition, we say that $\mathcal{Q}$ is *complete* with respect to $\mathcal{S}$.

Let us consider a probability law P in $\mathcal{Q}$ and two partitions X, Y in the structure $\mathcal{S}$, such that the joint XY belongs to $\mathcal{S}$. We denote by Greek letters $\alpha, \beta, ...$ the indices labelling the partition Y and by Latin letters $k, l, ...$ the indices of the partition X; the probability that $X = \xi_k, Y = \eta_\alpha$ is noted $p_{k,\alpha}$, then the probability of $X = \xi_k$ is equal to $p_k = \sum_\alpha p_{k,\alpha}$ and the probability of $Y = \eta_\alpha$ is equal to $q_\alpha = \sum_k p_{k,\alpha}$.

To simplify the notations, let us write $F = f(X; \mathbb{P})$, $G = f((Y, X); \mathbb{P})$, $H = f(Y; \mathbb{P})$, $F_\alpha = f(X; \mathbb{P}|(Y = \eta_\alpha))$, $H_k = f(Y; \mathbb{P}|(X = \xi_k))$.

The Hochschild co-cycle equation gives

$$(41) \qquad \sum_\alpha q_\alpha F_\alpha \left(\frac{p_{k_1,\alpha}}{q_\alpha}, ..., \frac{p_{k_m,\alpha}}{q_\alpha} \right) = G((p_{k,\alpha})) - H(q_{\alpha_1}, ..., q_{\alpha_n})$$

But we also have the relation obtained by exchanging X and Y, which gives

$$(42) \qquad \sum_k p_k H_k \left(\frac{p_{k,\alpha_1}}{p_k}, ..., \frac{p_{k,\alpha_n}}{p_k} \right) = G((p_{k,\alpha})) - F(p_{k_1}, ..., p_{k_m}).$$

Suppose that $p_{k,\alpha} = 0$ except when $\alpha = \alpha_1$ and $k = k_2, k_3, ..., k_m$ or $\alpha = \alpha_2$ and $k = k_1$; we put $p_{k_i,\alpha_1} = x_i; \quad i = 2, ..., m$ and $p_{k_1,\alpha_2} = x_1$, which implies that we have $x_1 + x_2 + ... + x_m = 1$. Then Equation (33) implies that each term H in Equation (42) is zero, because only one value of the image law is non-zero, thus we can replace the only term G by $F(p_{k_1}, ..., p_{k_m})$, and we get from Equation (41):

$$(43) \quad H(1 - x_1, x_1, 0, ..., 0) = F(x_1, x_2, ..., x_m) - (1 - x_1)F_{\alpha_1}\left(0, \frac{x_2}{1 - x_1}, ..., \frac{x_m}{1 - x_1}\right).$$

Only the term F for α_1 subsists because, the possible other one, for α_2, concerns a certitude.

Consequently, by imposing $x_2 = 1 - x_1 = a, x_3 = ... = x_m = 0$, we deduce the identity $H(a, 1 - a, 0, ..., 0) = F(1 - a, a, 0, ..., 0)$. This gives a recurrence equation to calculate F from the binomial case:

$$(44) \quad F(x_1, x_2, ..., x_m) = F(x_1, 1 - x_1, 0, ..., 0) + (1 - x_1)F\left(0, \frac{x_2}{1 - x_1}, ..., \frac{x_m}{1 - x_1}\right).$$

That is due to the fact that F_{α_1} is a special case of F, thus independent from Y and α_1.

Then coming back to the co-cycle equation, we obtain in particular a functional equation for the binomial variables.

Lemma 2. With the notations of the example 1 (*cf.* example 1), $\Omega = \{(00), (01), (10), (11)\}$, S_1 (resp. S_2) the projection pr_1 (resp. pr_2), on $E_1 = E_2 = \{0, 1\}$, $S = \{S_1, S_2\}$; then the (measurable) information co-homology of degree one is generated by the

entropy, i.e., there exists a constant C such that, for any X in $W(\Sigma)$, $P \in \mathcal{P}$, $f(X; P) = CH(X; P)$.

Proof. We consider a 1-cocycle f. We have $f(1; P) = 0$. Let us note $f_i(P) = f(S_i; P)$, and $f_{ijk}(u)$ the function $f(S_i; P|(S_j = k))$, the variable u representing the probability of the first point in the fiber $S_j = k$ in the lexicographic order. For each tableau 2×2, $P = (p_{00}, p_{01}, p_{10}, p_{11})$, the symmetry formula (36) gives

$$(p_{00} + p_{10})f_{120}(\frac{p_{00}}{p_{00} + p_{10}}) + (p_{01} + p_{11})f_{121}(\frac{p_{01}}{p_{01} + p_{11}}) - f_1(P)$$

$$(45) \quad = \quad (p_{00} + p_{01})f_{210}(\frac{p_{00}}{p_{00} + p_{01}}) + (p_{10} + p_{11})f_{211}(\frac{p_{10}}{p_{10} + p_{11}}) - f_2(P)$$

imposing $p_{10} = 0$, $p_{00} = u$, $p_{11} = v$, $p_{01} = 1 - u - v$ in this relation, we obtain the equation:

$$(1 - u)f_1(0, \frac{1 - u - v}{1 - u}, 0, \frac{v}{1 - u}) - f_1(u, 1 - u - v, 0, v)$$

$$(46) \quad = \quad (1 - v)f_2(\frac{u}{1 - v}, \frac{1 - u - v}{1 - v}, 0, 0) - f_2(u, 1 - u - v, 0, v).$$

By hypothesis, f_1, f_2 depend only on the image law by S_1, S_2 respectively, thus, again by noting a binomial probability from the value of the first element in lexicographic order, we get

$$(47) \quad (1 - u)f_1(\frac{1 - u - v}{1 - u}) - f_1(1 - v) = (1 - v)f_2(\frac{u}{1 - v}) - f_2(u).$$

By equating u to $1 - v$, we find that $f_1(u) = f_2(u)$; then we arrive to the following functional equation for $h = f_1 = f_2$:

$$(48) \quad h(u) - h(v) = (1 - v)h(\frac{u}{1 - v}) - (1 - u)h(\frac{v}{1 - u})$$

This is the functional equation which was considered by Tverberg in 1958 [52]. As a result of the works of Tverberg [52], Kendall [30] and Lee (1964, [36]), (see also Kontsevich, 1995 [35]), it is known that every measurable solution of this equation is a multiple of the entropy function:

$$(49) \quad h(x) = C(x \log(x) + (1 - x) \log(1 - x)).$$

From here it follows that, for any m-uple $(x_1, ..., x_m)$ of real numbers such that $x_1 + ... + x_m = 1$,

$$(50) \quad F(x_1, x_2, ..., x_m) = C \sum_i x_i \log(x_i).$$

The same is true for H and G with the appropriate number of variables.

A pair of variables X, Y, such that $X, Y, (XY)$ belong to $\mathcal{S}$, is called an edge of $\mathcal{S}$; we says this edge is *rich* if X and Y contain at least two elements and (X, Y) at least four elements which cross the elements of X and Y, in such a manner that the Lemma 2 applies if $\mathcal{Q}$ is complete. We say that $\mathcal{S}$ is *connected*, if every pair of elements X, X' in $\mathcal{S}$ can be joined by a sequence of edges. We say that $\mathcal{S}$ is *sufficiently rich* if each vertex belongs to at least one rich edge. By the the recurrence Equation (100), these two conditions guaranty that the constant C which appears in the Lemma 2 is the same for all rich edges. Then the same recurrence Equation (100) implies that the whole co-cycle is equal to CH. If $\mathcal{S}$ has m connected components, we get necessarily m independent constants.

Thus we have established the following result:

Theorem 1. For every connected structure of information $\mathcal{S}$, which is sufficiently rich, and every set of probability $\mathcal{Q}$, which is complete with respect to $\mathcal{S}$, the information co-homology group of degree one is one-dimensional and generated by the classical entropy.

The theorem applies to rich simplicial complexes, in particular to the full simplex $\mathcal{S} = W(\Sigma)$, which is generated by a family Σ of partitions $S_1, ..., S_n$, when $n \geq 2$, such that, for every i at least of the pairs (S_i, S_j) is rich.

Note that most of the axiomatic characterizations of entropy have used convexity, and recurrence over the dimension, see Khintchin [31], Baez *et al.* [3].

In our characterization, we assumed no symmetry hypothesis, this was a consequence of co-homology. Moreover, we do not assume any stability property relating to a higher dimensional simplex, this was also a consequence of the homological definition.

There exists a notion of symmetric information co-homology:

The group of permutations $\mathfrak{S}(\Omega, \mathcal{B})$, made by the permutations of Ω that respect the algebra $\mathcal{B}$, acts naturally on the set of partitions $\Pi(\Omega)$; in fact, if $X \in \Pi(\Omega)$ is made by the subsets $\Omega_1, ..., \Omega_k$, the partition $\sigma^* X$ is made by the subsets $\sigma^{-1}(\Omega_1), ..., \sigma^{-1}(\Omega_1)$, in such a manner that, if σ, τ are two permutations of Ω, we have $\tau^*(\sigma^* X) = (\sigma \circ \tau)^* X$.

We say that a classical information structure $\mathcal{S}$ on $(\Omega, \mathcal{B})$ is *symmetric* if it is closed by the action of the group of permutations $\mathfrak{S}(\Omega, \mathcal{B})$, *i.e.*, if $X \in \mathcal{S}$, and $\sigma \in \mathfrak{S}(\Omega)$, the partition $\sigma^* X$ also belongs to $\mathcal{S}$.

In the same way, we say that a probability functor $\mathcal{Q}$ is *symmetric*, if it is stable under local permutations, *i.e.*, if $X \in \mathcal{S}$ and $P \in \mathcal{Q}_X$, and if $\sigma \in \mathfrak{S}(\Omega/X)$, then the probability law $\sigma^* P = P \circ \sigma$ on Ω/X also belongs to $\mathcal{Q}_X$.

Remark that we also have $\tau^* \sigma^* P = (\sigma \circ \tau)^* P)$. Thus the actions of symmetric groups are defined here on the right. However, we have actions to the left by taking $\sigma_* = (\sigma^{-1})^*$. For the essential role of symmetries in information theory, see the article of Gromov in this volume.

A m-cochain $F_X : \mathcal{S}^m \times \mathcal{Q}_X \to \mathbb{R}$ is said symmetric, when, for every $X \in \mathcal{S}$, every probability $P \in \mathcal{Q}_X$, every collection of partitions $Y_1, ..., Y_m$ in $\mathcal{S}_X$, we have

$$(51) \qquad F_{\sigma_* X}(\sigma_* Y_1; ...; \sigma_* Y_m; \sigma_* P) = F_X(Y_1; ...; Y_m; P).$$

It is evident that symmetric cochains form a subcomplex of the information cochains complex; *i.e.*, the coboundary of a symmetric cochain being a symmetric cochain. Consequently we get a symmetric information co-homology, that we name $H^*_{\mathfrak{S}}(\mathcal{S}; \mathcal{Q})$.

In particular the entropy is a symmetric 1-cocycle.

The above proof of Theorem 1 applies to symmetric cocycle as well, thus, under the convenient hypothesis of connexity, richness, and completeness for $\mathcal{S}$ and $\mathcal{Q}$ we have $H^1_{\mathfrak{S}}(\mathcal{S}; \mathcal{Q}) = \mathbb{R}H$.

Remark that an equivalent way to look at symmetric information cochains, consists in enlarging the category $\mathcal{S}$ in a "symmetric category" $\mathcal{S}^{\mathfrak{S}}$, by putting an arrow associated to each element $\sigma_X \in \mathfrak{S}(\Omega/X)$ from X to $\sigma_* X$, and completing the category by composing the two kind of arrows, division and permutation. In

this case, the probability functor $\mathcal{Q}$ must behave naturally with respect to permutation, which implies it is symmetric. Moreover, the natural notion of functional sheaf and local cochains are a symmetric sheaf and symmetric cochains.

2.4. **Appendix. Complex of Possible Events.**

In each concrete situation, physical constraints produce exclusion rules between possible events, which select a sub-complex $\mathcal{Q}$ in the full probability simplex $\mathcal{P} = \Delta_N$ on Ω. The aim of this appendix is to make this remark more precise.

Let $A^0, A^1, A^2, A^3, ...$ the $N + 1$ vertices of the large simplex Δ_N, a point of Δ_N is interpreted as a probability $\mathbb{P}$ on the set of thee vertices; each vertex can be seen as an elementary event, and we will say that a general event A is *possible* for $\mathbb{P}$ when $\mathbb{P}(A)$ is different from zero. An event A is said *impossible* for $\mathbb{P}$ in the other case, that is when $\mathbb{P}(A) = 0$.

The star $S(A)$ of a vertex A of Δ_N is the complementary set of the opposite face to A, *i.e.*, it is the set of probabilities P in Δ_N such that A is possible, *i.e.*, has non-zero probability. The relative star $S(A|K)$ of A in subcomplex K is the intersection of the star of A with K.

We denote $F = (A, B, C, D, ...)$ the face of Δ_N whose vertices are $A, B, C, D, ...$ We note $L(F)$ the set of points p in Δ_N such that at least one of the points $A, B, C, D, ...$ is impossible for p. This is also the reunion of the faces which are opposite to the vertices $A, B, C, D, ...$. Then $L(F)$ is a simplicial complex. The complementary set in F of the interior of F, *i.e.*, the boundary of F, is the reunion of the intersections of F with all faces opposite to $A, B, C, D, ...$; it is also the set of probabilities p in F such that at least one of the points $A, B, C, D, ...$ is impossible for p, thus it is equal to $L(F) \cap F$. If G is a face containing F the complex $L(G)$ contains the complex $L(F)$.

Let K be a simplicial complex contained in a N-simplex; then K is obtained by deleting from Δ_N a set $E = E_K$ of open faces. Let $\dot{F} = F \backslash \partial F$ be an element of E, then each faces G of Δ_N containing F belongs to E, because K is a complex.

In this case K is contained in $L(F)$. In fact $L(F)$ is the smallest sub-complex of Δ_N which does not contain $\dot{F}$. This can be proved as follows: if p in K makes that every vertices of F is possible, it belongs to a face G such that every vertex of F is a vertex of G, thus K contains G which contains F. So, if K does not contain $\dot{F}$, K is contained in $L(F)$.

Let $L = L_K$ be the intersection of the $L(F)$, where F describe the faces in E_K. From what precedes we know that K is contained in L. However, every $\dot{F}$ in E is included in the complementary set of $L(F)$, thus it is included in the complementary set of L, which is the union of the complementary sets of the $L(F)$. Consequently the complementary set of K is included in the complementary set of L. Then $K = L$.

This discussion establishes the following result:

Theorem 2. A subset K of the simplex Δ_N is a simplicial sub-complex if and only if it is defined by a finite number of constraints of the type: "for any p in K, the fact that $A, B, C, ...$ are possible for p implies that D is impossible for p".

In other terms, more imaged but also more ambiguous, every sub-complex K is defined by constraints of the type: "if $A, B, C, ...$ are simultaneously allowed it is excluded that D can happen".

The statement of the theorem is just a rewriting of the discussion, using elementary propositional calculus: let K be a sub-complex of Δ_N, we have shown that K is the intersection of the $L(F)$ where the open face $\dot{F}$ is not in K, but if $A, B, C, D, ...$ denote the vertices of the face F, a point p belongs to $L(F)$ if and only if "(A is impossible for p) or (B is impossible for p) or ...", and this sentence is equivalent to "if (A is possible for p) and (B is possible for p) and ..., then (D is impossible for p)". This results from the equivalence between "(P implies Q) is true" and "(no P or Q) is true". Reciprocally any $L(F)$ is a simplicial complex, then every intersection of sets of the form $L(F)$ is a simplicial complex too.

3. Higher Mutual Informations. A Sketch

The topological co-boundary operator on $\mathcal{C}^*$, denoted by δ_t, is defined by the same formula as δ, except that the first term $Y_1.F(Y_2; ...; Y_n; \mathbb{P})$ is replaced by the term $F(Y_2; ...; Y_n; \mathbb{P})$ without Y_1:

(52)
$$\delta_t^m F(Y_1; ...; Y_{m+1}; P_X)$$
$$= F(Y_2; ...; Y_{m+1}; P_X) + \sum_1^m (-1)^i F(...; (Y_i, Y_{i+1}); ...; Y_{m+1}; P_X) + (-1)^{m+1} F(Y_1; ...; Y_m; P_X).$$

It is the coboundary of the bar complex for the trivial module $\mathcal{F}_t$, which is the same as $\mathcal{F}$ except no conditioning appears, *i.e.*, $Y.F = F$. Hence it is the ordinary simplicial co-homology of the complex $\mathcal{S}$ with local coefficients in $\mathcal{F}$.

Remark that this operator also preserves locality, because all the functions of $\mathbb{P}$ which comes in the development depends only on $(Y_2, ..., Y_n) * \mathbb{P}$, $(Y_1, ..., Y_n) * \mathbb{P}$ and $(Y_1, ..., Y_{n-1}) * \mathbb{P}$.

By definition a topological cocycle of information is a cochain F that satisfies $\delta_t F = 0$, and a topological co-boundary is an element in the image of δ_t.

It is easy to show that $\delta_t \circ \delta_t = 0$, which allows to define a co-homology theory that we will name topological co-homology.

Now assume that the information structure $\mathcal{S}$ is a set $W(\Sigma) = \Delta(n)$ generated by a family Σ of partitions $S_1, ..., S_n$, when $n \geq 2$.

Higher mutual information quantities were defined by Hu Kuo Ting [27] (see also Yeung [55]), generalizing the Shannon mutual information.

(53)
$$I_N(S_1; ...; S_N; \mathbb{P}) = \sum_{k=1}^{k=N} (-1)^{k-1} H_k(S_1; ...; S_N; \mathbb{P}),$$

where

(54)
$$H_k(S_1; ...; S_N; \mathbb{P}) = \sum_{I \subset [N]; card(I)=k} H(S_I; \mathbb{P}),$$

S_I denoting the joint partition of the S_i such that $i \in I$. We also define $I_1 = H$.

The definition of I_N makes evident it is a symmetric function, invariant by all permutation of the partitions $S_1, ..., S_N$.

For instance $I_2(S; T) = H(S) + H(T) - H(S, T)$ is the usual mutual information.

It is easily seen that $I_2 = \delta_t H$. The following formula generalizes this remark to higher mutual informations of even orders:

(55)
$$I_{2m} = \delta_t \delta \delta_t ... \delta \delta_t H,$$

where the right member contains $2m - 1$ terms.

And for odd mutual information we have

$$(56) \qquad\qquad I_{2m+1} = -\delta\delta_t\delta\delta_t...\delta\delta_t H,$$

where the right member contains $2m$ terms.

We deduce from here that higher mutual informations are co-boundaries for δ or δ_t according that their order is odd or even respectively.

The result which proves the two above formulas is the following:

Lemma 3. Let n be even or odd we have

$$(57) \quad I_N((S_0, S_1); S_2; ...; S_N; \mathbb{P}) = I_N(S_0; S_2; ...; S_N; \mathbb{P}) + S_0.I_N(S_1; S_2; ...; S_N; \mathbb{P})$$

This lemma can be proved by comparing the completely developed forms of the quantities. It seems to signify that, with respect to one variable, I_N satisfies the equation of information 1-cocycle, thus I_N seems to be a kind of "partial 1-cocycle"; however this is misleading, because the locality condition is not satisfied. In fact I_N is a N-cocycle, either for δ, either for δ_t depending on the parity of N.

For any N-cochain F we have

$$(58) \qquad (\delta - \delta_t)F(S_0; S_1; ...; S_N; \mathbb{P}) = ((S_0 - 1).F)(S_1; ...; S_N; P),$$

where $S_0 - 1$ denotes the sum of the two operators of mean conditioning and minus identity.

That implies:

$$(59) \quad (\delta\delta_t - \delta_t\delta)F(S_0; S_1; S_2; ...; S_N; \mathbb{P}) = ((1 + S_0 + S_1 - S_0 S_1).F)(S_2; ...; S_N; \mathbb{P}),$$

Remark 3. Reciprocally the functions I_N decompose the entropy of the finest joint partition:

$$(60) \qquad H(S_1, S_2, ..., S_N; \mathbb{P}) = \sum_{k=1}^{k=N} (-1)^{k-1} \sum_{I \subset [N]; card(I)=k} I_k(S_{i_1}; S_{i_2}; ...; S_{i_k}; \mathbb{P})$$

For example, we have $H(S, T) = I_1(S) + I_1(T) - I_2(S; T)$, and
$$(61)$$
$$H(S, T, U) = H(S) + H(T) + H(U) - I_2(S; T) - I_2(T; U) - I_2(S; U) + I_3(S; T; U).$$

Let us also note the recurrence formula whose proof is left to the reader (*cf.* Cover and Thomas [17]):

$$(62) \qquad I_{N+1}(S_0; S_1; ...; S_N) = I_N(S_1; ...; S_N) - S_0.I(S_1; ...; S_N).$$

4. Quantum Information and Projective Geometry

4.1. Quantum Measure, Geometry of Abelian Conditioning.
In finite dimensional quantum mechanics the role of the finite set Ω of atomic events is played by a complex vector space E of finite dimension.

In fact, to each set Ω, of cardinal N, is naturally associated a vector space of dimension N over $\mathbb{C}$, which is the space freely generated over $\mathbb{C}$ by the elements of Ω. Then we can identify E with $\mathbb{C}^N$, the canonical basis being the points x of Ω. In this case the canonical positive hermitian metric on E corresponds to the quadratic mean: if f and g are elements of E, we have

$$(63) \qquad\qquad h_0(f, g) = \langle f|g \rangle_0 = \int \bar{f}(\omega)g(\omega)d\omega = \frac{1}{N}\sum_j \overline{f_j}g_j$$

Remark that, in the infinite dimensional situation, the space which would play the role of E is the space of L^2 functions for a fixed probability P_0.

Probability laws $\mathbb{P}$, which are elements of the big simplex $\Delta(N)$, give other hermitian structures, the ones which are expressed by diagonal matrices, with positive coefficients, and trace equal to 1.

In the general quantum case, described by E, a quantum probability law is every positive non-zero hermitian product h. If a basis is chosen, h is described by an $N \times N$-matrix ρ. In the physical literature, every such ρ is called a *density of states*; and it is considered as a full description of the physical states of the finite quantum system. Usually ρ is normalized by $Tr(\rho) = 1$.

Note that this condition on the trace has no meaning for a positive hermitian form h if no additional structure is given, for instance a non-degenerate form h_0 of reference. Why is it so? Because *a priori* a hermitian form h on E is a map from E to $\overline{E}^*$, where $*$ denotes duality and *bar* denotes conjugation, the conjugate space $\overline{E}$ being the same set E, with the same structure of vector space over the real numbers as E, but with structure of vector space over the complex numbers changed by changing the sign of the action of the imaginary unit i. The complexification of the real vector space H of hermitian forms is $Hom_{\mathbb{C}}(E, \overline{E}^*) \cong E^* \otimes \overline{E}^*$. The space H is the set of fixed points of the $\mathbb{C}$-anti-linear map $u \mapsto {}^t \bar{u}$. A trace is defined for an endomorphism of the space E, as a linear invariant quantity on $E^* \otimes E$. Here we could take the trace over $\mathbb{R}$, because E and $\overline{E}$ are the same over $\mathbb{R}$, but the duality would be an obstacle, because even over the field $\mathbb{R}$, the spaces E and E^* cannot be identified, and there exits no linear invariant in $E^* \otimes E^*$, even over $\mathbb{R}$. In fact, a non-degenerate positive h_0 is one of the way to identify E and $\overline{E}^*$. A basis is another way, also defining canonically a form h_0. More precisely, when h_0 is given, every hermitian form h diagonalizes in an orthonormal basis for h_0, thus all the spectrum of h makes sense not only the trace.

This h_0 is tacitly assumed in most presentations. However it is better to understand the consequences of this choice. In non-relativistic quantum mechanics, it is not too grave, however in relativist quantum mechanics, it is; for instance, considering the system of two states as a spinor on the Lorentz space of dimension 4, the choice of h_0 is equivalent to the choice of a coordinate of time. See Penrose and Rindler [48].

A much less violent way to do is to consider hermitian structures h up to multiplication by a strictly positive number. This would have the same effect as fixing the trace equals to one, without introducing any choice. In quantum mechanics only non-zero positive h are considered, not necessarily positive definite, but non-zero. This indicates that a good space of states is not the set H_+ of all positive non-zero hermitian products but a convex part PH_+ of the real projective space of real lines in the vector space H of hermitian forms. In this space, the complex projective space $\mathbb{P}(E)$ of dimension $N - 1$ over $\mathbb{C}$ is naturally embedded, its image consists of the rank one positive hermitian matrices of trace 1; these matrices correspond to the orthogonal projectors on one dimensional directions in E.

When a basis of E is chosen, particular elements of $\mathbb{P}(E)$ are given by the generators of $\mathbb{C}^N$; they correspond to the Dirac distributions on classical states. We see here a point defended in particular by Von Neumann, that quantum states are projective objects not linear objects.

The classical random variables, *i.e.*, the measurable functions on Ω with values in $\mathbb{C}$, are generalized in Quantum Mechanics by the operators in E, they are all the endomorphisms, *i.e.*, any $N \times N$-matrix, and they are named *observables*. Classical observables are recovered by diagonal matrices, their action on E corresponding to the multiplication of functions. Real valued variables are generalized by hermitian operators. Again this supposes that a special probability law h_0 is given. If not "to be hermitian" for an operator has no meaning. (What could have a meaning for an operator is to be diagonalizable over $\mathbb{R}$, which is something else.)

Then if h_0 is chosen, the only difference between real observable and density of states is the absence of the positivity constraint.

By definition, the *amplitude*, or *expectation*, of the observable Z in the state ρ is the number given by the formula

$$(64) \qquad\qquad \mathbb{E}_\rho(Z) = Tr(Z\rho).$$

It is important to note that h_0 plays a role in this formula. Consequently the definition of expectation requires to fix an h_0 not only a ρ. This imposes a departure from the relativistic case, which shall not be surprising, since considerations in relativistic statistical physics show that the entropy, for instance, depends on the choice of a coordinate for time. Cf. Landau-Lifschitz, Fluid Mechanics, second edition [**?**].

The partitions of Ω associated to random variables are replaced in the quantum context by the spectral decompositions of the hermitian operators X. As h_0 is given, this decomposition is given by a set of positive hermitian commuting projectors of sum equal to the identity. The additional data for recovering the operator X is one real eigenvalue for each projector. The underlying fact from linear algebra is that every hermitian matrix is diagonalizable in a unitary basis, which means that

$$(65) \qquad\qquad Z = \sum_j z_j E_j,$$

where the number z_j are real, two by two different, and where the matrices E_j are hermitian projectors, which satisfy, for any j and $k \neq j$,

$$(66) \qquad\qquad E_j^2 = E_j; \quad E_j^* = E_j; \quad E_j E_k = E_k E_j = 0;$$

and

$$(67) \qquad\qquad \sum_j E_j = Id_N$$

When the hermitian operator Z commutes with the canonical projectors on the axis of $\mathbb{C}^N$, its spectral measure gives an ordinary partition of the canonical basis, and we recover the classical situation.

Note that the extension of the notion of partition is given by any decomposition of the vector space E in orthogonal sum, not necessarily compatible with a chosen basis. Again this assumes a given positive definite h_0.

To generalize what we presented in the classical setting, quantum information theory must use only the spectral support of the decomposition, not the eigenvalues.

It would have been tempting to consider any decomposition of E in direct sum as a possible observable, however not every linear operator, or projective transformation, corresponds to such a decomposition, due to the existence of non-trivial nilpotent operators. What could be their role in quantum information? Moreover, the presence of h_0 fully justifies the limitation to orthogonal decompositions.

In the general case, hermitian but not necessarily diagonal, we define the probability of the elementary events $Z = z_j$ by the following formula

$$
(68) \qquad \mathbb{P}_\rho(Z = z_j) = Tr(E_j^* \rho E_j)
$$

And we define the conditional probability $\rho|(Z = z_j)$ by the formula

$$
(69) \qquad \rho|(Z = z_j) = E_j^* \rho E_j / Tr(E_j^* \rho E_j).
$$

One can notice that this definition can be extended to any projector, not necessarily hermitian. By definition, the conditioning of ρ by a projector Y is the matrix $Y^* \rho Y$, normalized to be of trace 1. However, here, as it is done in most of the texts on Quantum Mechanics, we will mostly restrict ourselves to the case of hermitian projectors, $i.e.$, $Y^* = Y$.

Remark 4. What justifies these definitions of probability and conditioning? First they allow to recover the classical notions when we restrict to diagonal densities and diagonal observables, $i.e.$, when ρ is diagonal, real, positive, of trace 1, Z is diagonal, and the E_j are diagonals, in which case they give a partition of Ω. The mean of Z is its amplitude. The probability of the event $Z = z_j$ is the sum of the probabilities $p(\omega) = \rho_{\omega\omega}$ for ω in the image of E_j; this the trace of ρE_j. Moreover, the conditioning by this event is the probability obtained by projection on this image, as prescribed by the above formula.

Second, pure states are defined as rank one hermitian matrices. In this case ρ is the orthogonal projection on a vector ψ of norm equal to 1 (the finite dimensional version of the Schrodinger wave vector), the exact relation is

$$
(70) \qquad \rho = |\psi\rangle\langle\psi|
$$

or, in coordinates, if ψ has for coordinates the imaginary numbers $\psi(\omega)$, we have

$$
(71) \qquad \rho_{\omega\omega'} = \overline{\psi(\omega)}\psi(\omega').
$$

Let Z be any hermitian operator, the result of quantum experiments indicate that the probability of the event $Z = z_j$, for the state ψ, is equal to

$$
(72) \qquad P_j = \langle\psi|E_j\psi\rangle.
$$

But this quantity can also be written

$$
(73) \qquad P_j = Tr_\mathbb{C}(\langle\psi|E_j\psi\rangle) = Tr_E(|\psi\rangle\langle\psi|E_j) = Tr(\rho E_j).
$$

Starting from this formula and the fact any ρ can be written as a classical mixture of commuting quantum pure states,

$$
(74) \qquad \rho = \sum_a p_a|\psi_a\rangle\langle\psi_a|,
$$

we get the general formula of a quantum probability that we recalled.

Moreover, physical experiments indicate that after the measurement of an observable Z, giving the quantity z_j, the system is reduced to the space E_j, and every pure state ψ is reduced to its projection $E_j\psi$, which is compatible with the above definition of conditioning for pure states. Here again, the general formula can be deduced by Equation (74). The division by the probability is achieved to normalize to a trace 1. Thus conditioning in general is given by orthogonal projection in E, and it corresponds to the operation of measurement.

However, as claimed in particular by Roger Balian [4], the fact that the decomposition in pure states is non-unique implies that pure states cannot be so pertinent for understanding quantum information.

Definition 3. The density of states associated to a given variable Z and a given density ρ is given by the sum:

$$(75) \qquad \rho_Z = \sum_j \mathbb{P}_\rho(Z = z_j)\rho|(Z = z_j) = \sum_j E_j^* \rho E_j,$$

where $(E_j)_{j \in J}$ designates the spectral decomposition of Z, also named spectral measure of Z. Thus ρ_Z is usually seen as representing the density of states after the measurement of the variable Z. This formula is usually interpreted by saying that the statistical analysis of the repeated measurements of the observable Z transforms the density ρ into the density ρ_Z.

Remark that ρ_Z is better understood as being a collection of conditional probabilities $\rho|(Z = z_j)$, indexed by j.

In quantum physics as in classical physics the symmetries, discrete and continuous, have always played a fundamental role. For example, in quantum mechanics, a fundamental principle is the unitarity of the evolution in time, which claims that the states evolve as $\rho_t = U_t\rho$ and that the observables evolve as $Z_t = U_t Z U_t^{-1}$, with U_t respecting the fundamental scalar product h_0. In fact, as we already mentioned, a deeper principle associates the choice of a time coordinate t to the choice of h_0, which gives birth to a unitary group $U(E; h_0)$, isomorphic to $U_N(\mathbb{C})$. For stationary systems the family $(U_t)_{t \in \mathbb{R}}$ forms a one parameter group, *i.e.*, $U_{t+s} = U_t U_s = U_s U_t$, and there exists a hermitian generator H of U_t in the sense that $U_t = \exp(2\pi i t H/h)$; by definition, this particular observable H is the energy, the most important observable. Even if we have a privileged basis, like Ω in the relation with classical probability, the consideration of another basis which makes the energy H diagonal is of great importance. In the stationary case, a *symmetry* of the dynamical system is defined as any unitary operator, which commutes with the energy H. The set of symmetries forms a Lie group G, a closed sub-group in U_N. The infinitesimal generators are considered as hermitian observables (obtained by multiplying the elements of the Lie algebra $L(G)$ by i); in general they do not commute between themselves.

All these axioms extend to the infinite dimensional situation when E has a structure of an Hilbert space, but the spectral analysis of the un-bounded operators is more delicate and diverse than the analysis in finite dimension. Three kinds of spectrum appear, discrete, absolutely continuous and singular continuous. The symmetries could not form a Lie group in general, and so on.

In our simple case of elementary quantum probability, without fixed dynamics, the classical symmetries of the set of probabilities are given by the permutations of Ω, the vertices of $\Delta(N)$. They correspond to the unitary matrices which have one and only one non-zero element in each line and each column. They do not diagonalize in the same basis because they do not commute, but they form a group $\mathfrak{S}_N$. Another subgroup of U_N is natural for semi-classical study, it is the diagonal torus $\mathbb{T}^N$, its elements are the diagonal matrices with elements of modulus 1, they correspond to sets of angles. The group $\mathfrak{S}_N$ normalizes the torus $\mathbb{T}^N$, *i.e.*, for each permutation σ and each diagonal element Z, the matrix $\sigma Z \sigma^{-1}$ is also diagonal; its

elements are the same as the elements of Z but in a different orders. The subgroup generated by $\mathfrak{S}_N$ and $\mathbb{T}^N$ is the full normalizer of $\mathbb{T}^N$.

One of the strengths of the quantum theory, with respect to the classical theory, is that it gives a similar status to the states, the observables and the symmetries. States are hermitian forms, generalizing points in the sphere (or in the projective space) which are pure states, observables are hermitian operators, or better spectral decompositions, and symmetries are unitary operators, infinitesimal symmetries being anti-hermitian matrices.

All classical groups should appear in this framework. First, by choosing a special structure on E we restrict the linear group $GL_N(\mathbb{C})$ to an algebraic subgroup $G_{\mathbb{C}}$. For instance, by choosing a symmetric invertible bilinear form on E we obtain $O_N(\mathbb{C})$, or, when N is even, by choosing an antisymmetric invertible bilinear form on E we obtain $Sp_N(\mathbb{C})$. In each of these cases there exists a special maximal torus (formed by the complexification of a maximal abelian subgroup T of unitary operators in $G_{\mathbb{C}}$), and a Weyl group, which is the quotient of the normalizer $N(T)$ by the torus T itself. This Weyl group generalizes the permutation group when more algebraic structures are given in addition to the linear structure. The compact group of symmetries is the intersection G of $G_{\mathbb{C}}$ with U_N. In fact, given any compact Lie group G_c, and any faithful representation r_c of G_c in $\mathbb{C}^N$, we can restrict real observables to generators of elements in C_c, and general observables to complex combinations of these generators, which integrate in a reductive linear group G. The spectral decomposition corresponds to the restriction to parabolic sub-groups of $G_{\mathbb{C}}$. The densities of states are restricted to the Satake compactification of the symmetric space $G_{\mathbb{C}}/G_c$ [10].

4.2. Quantum Information Structures and Density Functors.

To define information quantities in the quantum setting, we have *a priori* to consider families of operators $(Y_1, Y_2, ..., Y_m)$ as joint variables. However, the efforts made in Physics and Mathematics were not sufficient to attribute a clear probability to the joint events $(Y_1 = y_1, Y_2 = y_2, ..., Y_m = y_m)$, when $Y_1, ..., Y_m$ do not commute; we even suspect that this difficulty is revelator of a principle, that information requires a form of commutativity. Thus, in our study, we will adopt the convention that *every time we consider joint observables, they do commute*. Hence we will consider only collections of commuting hermitian observables; their natural amplitudes in a given state are vectors in $\mathbb{R}^m$. However we do not exclude the consideration in our theory of sequences $(Y_1; ...; Y_m)$ such that the Y_i do not commute.

A joint observable $(Y_1, Y_2, ..., Y_m)$ define a linear decomposition of the total space E in direct orthogonal sum

$$(76) \qquad E = \bigoplus_{\alpha \in A} E_\alpha,$$

where $E_\alpha; \alpha \in A$ is the collection of joint eigenspaces of the operators Y_j. Note that any orthogonal decomposition can be defined by a unique operator.

Another manner to handle the joint variables is to consider linear families of commuting operators

$$(77) \qquad Y(\lambda_1, ..., \lambda_m) = \lambda_1 Y_1 + ... + \lambda_m Y_m,$$

or in equivalent terms, linear maps from $\mathbb{R}^m$ to $End(E)$. Then assigning a probability number and perform probability conditioning can be seen as functorial operations.

In what follows we denote indifferently by E_α the subspace of E or the orthogonal projection on this subspace.

From the point of view of information, two sets of observables are equivalent if they give the same linear decomposition of E. We say that a decomposition $E_\alpha; \alpha \in A$ refines a decomposition $E'_\beta; \beta \in B$, when each E'_β is a sum of spaces E_α for α in a subset A_β of A. In such a case, we say that $E_\alpha; \alpha \in A$ *divides* $E'_\beta; \beta \in B$.

For instance, for commuting decompositions Y, Z it is possible to define the joint variable, as the less fine decomposition which is finer than Y and Z.

We insist that only decompositions have a role in information study at this moment. We will see that observation trees in the last section imposes to consider a supplementary structure, which consists in an *ordering* of the factors in the decomposition.

An *information structure* on E is a set $\mathbf{S}$ of decompositions X of E in direct sum, such that when Y and Z are elements of $\mathbf{S}$ which refine $X \in \mathbf{S}$, then Y, Z commute and the finer decomposition (Y, Z) they generate belongs to $\mathbf{S}$. In this text, we will only consider orthogonal decompositions.

Remark: in fact, the necessity of this condition in the quantum context was the original motivation to introduce the definition of classical information structure, as exposed in the first section. This can be seen as a comfortable flexibility in the classical context, or as a step from classical to quantum information theory.

As in the classical case, an information structure gives a category, denoted by the letter $\mathbf{S}$, whose objects are the elements of $\mathbf{S}$, and whose arrows $X \to Y$ are given by the divisions $X|Y$ between the decompositions in $\mathbf{S}$.

In what follows we always assume that 1, which corresponds to the trivial partition E, belongs to $\mathbf{S}$, and is a final object. If not we will not get a topos.

Note that we are not the first to use categories and topos to formulate quantum or classical probability. In particular Doring and Isham propose a reformulation of the whole quantum and classical physics by using topos theory, see [18] and references inside. This theory followed remarkable works of Isham, Butterfield and Hamilton, made beween 1998 and 2002, and was further developed by Flori, Heunen, Landsman, Spitters, specially in the direction of a quantum logic. A common point between these works and our work is the consideration of sheaves over the category made by the partial ordering in commutative subalgebras. However, Doring *et al.* consider only the set of maximal algebras, and do not look at decompositions, *i.e.*, they consider also the spectral values. In [18], Doring and Isham defined topos associated to quantum and classical probabilities. However, they focused on the definition of truth values in this context. For instance, in the classical setting, the topos they define is the topos of ordinary topological sheaves over the space $(0,1)_L$ which has for open sets the intervals $]0,r[$ for $0 \le r \le 1$, and particular points in their topos are given by arbitrary probabilized spaces, which is far from the objects we consider, because our classical topos are attached to sigma-algebras over a given set. In fact, our aim is more to develop a kind of geometry in this context, by using homological algebra, in the spirit of Artin, Grothendieck, Verdier, when they developed topos for studying the geometry of schemes.

Example 5. The most interesting structures $\mathbf{S}$ seem to be provided by the quantum generalization of the simplicial information structure in classical finite probability. A finite family of commuting decompositions $\Sigma = \{S_1, ..., S_n\}$ is given, they

diagonalize in a common orthogonal basis, but it can happen that not all diagonal decompositions associated to the maximal torus belongs to the set of joints $W(\Sigma)$. In such a case a subgroup G_Σ appears, which corresponds to the stabilizer of the finest decomposition $S_{[n]} = (S_1...S_n)$. This group is in general larger than a maximal torus of U_N, it is a product of unitary groups (corresponding to common eigenvalues of observables in $W(\Sigma)$), and it is named a Levy subgroup of the unitary group. In addition we consider a closed subgroup G in the group $U(E; h_0)$ (which could be identified with U_N), and all the conjugates gYg^{-1} of elements of $W(\Sigma)$ by elements of G; this gives a manifold of commutative observable families $\Sigma_g; g \in G$. More generally we could consider several families $\Sigma_\gamma; \gamma \in \Gamma$ of commuting observables, where Γ is any set. It can happen that an element of Σ_γ is also an element of Σ_λ for $\lambda \neq \gamma$. The family $\Gamma * \Sigma$ of the Σ_γ when γ describes the set Γ forms a quantum information structure. The elements of this structure are (perhaps ambiguously) parameterized by the product of an abstract simplex $\Delta(n)$ with the set Γ (in particular $\Gamma = G$ for conjugated families).

A simplicial information structure is a subset of $\Gamma * \Sigma$ which corresponds to a family K_γ of simplicial sub-complexes of $\Delta(n)$. In the invariant case, when $\Gamma = G$, several restrictions could be usefull, for instance using the structure of the manifold of the conjugation classes of G_Σ under G. The simplest case is given by taking the same complex K for all conjugates $g\Sigma g^{-1}$. By definition this latter case is a simplicial invariant family of quantum observables.

An *event* associated to $\mathbf{S}$ is a subspace E_A, which is an element of one of the decompositions $X \in \mathbf{S}$. For instance, if $Y = (Y_1, ..., Y_m)$, the joint event $A = (Y_1 = y_1, Y_2 = y_2, ..., Y_m = y_m)$ gives the space E_A which is the maximal vector subspace of E where A happens, *i.e.*,

$$(78) \qquad (f \in E_A) \Leftrightarrow (Y_1(f) = y_1 f, Y_2(f) = y_2 f, ..., Y_m(f) = y_m f).$$

We say that A is measurable for a decomposition Y whenever it is obtained by unions of elements of Y.

The role of the Boolean algebra $\mathcal{B}$ introduced in the first section, could have been accounted here by a given decomposition $\mathbf{B}$ of E such that any decomposition in $\mathbf{S}$ is divided by $\mathbf{B}$.

However this choice of $\mathbf{B}$ is too rigid, in particular it forbids invariance by the unitary group $U(h_0)$. Thus we decided that a better analog of the Boolean algebra $\mathcal{B}$ is the set $U\mathbf{B}$ of all decompositions that are deduced from a given $\mathbf{B}$ by unitary transformations.

On the side of density of states, *i.e.*, quantum probabilities, we can consider a subspace $\mathbf{Q}_1$ of the space $\mathbf{P} = \mathbb{P}\mathbf{H}_+$ of hermitian positive matrices modulo multiplication by a constant. Concretely, we identify the elements of $\mathbf{Q}_1$ with positive hermitian operators ρ such that $Tr\rho = 1$. The space $\mathbf{P}$ is naturally stratified by the rank of the form; the largest cell $\mathbb{P}\mathbf{H}_{++}$ corresponds to the non-degenerate forms; the smallest cells correspond to the rank one forms, which are called pure states in Quantum Mechanics.

We will only consider subsets $\mathbf{Q}_1$ of $\mathbf{P}$ which are *adapted* to $\mathbf{S}$, *i.e.*, which satisfy that if ρ belongs to $\mathbf{Q}_1$, the conditioning of ρ by elements of $\mathbf{S}$ also belongs to $\mathbf{Q}_1$. This means that $\mathbf{Q}_1$ is closed by orthogonal projections on all the elements E_A of the orthogonal decompositions X belonging to $\mathbf{S}$. Note that a subset of $\mathbf{P}$ which

is closed by all orthogonal projections is automatically adapted to any information category $\mathbf{S}$.

Remind that, if ρ is a density of states and E_A is an elementary event (*i.e.*, a subspace of E), we define the conditioning of ρ by A by the hermitian matrix

(79) $$\rho|A = E_A^* \rho E_A / Tr(E_A^* \rho E_A).$$

And we define the *probability of the event E_A* for ρ as the trace:

(80) $$\mathbb{P}_\rho(A) = Tr(E_A^* \rho E_A),$$

In the same manner we define the density of a joint observable by

(81) $$\rho_Y = \sum_A \mathbb{P}_\rho(A)\rho|A = \sum_A E_A^* \rho E_A,$$

A nice reference studying important examples is Paul-Andre Meyer, Quantum probability for probabilists [45].

If X is an orthogonal decomposition of E, we can associate to it a subset $\mathbf{Q}_X$ of $\mathbf{Q}_1$, which contains at least all the forms ρ_X where ρ belongs to $\mathbf{Q}_1$. The natural axiom that we assume for the function $X \mapsto \mathbf{Q}_X$, is that for each arrow of division $X \to Y$, the set $\mathbf{Q}_Y$ contains the set $\mathbf{Q}_X$; then we note Y_* the injection from $\mathbf{Q}_X$ to $\mathbf{Q}_Y$. The fact that $\mathbf{Q}_X$ is stable by conditioning by every element of a decomposition Y which is less fine than X is automatic; it follows from the fact that $\mathbf{Q}_1$ is adapted to $\mathbf{S}$. We will use conditioning in this way.

In what follows we denote by the letter $\mathbf{Q}$ such a functor $X \mapsto \mathbf{Q}_X$ from the category $\mathbf{S}$ to the category of quantum probabilities, with the arrows given by direct images. The set $\mathbf{Q}_1$ is the value of the functor $\mathbf{Q}$ for the certitude 1. We must remind that many choices are possible for the functor when $\mathbf{Q}_1$ is given; the two extreme being the functor $\mathbf{Q}^{max}$ where $\mathbf{Q}_X = \mathbf{Q}_1$ for every X, and the functor $\mathbf{Q}^{min}$ where $\mathbf{Q}_X$ is restricted to the set of forms ρ_X where ρ describes $\mathbf{Q}_1$; in this last case the elements of $\mathbf{Q}_X$ are positive hermitian forms on E, which are decomposed in blocs according to X.

From the physical point of view, $\mathbf{Q}^{min}$ appears to have more sense than $\mathbf{Q}^{max}$, but we prefer to consider both of them.

A special probability functor, which will be noted $\mathbf{Q}^{can}(\mathbf{S})$, is canonically associated to a quantum information structure $\mathbf{S}$:

Definition 4. The *canonical density functor* $\mathbf{Q}_X^{can}(\mathbf{S})$ is made by all positive hermitian forms matched to X, i.e., all the forms ρ_X when ρ describes $P\mathbf{H}_+$.

It is equal to the functor $\mathbf{Q}^{min}$ associated to the full set $\mathbf{Q}_1 = P\mathbf{H}_+$. When the context is clear, we will simply write $\mathbf{Q}^{can}$.

An important difference appears between the quantum and the classical frameworks: if X divides Y, there exist more (quantum) probability laws in $\mathbf{Q}_Y$ than in $\mathbf{Q}_X$, but there exist less classical laws at the place Y than at the place X, because classical laws are defined on smaller sigma-algebras.

In particular, the trivial partition has only one classical state, which is $Tr(\rho) = 1$, but it has the richest structure in terms of quantum laws, any hermitian positive form.

Let us consider the classical probabilities, *i.e.*, the maps that associate the number $P_\rho(A)$ to an event A; then, for an event which is measurable for Y, the law $Y_* \rho_X$ gives the same result than the law ρ_X.

Remark: This points to a generalized notion of direct image, which is a correspondence $q_X Y_*$ between $\mathbf{Q}_X$ and $\mathbf{Q}_Y$, not a map: we say that the pair (ρ_X, ρ_Y) in $\mathbf{Q}_X \times \mathbf{Q}_Y$ belongs to $q_X Y_*$, if for any event which is measurable for Y, we have the equality of probabilities

(82) $$\mathbb{P}_{\rho_X}(A) = \mathbb{P}_{\rho_Y}(A).$$

Let us look at the relation of quantification, between a classical information structure and a quantum one:

Consider a maximal family of commuting observables $\mathcal{S}$ in the quantum information structure $\mathbf{S}$, *i.e.*, the full subcategory associated to an initial object X_0. This family is a classical information structure. Conversely, if we start with a classical information structure $\mathcal{S}$, made by partitions of a finite set Ω, we can always consider it as a quantum structure associated to the vector space $E = \mathbb{C}^{\Omega}$ freely generated over $\mathbb{C}$ by the elements of Ω. Note that E comes with a canonical positive definite form h_0, and, to be interesting from the quantum point of view, it is better to extend $\mathcal{S}$ by applying to it all unitary transformations of E, generating a quantum structure $\mathbf{S} = U\mathcal{S}$.

Remark 5. Suppose that $\mathbf{S}$ is unitary invariant, we can define a larger category $\mathbf{S}^U$ by taking as arrows the isomorphisms of ordered decomposition, and close by all compositions of arrows of $\mathbf{S}$ with them. Such an invariant extended category $\mathbf{S}^U$ is not far to be equivalent to the category $\mathcal{S}^{\mathfrak{S}}$, made by adding arrows for permutations of the sets Ω/X (cf. above section), from the point of view of category theory: let us work an instant, as we will do in the last part of this paper, with ordered partitions of Ω, being itself equipped with an order, and ordered orthogonal decompositions of E. In this case we can associate to any ordered partition $X = (E_1, ..., E_m)$ of E, the unique ordered partition Ω compatible with the sequence of dimensions and the order of Ω. It gives a functor τ from $\mathbf{S}$ to $\mathcal{S}$ such that $\iota \circ \tau = Id_{\mathcal{S}}$, where ι denotes the inclusion of $\mathcal{S}$ in $\mathbf{S}$. These two functors are extended, preserving this property, to the categories $\mathbf{S}^U$ and $\mathcal{S}^{\mathfrak{S}}$. In fact, the functor ι sends a permutation to the unitary map which acts by this permutation on the canonical basis, and the functor τ sends a unitary transformation g between $X \in \mathbf{S}$ and $gXg^* \in \mathbf{S}$ to the permutation it induces on the orthogonal decompositions. Moreover, consider the map f which associates to any $X \in \mathbf{S}^U$ the unique morphism from the decomposition $\iota \circ \tau(X)$ to X; it is a natural transformation from the functor $\iota \circ \tau$ to the functor $Id_{\mathbf{S}^U}$, which is invertible, then it defines an equivalence of category between $\mathcal{S}^{\mathfrak{S}}$ and $\mathbf{S}^U$. However a big difference begins with probability functors.

Let $\mathbf{Q}$ be a quantum density functor adapted to $\mathbf{S}$, and note $\iota^* \mathbf{Q}$ the composite functor on $\mathcal{S}$; we can consider the map Q which associates to $X \in \mathcal{S}$ the set of classical probabilities $\mathbb{P}_{\rho}$ for $\rho \in \mathbf{Q}_X$. If X divides Y, the fact that the direct image $Y_* \mathbb{P}(\rho)$ of $\rho \in \mathbf{Q}_X$ coincides with the law $\mathbb{P}_{Y_*(\rho)}$ gives the following result:

Lemma 4. $\rho \mapsto \mathbb{P}_{\rho}$ is a natural transformation from the functor $\iota^* \mathbf{Q}$ to the functor Q.

Definition 5. This natural transformation is called the *Trace*, and we denote by Tr_X its value in X, *i.e.*, $Tr_X(\rho) = \mathbb{P}_{\rho}$, seen as a map from $\mathbf{Q}_X$ to $\mathcal{Q}_X$.

In general there is no natural transformation in the other direction, from $\mathcal{Q}_X$ to $\mathbf{Q}_X$.

Remark that the trace sends a unitary invariant functor to a symmetric functor.

4.3. **Quantum Information Homology.**

As in the classical case, we can consider the ringed site given by the category $\mathbf{S}$, equipped with the sheaf of monoids $\{\mathbf{S}_X; X \in \mathbf{S}\}$. In the ringed topos of sheaves of $\mathbf{S}$-modules, the choice of a probability functor $\mathbf{Q}$ generates remarkable elements in this topos, formed by the functional space $\mathbf{F}$ of measurable functions on $\mathbf{Q}$ with values in $\mathbb{R}$. The action of the monoid (or the generated ring) being given by averaged conditioning, and the arrows being given by transposition of direct images. Then, the quantum information co-homology is the topos co-homology:

$$(83) \qquad H^m(\mathbf{S}, \mathbf{Q}) = Ext_{\mathbf{S}}^m(\mathbb{R}; \mathbf{F})$$

However, as in the classical case, we can define directly the co-homology with a bar resolution of the constant sheaf, as follows:

A set of functions F_X of m observables $Y_1, ..., Y_m$ divided by X, and one density ρ indexed by $X \in \mathbf{S}$, is said *local*, when for any decomposition X dividing a decomposition Y, we have, for each ρ in $\mathbf{Q}_X$,

$$(84) \qquad F_X(Y_1; ...; Y_m; \rho) = F_X(Y_1; ...; Y_m; Y_*(\rho)).$$

For $m = 0$ this equation expresses that the family F_X is an element of the topos.

For every m, a collection $F_X, X \in \mathbf{S}$ is a natural transform F from a free functor $\mathbf{S}_m$ to the functor $\mathbf{F}$.

Be careful that in the quantum context, it is not true in general that locality is equivalent to the condition saying that the value $F_X(Y_1; ...; Y_n; \rho)$ depends only on the family of conditioned densities $E_{A_i}^* \rho E_{A_i}; i = 0, ..., m$, where A_i is one of the possible events defined by Y_i.

In fact it depends on the choice of $\mathbf{Q}$; for instance it is false for a $\mathbf{Q}^{max}$, but it is true for a $\mathbf{Q}^{min}$.

The counter-example in the case of $\mathbf{Q}^{max}$ is given by a function $F(\rho)$ which is independent of X. It is local (in the sense of topos that we adopt) but it is non-local in the apparently more natural sense that it depends only of ρ_X. This is important to have this quantum particularity in the mind for understanding the following discussion.

As in the classical case, the action of observables on local functions is given by the average of conditioning, in the manner of Shannon, but using the Von Neumann conditioning:

$$(85) \qquad Y.F(Y_0; ...; Y_m; \rho) = \sum_A Tr(E_A^* \rho E_A) F(Y_0; ...; Y_m; \rho|A)$$

where the E_A's are the spectral projectors of the bundle Y. In this definition there is no necessity to assume that Y commutes with the Y_j's.

Remind that, when $E_A^* \rho E_A$ is non-zero, $\rho|A$ is equal to $E_A^* \rho E_A / Tr(E_A^* \rho E_A)$, and verifies the normalization condition that the trace equals to one. When $E_A^* \rho E_A$ is equal to zero, the factor $Tr(E_A^* \rho E_A)$ is zero, then by convention the corresponding term F is absent.

The proof of the Lemma 1 applies without significant change to prove that the above formula defines an action of the monoid functor $\mathbf{S}_X$.

Then, the definition of co-homology is given exactly as we have done for the classical case, by introducing the Hochschild operator:

(86)
$$\widehat{\delta}^m F(Y_1;...;Y_{m+1};\rho)$$
$$=Y_1.F(Y_2;...;Y_{m+1};\rho)+\sum_1^m(-1)^iF(...;(Y_i,Y_{i+1});...;Y_{m+1};\rho)+(-1)^{m+1}F(Y_1;...;Y_m;\rho).$$

The Von-Neumann entropy is defined by the following formula

(87)
$$S(\rho)=\mathbb{E}_\rho(-\log_2(\rho))=-Tr(\rho\log_2(\rho)).$$

For any density functor $\mathbf{Q}$ which is adapted to $\mathbf{S}$, the Von-Neumann entropy defines a local 0-cochain, that we will call S_X, and is simply the restriction of S to the set $\mathbf{Q}_X$. If ρ belongs to $\mathbf{Q}_X$ and if X divides Y, the law $Y_*\rho$, which is the same hermitian form as ρ belongs to $\mathbf{Q}_Y$ by functoriality, thus $S(Y_*\rho)=S(\rho)$ is translated by $S_X(\rho)=S_Y(Y_*\rho)$. This 0-cochain will be simply named the Von Neumann entropy.

In the case of $\mathbf{Q}^{max}$, S_X gives the same value at all places X. In the case of $\mathbf{Q}^{min}$ it coincides with $S(\rho_X)$, where ρ_X denotes the restriction to the decomposition X.

Be careful: $\rho \mapsto S(\rho_X)$ is not a local 0-cochain for $\mathbf{Q}^{max}$. In fact in the case of $\mathbf{Q}^{max}$ we have the same set $\mathbf{Q}=\mathbf{Q}_X$ for every place X, thus, if we take for X a strict divisor of Y and if we take a density ρ such that, for the restrictions of ρ, the spectrum of ρ_Y and ρ_X are different, then, in general, we do not have $S_X(\rho)=S_Y(Y_*\rho)$, even if, as it is the case in the quantum context, $Y_*\rho=\rho$.

Remark that in the case of $\mathbf{Q}^{max}$, where every function of ρ independent of X is a cochain of degree zero, the particular functions which depends only on the spectrum of ρ are invariant under the action of the unitary group, and they are the only 0-cochains which are invariant by this group.

Definition 6. Suppose that $\mathbf{S}$ and $\mathbf{Q}$ are invariant by the unitary group, as is $U\mathbf{B}$, we say that an m-cochain F is *invariant*, if for every X in $\mathbf{S}$ dividing $Y_1,...,Y_m$ in $\mathbf{S}$, every ρ in $\mathbf{Q}_X$ and every g in the group $U(h_0)$, we have

(88)
$$F_{g.X}(g.Y_1,...,g.Y_m;g.\rho)=F_X(Y_1;...;Y_m;\rho);$$

where $g.X=gXg^*$, $g.Y_i=gY_ig^*;i=1,...,m$ and $g.\rho=g\rho g^*$.

This is compatible with the naturality assumption (functoriality by direct images), because direct image is a covariant operation.

Note that conditioning is also covariant if we change all variables and laws coherently. Thus the action of the monoids $\mathbf{S}_X$ on cochains respects the invariance.

Then the coboundary $\hat{\delta}$ preserves invariance. Thus the co-homology of the invariant co-chains is well defined. We call it the *invariant information co-homology*, and we will denote it by $H_U^*(\mathbf{S};\mathbf{Q})$, U for unitary.

Invariant co-cochains form a subcomplex of ordinary cochains, then we have a well defined map from $H_U^*(\mathbf{S};\mathbf{Q})$ to $H^*(\mathbf{S};\mathbf{Q})$.

The invariant 0-co-chains depend only on the spectrum of ρ in the sets $\mathbf{Q}_X$.

The invariant co-homology is probably a more natural object from the point of view of Physics. It is also on this co-homology that we were able to obtain constructive results.

The classical entropy of the decomposition $\{E_j\}$ and the quantum law ρ is

$$(89) \qquad H(X;\rho) = -\sum_j Tr(E_j^* \rho E_j) \log_2(Tr(E_j^* \rho E_j))$$

In general it is not true that $H(X;\rho) = H(Y;Y_*\rho)$ when X divides Y. Thus the Shannon (or Gibbs) entropy is *not* a local 0-cochain, but it is a local 1-cochain, *i.e.*, if $X \to Y \to Z$ we have

$$(90) \qquad H_X(Z;\rho_X) = H_Y(Z;Y_*\rho_X),$$

Moreover it is a spectral 1-cochain for any $\mathbf{Q}^{min}$.

The following result is well known, *cf.* Nielsen and Chuang [46].

Lemma 5. Let X, Y be two commuting families of observables; we have

$$(91) \qquad S_{(X,Y)}(\rho) = H(Y;\rho) + Y.S_X(\rho)$$

Proof. We denote by $\alpha, \beta, ...$ the indices of the different values of X, by $k, l, ...$ the indices of the different values of Y, and by $i, j, ...$ the indices of a basis $I_{k,\alpha}$ of eigenvectors of the conditioned density $\rho_{k,\alpha} = E_{k,\alpha}^* \rho E_{k,\alpha}$ constrained by the projectors $E_{k,\alpha}$ of the pair (Y, X). The probability $p_k = P_\rho(X = \xi_k)$ is equal to the sum over i, α of the eigenvalues $\lambda_{i,k,\alpha}$ of $\rho_{k,\alpha}$. We have

$$\begin{aligned}
Y.S(X;\rho) &= -\sum_k p_k \sum_{i,\alpha} \frac{\lambda_{i,k,\alpha}}{p_k} \log_2\left(\frac{\lambda_{i,k,\alpha}}{p_k}\right) \\
&= -\sum_{i,k,\alpha} \lambda_{i,k,\alpha} \log_2(\lambda_{i,k,\alpha}) + \sum_{i,k,\alpha} \lambda_{i,k,\alpha} \log_2(p_k) \\
&= -\sum_{i,k,\alpha} \lambda_{i,k,\alpha} \log_2(\lambda_{i,k,\alpha}) + \sum_k p_k \log_2(p_k).
\end{aligned}$$

Remark 6. Taking $X = 1$, or any scalar matrix, the preceding Lemma 5 expresses the fact that classical entropy is a derived quantity measuring the default of equivariance of the quantum entropy:

$$(92) \qquad H(Y;\rho) = S_Y(\rho) - (Y.S_Y)(\rho).$$

Lemma 6. For any $X \in \mathbf{S}$, dividing $Y \in \mathbf{S}$ and $\rho \in \mathbf{Q}_X$,

$$(93) \qquad \hat{\delta}(S_X)(Y;\rho) = -H_X(Y;\rho).$$

Proof. This is exactly what says the Lemma 5 in this particular case, because in this case $(X, Y) = X$, and, by definition, we have $\hat{\delta}(S_X)(Y;\rho) = Y.S_X(\rho) - S_X(\rho)$. To insist, we give a direct proof with less indices for this case:

$$\begin{aligned}
Y.S_X(\rho) &= -\sum_i p_i \sum_k \frac{\lambda_{ik}}{p_i} \log_2 \frac{\lambda_{ik}}{p_i} \\
&= -\sum_{ik} \lambda_{ik} \log_2 \lambda_{ik} + \sum_{ik} \lambda_{ik} \log_2 p_i \\
&= S_X(\rho) + \sum_i \log_2 p_i \sum_k \lambda_{ik} = S_X(\rho) + \sum_i (\log_2 p_i) p_i \\
&= S_X(\rho) - H_X(Y;\mathbb{P}_\rho) = S_X(\rho) - H_X(Y;\rho).
\end{aligned}$$

The Lemma 6 says that (up to the sign) the Shannon entropy is the co-boundary of the Von-Neumann entropy. This implies that the Shannon entropy is a 1-co-cycle, as in the classical case, but now it gives zero in co-homology.

Note that the result is true for any $\mathbf{Q}$, thus for $\mathbf{Q}^{min}$ and for $\mathbf{Q}^{max}$ as well.

Consider a maximal observable X_0 in $\mathbf{S}$, *i.e.*, a maximal set of commuting observables in $\mathbf{S}$, the elements of this maximal partition form a finite set Ω_0. If $\mathbf{S}$ is invariant by the group $U(E; h_0)$, all the maximal observables are deduced from X_0 by applying a unitary base change. Suppose that the functor $\mathbf{Q}$ is invariant also; then we get automatically a symmetric classical structure of information $\mathcal{S}$ on Ω_0, given by the elements of $\mathbf{S}$ divided by X_0. And $\mathcal{S}$ is equipped with a symmetric classical functor of probability, given by the probability laws associated to the elements of $\mathcal{S}$.

Remind that we defined the *trace* from quantum probabilities to classical probabilities, by taking the classical $\mathbb{P}_\rho$ for each ρ, and we noticed that the trace is compatible with invariance and symmetry by permutations.

Definition 7. To each classical co-chain F^0 we can associate a quantum co-chain $F = tr^* F^0$ by putting

(94) $$tr^*(F)_X(Y_1; ...; Y_m; \rho) = F_X^0(Y_1; ...; Y_m; tr_X(\rho)).$$

The following result is straightforward:

Proposition 3. (i) The trace of co-chains defines a map of the classical information Hochschild complex to the quantum one, which commutes with the co-boundaries, i.e., the map tr^* defines a map from the classical information Hochschild complex to the quantum Hochschild complex; (ii) this map sends symmetric cochains to invaraint cochains; it induces a natural map from the symmetric classical information co-homology $H_\mathcal{S}^*(\mathcal{S}; \mathcal{Q})$ to the invariant quantum information co-homology $H_U^*(\mathbf{S}; \mathbf{Q})$.

The Lemma 6 says that the entropy class goes to zero.

Remark 7. In a preliminary version of these notes, we considered the expression $s(X; \rho) = S(\rho_X) - S(\rho)$ and showed it satisfies formally the 1-cocycle equation. But we suppress this consideration now, because s is not local, thus it plays no interesting role in homology. For instance in $\mathbf{Q}^{min}$, $S(\rho_X)$ is local but $S(\rho)$ is not and in $\mathbf{Q}^{max}$, $S(\rho)$ is local but $S(\rho_X)$ is not.

Definition 8. In an information structure $\mathbf{S}$ we call *edge* a pair of decompositions (X, Y) such that X, Y and XY belong to $\mathbf{S}$; we say that an edge is *rich* when both X and Y have at least two elements and XY cuts those two in four distinct subspaces of E. The structure $\mathbf{S}$ is connected if every two points are joined by a sequence of edges, and it is *sufficiently rich* when every point belongs to a rich edge. We assume a maximal set of subspaces $U\mathbf{B}$ is given in the Grassmannian of E, in such a way that the maximal elements X_0 of $\mathbf{S}$ (i.e., initial in the category) are made by pieces in $U\mathbf{B}$. The density functor $\mathbf{Q}$ is said *complete* with respect to $\mathbf{S}$ (or $U\mathbf{B}$) if for every X, the set $\mathbf{Q}_X$ contains the positive hermitian forms on the blocs of X, that give scalar blocs $\rho_{\alpha\beta}$ for two elements E_α, E_β of a maximal decomposition. (All that is simplified when we choose a basis, and take maximal commutative subalgebras of operators, but we want to be free to consider simplicial complexes.)

Theorem 3. (i) for any unitary invariant quantum information structure $\mathbf{S}$, which is connected and sufficiently rich, and for the canonical invariant density functor $\mathbf{Q}^{can}(\mathbf{S})$, (i.e., the density functor which is minimal and complete with respect to $\mathbf{S}$), the invariant information co-homology of degree one $H_U^1(\mathbf{S};\mathbf{Q})$ is zero. (ii) Under the same hypothesis, the invariant co-homology of degree zero has dimension one, and is generated by the constants. Then, up to an additive constant, the only invariant 0-cochain which has the Shannon entropy as co-boundary is (minus) the Von-Neumann entropy.

Proof. (I) Let X, Y be two orthogonal decompositions of E belonging to $\mathbf{S}$ such that (X, Y) belongs to $\mathbf{S}$, and ρ an element of $\mathbf{Q}$. We name $A_{k_i}; i = 1, ..., m$ the summands of X, and $B_{\alpha_j}; j = 1, ..., l$ the summands of Y; the projections $E_{k_i}\rho E_{k_i}; i = 1, ..., m$ resp. $E_{\alpha_j}\rho E_{\alpha_j}; j = 1, ..., l$ of ρ on the summands of X, resp. Y are denoted by $\rho_{k_i}; i = 1, ..., m$ and $\rho_{\alpha_j}; j = 1, .., l$ respectively. The projections by the commutative products $E_{k_i}E_{\alpha_j}$ are denoted by $\rho_{k_i, \alpha_j}; i = 1, ..., m, j = 1, .., l$.

Let f be a 1-cocycle, we write $f(X; \rho) = F(\rho)$, $f(Y; \rho) = H(\rho)$ and $G(\rho) = f(X, Y; \rho)$. Note that in $\mathbf{Q}^{min}$, F is a function of the ρ_{k_i}, H a function of the ρ_{α_j} and G a function of the ρ_{k_i, α_j}, but there is no necessity too assume this property; we can always consider these functions restricted to diagonal blocs, which are arbitrary due to the completeness hypothesis.

For any positive hermitian ρ', we write $\rho'|\alpha$, resp. $\rho'|i$ the form conditioned by the event B_α resp. A_i.

The co-cycle equation gives the two following equations, that are exchanged by permuting X and Y:

(95) $$\sum_{\alpha_j} Tr(\rho_{\alpha_j})F((\rho_{k_i}|\alpha_j); i = 1, ..., m) = G((\rho_{k_i, \alpha_j}); i, j) - H((\rho_{\alpha_j}); j),$$

(96) $$\sum_i Tr(\rho_{k_i})H((\rho_{\alpha_j}|k_i); j) = G((\rho_{k_i, \alpha_j}); i, j) - F((\rho_{k_i}); i).$$

Now we consider a particular case, where the small blocs $\rho_{k, \alpha}$ are zero except for (k_1, α_2) and (k_j, α_1) for $j = 2, ..., m$. We denote by h_1 the forme ρ_{k_1, α_2} and by h_i the form ρ_{k_i, α_1}, for $i = 2, ..., m$. Remark that $Tr(h_1 + h_2 + ... + h_m) = 1$.

(II) As in the classical case, it is a general fact for a 1-cocycle f and any variable Z the value $f(Z; \rho)$ is zero if ρ is zero outside one of the orthogonal summand C_a of Z; because the equation $f_X(Z, Z; \rho) = f_X(Z; \rho) + Z.f_X(Z; \rho)$ implies $Z.f_X(Z, \rho) = 0$, and if ρ has only one non-zero factor ρ_a, we have

(97)
$$Z.f(Z; \rho) = \sum_b Tr(\rho_b)f(Z; \rho_b/Tr(\rho_b)) = Tr(\rho_a)f(Z; \rho_a/Tr(\rho_a)) = 1.f(Z; \rho_a).$$

Therefore in the particular case that we consider, we get for any i that $H((\rho_{\alpha_j}|k_i); j) = 0$. Consequently the Equation (96) equals the term in G to the term in F, and we can report this equality in the first equation. By denoting $1 - x_1 = Tr(\rho_{\alpha_1})$, this gives
(98)
$$H((\rho_{\alpha_j}); j = 1, 2) = F((\rho_{k_i}); i = 1, ..., m) - (1 - x_1)F((0, \frac{h_2}{1 - x_2}, ..., \frac{h_m}{1 - x_m})).$$

Now if we add the condition $h_3 = ... = h_m = 0$ we have $F(0, h_2/(1-x_1), 0, ..., 0) = 0$ for the reason which eliminated the $H((\rho_{\alpha_j}|k_i); j)$; thus we obtain

(99) $$H(\rho_{\alpha_1}); j = 1, 2) = F((\rho_{k_1}); i = 1, 2).$$

This is a sufficiently strong constraints for implying that both terms are functions of h_1, h_2 only, and that of course they coincide as functions of these small blocs. First this gives a recurrence equation, which, as in the classical case is able to reconstruct $F((\rho_{k_i}); i = 1, ..., m)$ from the case of two blocs:

(100)
$$F(X; (\rho_{k_i}); i = 1, ..., m) = F(X; (\rho_{k_1}, \rho_{k_2}, 0, ..., 0) - (1-x_1)F(X; (0, \frac{h_2}{1 - x_2}, ..., \frac{h_m}{1 - x_m})).$$

(III) We are left with the study of two binary variables Y, Z, forming a rich edge.

The blocs of ρ adapted to the joint ZY are denoted by $\rho_{00}, \rho_{01}, \rho_{10}, \rho_{11}$, where the first index refers to Y and the second index refers to Z, but the blocs that are allowed for Y and Z are more numerous than four; there exist out of diagonal blocs, and their role will be important in our analysis. For Y we have matrices ρ_0^0 and ρ_1^0, and for Z we have matrices ρ_0^1 and ρ_1^1;

(101) $$\rho_0^0 = \begin{pmatrix} \rho_{00} & \rho_{001}^0 \\ \rho_{010}^0 & \rho_{01} \end{pmatrix} \quad \rho_1^0 = \begin{pmatrix} \rho_{10} & \rho_{101}^0 \\ \rho_{111}^0 & \rho_{11} \end{pmatrix}$$

(102) $$\rho_0^1 = \begin{pmatrix} \rho_{00} & \rho_{001}^1 \\ \rho_{010}^1 & \rho_{01} \end{pmatrix} \quad \rho_1^1 = \begin{pmatrix} \rho_{10} & \rho_{101}^1 \\ \rho_{111}^1 & \rho_{11} \end{pmatrix}$$

They are disposed in sixteen blocs for ρ, but certain of them, noted with stars, cannot be seen from ρ_Y or ρ_Z:

(103) $$\rho = \begin{pmatrix} \rho_{00} & \rho_{001}^0 & \rho_{001}^1 & \rho_{001}^* \\ \rho_{010}^0 & \rho_{01} & \rho_{101}^* & \rho_{001}^1 \\ \rho_{010}^1 & \rho_{010}^* & \rho_{10} & \rho_{101}^0 \\ \rho_{111}^* & \rho_{111}^1 & \rho_{111}^0 & \rho_{11} \end{pmatrix}$$

Now the co-cycle equations are

(104) $$F(Y, Z; \rho) = Y.F(Z; \rho) + F(Y; \rho) = Z.F(Y; \rho) + F(Z; \rho),$$

giving the symmetrical relation:

(105) $$Y.F(Z; \rho) - F(Z; \rho) = Z.F(Y; \rho) - F(Y; \rho).$$

The conditioning makes many blocs disappear. Then, by denoting with latin letters the corresponding traces, and taking in account explicitly the blocs that must count, the symmetrical identity gives, for any ρ, the following developed equation:

$$
\begin{aligned}
& (p_{00} + p_{01})F_Z(\frac{\rho_{00}}{p_{00} + p_{01}}, \frac{\rho_{01}}{p_{00} + p_{01}}, 0, 0) + (p_{10} + p_{11})F_Z(0, 0, \frac{\rho_{10}}{p_{10} + p_{11}}, \frac{\rho_{01}}{p_{10} + p_{11}}) \\
& - F_Z(\rho_{00}, \rho_{001}^1, \rho_{01}, \rho_{001}^1, \rho_{010}^1, \rho_{10}, \rho_{111}^1, \rho_{11}) \\
= \ & (106) (p_{00} + p_{10})F_Y(\frac{\rho_{00}}{p_{00} + p_{10}}, 0, \frac{\rho_{10}}{p_{00} + p_{11}}, 0) + (p_{01} + p_{11})F_Y(0, \frac{\rho_{10}}{p_{01} + p_{11}}, 0, \frac{\rho_{11}}{p_{01} + p_{11}}) \\
& - F_Y(\rho_{00}, \rho_{001}^0, \rho_{01}, \rho_{001}^0, \rho_{010}^0, \rho_{10}, \rho_{111}^0, \rho_{11}).
\end{aligned}
$$

(IV) Now we make appeal to the invariance hypothesis: let us apply a unitary transformation g which respects the two summands of Y but does not necessarily respect the summands of Z we replace Z by gZg^*, and ρ by $g\rho g^*$, the value of $F_Y(\rho_0^0, \rho_1^0)$ does not change. Our claim is that the only function F_Y which is

compatible with the Equation (106) for every ρ are functions of the traces of the blocs.

For the proof, we assume that all the blocs are zero except the eight blocs concerning Y. In this case, we see that the last function $-F_Y$ of the right member, involves the eight blocs, but all the other functions involve only the four diagonal blocs. Thus our claim follows from the following result:

Lemma 7. A measurable function f on the set H of hermitian matrices which is invariant under conjugation by the unitary group U_n and invariant by the change of the coefficient a_{1n}, the farthest from the diagonal, is a function of the trace.

Proof. An invariant function for the adjoint representation is a function of the traces of the exterior powers $\Lambda^k(\rho)$, but these traces are coefficients in the basis $e_{i_1} \wedge e_{i_1} \wedge ... \wedge e_{i_k}$, and the elements divisible by $e_1 \wedge e_n$ cannot be neglected, as soon as $k \geq 2$.

Therefore the co-cycle F_Y, F_Z comes from the image of tr^* in proposition 3. Then the recurrence relation (100) implies that the same is true for the whole co-cycle F.

(V) For concluding the proof of (i), we appeal to the Theorem 1, that the only non-zero cocycles in this context, connected and sufficiently rich, are multiples of the classical entropy. However, the Lemma 5 says that the entropy is a co-boundary.

(VI) To prove (ii), we have to show that every 0-cocycle $X \mapsto f_X(\rho)$, which depends only on the spectrum of ρ, is a constant. We know that a spectral function is a measurable function $\varphi(\sigma_1, \sigma_2, ...)$ of the elementary symmetric functions $\sigma_1 = \sum_i \lambda_i, \sigma_2 = \Sigma_{i<j} \lambda_i \lambda_j, ...$

And, to be a 0-cocycle, f must verify, for every pair of decompositions, $X \to Y$, the equation

$$(107) \qquad f_X(\rho) = \sum_i P_\rho(Y = i) f_X(\rho|(Y = i)).$$

Explicitly, if $f_X(\rho) = \varphi_X(\sigma_1, \sigma_2, ...)$,

$$(108) \qquad \varphi_X(\sigma_1, \sigma_2, ...) = \sum_i \sigma_1(\lambda_{k,i}) \varphi_X(\sigma_1(\lambda_{ki}), ...)$$

where each bloc $\rho|i$ has the spectrum $\{\lambda_{k,i}; k \in J_i\}$. For a sufficiently rich edge $X = YZ$, we have with four eigenvalues repeated as it must be to fulfill the dimensions:

$$
\begin{aligned}
(109) \quad & f(\lambda_{00}^{(n_{00})}, n_{00}^{(n_{00})}, \lambda_{01}^{(n_{01})}, \lambda_{10}^{(n_{10})}, \lambda_{11}^{(n_{11})}) \\
= \quad & (n_{00}\lambda_{00} + n_{01}\lambda_{01}) f\left(\frac{\lambda_{00}^{(n_{00})}}{n_{00}\lambda_{00} + n_{01}\lambda_{01}}, \frac{\lambda_{01}^{(n_{01})}}{n_{00}\lambda_{00} + n_{01}\lambda_{01}}\right) \\
& + (n_{10}\lambda_{10} + n_{11}\lambda_{11}) f\left(\frac{\lambda_{10}^{(n_{10})}}{n_{10}\lambda_{10} + n_{11}\lambda_{11}}, \frac{\lambda_{11}^{(n_{11})}}{n_{10}\lambda_{10} + n_{11}\lambda_{11}}\right),
\end{aligned}
$$

and

$$
\begin{aligned}
(110) \quad & f(\lambda_{00}^{(n_{00})}, n_{00}^{(n_{00})}, \lambda_{01}^{(n_{01})}, \lambda_{10}^{(n_{10})}, \lambda_{11}^{(n_{11})}) \\
= \quad & (n_{00}\lambda_{00} + n_{10}\lambda_{10}) f\left(\frac{\lambda_{00}^{(n_{00})}}{n_{00}\lambda_{00} + n_{10}\lambda_{10}}, \frac{\lambda_{10}^{(n_{10})}}{n_{00}\lambda_{00} + n_{10}\lambda_{10}}\right) \\
& + (n_{01}\lambda_{01} + n_{11}\lambda_{11}) f\left(\frac{\lambda_{01}^{(n_{01})}}{n_{01}\lambda_{01} + n_{11}\lambda_{11}}, \frac{\lambda_{11}^{(n_{11})}}{n_{01}\lambda_{01} + n_{11}\lambda_{11}}\right),
\end{aligned}
$$

By equating the two second members, taking $\lambda_{01} = \lambda_{00} = 0$, and varying $\lambda_{10}, \lambda_{11}$, we find that $f(x, y)$ is the sum of a constant and a linear function.

At the end, f_X must be the sum of a constant and a linear function for every X. However, a linear symmetric function is a multiple of σ_1. As ρ is normalized by the condition $Tr(\rho) = 1$, only the constant survives.

Remark 8. In his book "Structure des Systemes Dynamiques", J-M. Souriau [50] showed that the mass of a mechanical system is a degree one class of co-homology of the relativity group with values in its adjoint representation; this class being non-trivial for classical Mechanics, with the Galileo group, and becoming trivial for Einstein relativistic Mechanics, with the Lorentz-Poincare group. Even if we are conscious of the big difference with our construction, the above result shows the same thing happens for the entropy, but going from classical statistics to quantum statistics.

From the philosophical point of view, it is important to mention that the main difference between classical and quantum information co-homology in degree less than one, is the fact that the certitude, 1, becomes highly non-trivial in the quantum context. This point is discussed in particular by Gabriel Catren [15]. In geometric quantization the first ingredient, discovered by Kirillov, Kostant and Souriau in the sixties, is a circular bundle over the phase space that allows a non-trivial representation of the constants. The second ingredient also discovered by the same authors, is the necessity to choose a polarization, which correspond to the choice of a maximal commutative Poisson sub-algebra of observable quantities. This second ingredient appears in our framework through the limitations of information categories to collection of commutative Boolean algebras, coming from the impossibility to define manageable joints for arbitrary pair of observables.

5. PRODUCT STRUCTURES, KULLBACK-LEIBLER DIVERGENCE, QUANTUM VERSION

In this short section, we use both the homogeneous bar-complex and the non-homogeneous complex.

A natural extension of the information co-cycles is to look at the measurable functions

$$(111) \qquad F(X_0; X_1; ...; X_m; P_0; P_1, P_2, ..., P_n; X),$$

of several probability laws P_j (or density of states respectively) on Ω (or E respectively) belonging to the space $\mathcal{Q}_X$ that are absolutely continuous with respect to P_0, and several decompositions Y_i less fine than X. To be homogeneous co-chains these functions have to behave naturally under direct image $Y_*(P_i)$, and to satisfy the equivariance relation:

$$F((Y, X_0); (Y, X_1); ...; (Y, X_m); P_0; P_1, P_2, ..., P_n; X)$$
$$(112) \qquad = \quad Y.F(X_0; X_1; ...; X_m; P_0; P_1, P_2, ..., P_n; X),$$

for any $Y \in \mathcal{S}_X$ (resp. $\mathbf{S}_X$), where

$$Y.F(X_0; X_1; ...; X_m; P_0; P_1, P_2, ..., P_n; X)$$
$$(113) = \int_{E_Y} dY_* P_0(y) F(X_0; X_1; ...; X_m; P_0 | Y = y; P_1 | Y = y, ..., P_n | Y = y; X).$$

Note that a special role is played by the law P_0, which justifies the coma notation.

The proof of the Lemma 1 in Section 2.1 extends without modification to show that this defines an action of semi-group.

Then we define the homogeneous co-boundary operator by

$$
\begin{aligned}
&\delta F(X_0; X_1; ...; ...; X_m; X_{m+1}; P_0; P_1, P_2, ..., P_n; X) \\
(114) \quad &= \sum_i (-1)^i F(X_0; ...; \widehat{X_i}; ...; X_m; X_{m+1}; P_0; P_1, P_2, ..., P_n; X).
\end{aligned}
$$

The co-cycles are the elements of the kernel of δ and the co-boundaries the elements of the image of δ (with a shift of degree). The co-homology groups are the quotients of the spaces of co-cycles by the spaces of co-boundaries.

This co-homology is the topos co-homology $H_{\mathcal{S}}^*(\mathbb{R}, \mathcal{F}_n)$, of the module functor $\mathcal{F}_n$ of measurable functions of $n + 1$-uples of probabilities, in the ringed topos $\mathcal{S}$ (resp. $\mathbf{S}$ in the quantum case).

There is also the non-homogeneous version: a m-cocycle is a family of functions $F_X(X_1; ...; ...; X_m; P_0; P_1, P_2, ..., P_n)$ which behave naturally under direct images, without equivariance condition.

The co-boundary operator is copied on the Hochschild operator: then we define the homogeneous co-boundary operator by

$$
\begin{aligned}
&\widehat{\delta} F_X(X_0; X_1; ...; ...; X_m; P_0; P_1, P_2, ..., P_n) \\
(115) \quad &= (X_0.F_X)(X_1; ...; ...; X_m; P_0; P_1, P_2, ..., P_n) \\
&\quad + \sum_i (-1)^{i+1} F(X_0; ...; \widehat{X_i}; ...; X_m; P_0; P_1, P_2, ..., P_n; X).
\end{aligned}
$$

Let us recall the definition of the Kullback-Leibler divergence (or relative entropy) between two classical probability laws P, Q on the same space Ω, in the finite case:

$$
(116) \qquad\qquad H(P; Q) = -\sum_i p_i \log \frac{q_i}{p_i}.
$$

Over an infinite set, it is required that Q is absolutely continuous with respect to P with a L^1-density dQ/dP, and the definition is

$$
(117) \qquad\qquad H(P; Q) = -\int_\Omega dP(\omega) \log \frac{dQ(\omega)}{dP(\omega)}.
$$

When $dQ(\omega)/dP(\omega) = 0$, the logarithm is $-\infty$ and due to the sign minus, we get a contribution $+\infty$ in H, thus, if this happens with probability non-zero for P the divergence is infinite positive. To get a finite number we must suppose also that P is absolutely continuous with respect to Q, i.e., P and Q are equivalent.

The analogous formula defines the quantum Kullback-Leibler divergence (or quantum relative entropy), cf. Nielsen-Chuang [46], between two density of states ρ, σ on the same Hilbert space E, in the finite dimensional case:

$$
(118) \qquad\qquad S(\rho; \sigma) = -Tr(\rho(\log \sigma - \log \rho)).
$$

In the case of an infinite dimensional Hilbert space, it is required that the trace is well defined.

These quantities are positive or zero, and they are zero only in the case of equality of the measures (resp. the densities of states). It is the reason why it is frequently used as a measure of distance between two laws.

Proposition 4. The map which associates to X in $\mathcal{S}$, Y divided by X, and two laws P, Q the quantity $H(Y_*P; Y_*Q)$ defines a non-homogeneous 1-cocycle, denoted $H_X(Y; P; Q)$.

Proof. As we already know that the classical Shannon entropy is a non-homogeneous 1-cocycle, it is sufficient to prove the Hochschild relation for the new function

$$(119) \qquad H_m(Y; P; Q) = -\sum_i p_i \log q_i.$$

Let us denote by p_{ij} (resp. q_{ij}) the probability for P (resp. Q) of the event $Y = x_i, Z = y_j$, and by p^j (resp. q^j) the probability for P (resp. Q) of the event $Z = y_j$; then the probability p_i^j (resp. q_i^j) of $Y = x_i$ knowing that $Z = y_j$ for P (resp. for Q) is equal to p_{ij}/p^j (resp. q_{ij}/q^j), and we have

$$(120) \qquad H_m((Z,Y); P, Q) = -\sum_i \sum_j p_{ij} \log q_{ij}$$

$$(121) \qquad = -\sum_j p^j \sum_i p_i^j \log(q^j q_i^j)$$

$$(122) \qquad = -\sum_j p^j \log q^j \left(\sum_i p_i^j\right) - \sum_j p^j \sum_i p_i^j \log q_i^j$$

$$(123) \qquad = -\sum_j p^j \log q^j - \sum_j p^j \sum_i p_i^j \log q_i^j;$$

the first term on the right is $H_m(Z; P; Q)$ and the second is $(Z.H_m)(Y; P; Q)$, Q.E.D.

This defines a homogeneous co-cycle for pairs of probability laws $H_X(Y; Z; P; Q) = H_X(Y; P; Q) - H_X(Z; P; Q)$, named *Kullback-divergence variation.*

In the quantum case, for two densities of states ρ, σ we define in the same manner a classical Kullback-Leibler divergence $H_X(Y; \rho; \sigma)$ by the formula

$$(124) \qquad H_X(Y; \rho; \sigma) = \sum_k (Tr(\rho_k \log(Tr(\rho_k)) - \log(Tr(\sigma_k))));$$

where the index k parameterizes the orthogonal decomposition E_k associated to Y and where ρ_k (resp. σ_k) denotes the matrix $E_k^* \rho E_k$ (resp. $E_k^* \sigma E_k$). It is the Kullback-Leibler divergence of the classical laws associated to the direct images ρ and σ respectively.

But in the case of quantum information theory, we can also define a quantum divergence, for any pair densities of states (ρ, σ) in $\mathbf{Q}_X$,

$$(125) \qquad S_X(\rho; \sigma) = -Tr(\rho \log \sigma).$$

Lemma 8. For any pair (X, Y) of commuting hermitian operators, such that Y divides X, the function S_X satisfies the relation

$$(126) \qquad S(X, Y)(\rho; \sigma) = H_Y(X; \rho; \sigma) + X.S_Y(\rho; \sigma);$$

where H_X of two variables denotes the mixed entropy, defined by Equation (119).

Proof. As in the proof of the Lemma 4, we denote by $\alpha, \beta, ...$ (resp. $k, l, ...$) the indices of the orthogonal decomposition Y (resp. X), and by $i, j, ...$ the indices of a basis $\phi_{i,k,\alpha}$ of the space $E_{k,\alpha}$ made by eigenvectors of the matrix $\varrho_{k,\alpha} = E_{k,\alpha}^* \rho E_{k,\alpha}$ belonging to the joint operator (X, Y). In a general manner if M is

an endomorphism of $E_{k,\alpha}$ we denote by $M_{i,k,\alpha}$ the diagonal coefficient of index (i, k, α). The probability p_k (resp. q_k) for ρ (resp. σ) of the event $X = \xi_k$ is equal to the sum over i, α of the eigenvalues $\lambda_{i,k,\alpha}$ of $\rho_{k,\alpha}$ (resp. $\mu_{i,k,\alpha}$ of $\sigma_{k,\alpha}$). And the restricted density ρ^{Y_k} (resp. σ^{Y_k}), conditioned by $X = \xi_k$, is the sum over α of $\varrho_{k,\alpha}$ (resp. of $\sigma_{k,\alpha}$) divided by p_k (resp. q_k). We have

$$X.S_Y(\rho; \sigma) = -\sum_k p_k Tr(\rho_{Y_k} \log \sigma^{Y_k}) \tag{127}$$

$$= -\sum_k p_k \sum_{i,\alpha} \frac{\lambda_{i,k,\alpha}}{p_k} (\log \frac{\sigma_k}{q_k})_{i,k,\alpha} \tag{128}$$

$$= \sum_{i,k,\alpha} \lambda_{i,k,\alpha} \log q_k - \sum_{i,k,\alpha} \lambda_{i,k,\alpha} (\log \sigma_k)_{i,k,\alpha} \tag{129}$$

$$= \sum_k p_k \log q_k - Tr(\rho_{k,\alpha} \log(\sigma_{k,\alpha}) \tag{130}$$

$$= -H_Y(X; \rho; \sigma) + S_{(X,Y)}(\rho; \sigma). \tag{131}$$

As a corollary, with the argument proving the Lemma 5 from the Lemma 4, we obtain that the classical Kullback divergence is minus the co-boundary of the 0-cochain defined by the quantum divergence.

This shows that the generating function of all the co-cycles we have considered so far is the quantum 0-cochain for pairs $S(\rho; \sigma) = -Tr(\rho \log \sigma)$.

6. Structure of Observation of a Finite System

Up to now the considered structures and the interventions of entropy can be considered as forming a kind of statics in information theory. The aim of this section is to indicate the elements of dynamics which could correspond. This more dynamical study could be more adapted to the known intervention of entropy in the theory of dynamical systems, as defined by Kolmogorov and Sinai.

6.1. **Problems of Discrimination.**
The problem of *optimal discrimination* consists in separating the various states of a system, by using in the most economical manner, a family of observable quantities. One can also only want to detect a state satisfying a certain chosen property. A possible measure of the cost of discrimination is the number of step before ending the process.

First, let us define more precisely what we mean by a system, a state, an observable quantity and a strategy for using observations. As before, for simplicity, the setting is finite sets.

The symbol $[n]$ denotes the set $\{1, ..., n\}$. We have n finite sets M_i of respective cardinalities m_i, and we consider the set M of sequences $x_1, ..., x_n$ where x_i belongs to M_i; by definition a *system* is a subset X of M and a *state* of the system is an element of X. The set of (classical) observable quantities is a (finite) subset A of the functions from X to $\mathbb{R}$.

A use of observables, named an *observation strategy*, is an oriented tree Γ, starting at its root, that is the smallest vertex, and such that each vertex is labelled by an element of A, and each arrow (naturally oriented edge) is labelled by a possible value of the observable at the initial vertex of the arrow.

For instance, if F_0 marks the root s_0, it means that we aim to measure $F_0(x)$ for the states; then branches issued from t_0 are indexed by the values v of F_0, and to each branch $F_0 = v$ corresponds a subset X_v of states, giving a partition of X. If $F_{1,v}$ is the observable at the final vertex α_v of the branch $F_0 = v$, the next step in the program is to evaluate $F_{1,v}(x)$ for $x \in X_v$; then branches issued from α_v corresponds to values w of $F_{1,v}$ restricted to X_v, and so on.

For each vertex s in Γ we note $\nu(s)$ the number of edges that are necessary for joining s to the root s_0. The function ν with values in $\mathbb{N}$ is called the *level* in the tree.

It can happen that a set X_v consists of one element only; in this case we decide to extend the tree to the next levels by a branch without bifurcation, for instance by labelling with the same observable and the same value, but it could be any labelling, and its value on X_v. In such a way, each level k gives a well defined partition π_k of X.

The level k also defines a sub-tree Γ_k of Γ, such that its final branches are bearing π_k. This gives a sequence $\pi_0, \pi_1, ..., \pi_l$ of finer and finer partitions of X, *i.e.*, a growing sequence of partitions (if the ordering on partition is the opposite of the sense of arrows in the information category $\Pi(X)$). The tree is said *fully discriminant* if the last partition π_l, which is the finest is made by singletons.

The minimal number of steps that are necessary for separating the elements of X, or more modestly for detecting a certain part of states, can be seen as a measure of complexity of the system with respect to the observations A. A refined measure could take in account the cost of use of a given observable, for instance the difficulty to compute its values.

Standard examples are furnished by weighting problems: in this case the states are mass repartitions in n objects, and allowed observables are weighting, which are functions of the form

$$(132) \qquad F_{I,J}(x) = \sum_{i \in I} x_i - \sum_{j \in J} x_j$$

where I et J are disjoint subsets of $[n]$.

We underline that such a function, which requires the choice of two disjoint subsets in $[n]$, makes use of the definition of M as a set of sequences, not as an abstract finite set.

The kind of problems we can ask in this framework were studied for instance in "*Problemes plaisants et delectables qui se font par les nombres*" from Bachet de Meziriac (1612, 1624) [2].

The starting point of our research in this direction was a particular classical problem signaled to us by Guillaume Marrelec: given n objects $\xi_1, ..., \xi_n$, if we know that m have the same mass and $n - m$ have another common mass, how many measures must be performed, to separate the two groups and decide which is the heavier?

Even for $m = 1$ the solution is interesting, and follows a principle of choice by maximum of entropy. In the present text we only want to describe the general structures in relation to this kind of problem without developing a specific study, in particular we want to show that the co-homological nature of the entropy extends to a more dynamical context of discrimination in time.

Remark 9. The discrimination problem is connected with the *coding* problem. In fact a finite system X (as we defined it just before) is nothing else than a particular set of words of length n, where the letter appearing at place i belongs to an alphabet M_i. Distinguishing between different words with a set A of variables f, is nothing else than rewriting the words x of X with symbols v_f (labelling the image $f(X)$). To determine the most economical manner to do that, consists to find the smallest maximal length l of words in the alphabet $(f, v_f); f \in A, v_f \in f(X)$ translating all the words x in X. This translation, when it is possible, can be read on the branches of a fully discriminating rooted tree, associated to an optimal strategy, of minimal level l. The word that translate x being the sequence $(F_0, v_0), (F_1, v_1), ..., (F_k, v_k)$, $k \leq l$, of the variables put on the vertices along the branch going from 0 to x, and the values of these variables put along the edges of this branch.

6.2. **Observation Trees. Galois Groups and Probability Knowledge.**
More generally, we consider as in the first part (resp. in the second part) a finite set Ω, equipped with a Boolean algebra $\mathcal{B}$ (resp. a finite dimensional complex vector space E equipped with a positive definite hermitian form h_0 and a family of direct decompositions in linear spaces $U\mathbf{B}$). In each situation we have a natural notion of *observable quantity*: in the case of Ω it is a partition Y compatible with $\mathcal{B}$ (*i.e.*, less fine than $\mathcal{B}$) with numbering of the parts by the integers $1, .., k$ if Y has k elements; in the case of E it is a decomposition Y compatible with $U\mathbf{B}$ (*i.e.*, each summand is direct sum of elements of one of the decompositions $u\mathbf{B}$; for $u \in U(h_0)$), with a numbering of the summands by the integers $1, .., k$ if Y has k elements. We also have a notion of *probability*: in the case of (Ω, Y) it is a classical probability law P_Y on the quotient set Ω/Y; in the case of (E, Y) it is a collection of non-negative hermitian forms $h_{Y,i}$ on each summands of Y.

We will consider information structures, denoted by the symbol $\mathbf{S}$, for both cases (which could be distinguished by the typography, $\mathcal{S}$ or $\mathbf{S}$, if necessary): they are categories made by objects that are observables and arrows that are divisions, satisfying the condition that if $X \in \mathbf{S}$ divides Y and Z in $\mathbf{S}$, then the joint (Y, Z) belongs to $\mathbf{S}$.

We will also consider probability families adapted to these information structures; they form a covariant functor $X \mapsto Q_X$ (which can be typographically distinguished in the two cases by $\mathcal{Q}_X$ and $\mathbf{Q}_X$) of direct images. When $\mathcal{S}$ is a classical subcategory of the quantum structure $\mathbf{S}$, we suppose that we have a trace transformation from $\iota^*\mathbf{Q}$ to $\mathcal{Q}$, and if $\mathbf{S}$ and $\mathbf{Q}$ are unitary invariant, we remind that, thanks to the ordering, we have an equivalence of category between $\mathbf{S}^U$ and $\mathcal{S}$, and a compatible morphism from the functional module $\mathcal{F}_{\mathcal{Q}}$ to the functional module $\mathcal{F}_{\mathbf{Q}}$.

Except the new ingredient of orderings, they are familiar objects for our reader. The letter $\mathbf{X}$ will denote both cases Ω and E, then the letters $\mathbf{S}, \mathbf{B}, \mathbf{Q}$ will denote respectively $\mathcal{S}, \mathcal{B}, \mathcal{Q}$ or $\mathbf{S}, U\mathbf{B}, \mathbf{Q}$. Be careful that now all observable quantities are ordered, either partitions, either direct decomposition. We will always assume the compatibility condition between $\mathbf{Q}$ and $\mathbf{S}$, meaning that every conditioning of $P \in \mathbf{Q}$ by an event associated to an element of $\mathbf{S}$ belongs to $\mathbf{Q}$.

In addition we choose a subset A of observables in $\mathbf{S}$, which play the role of allowed elementary observations.

We say that a bijection σ from Ω to itself, measurable for $\mathcal{B}$, *respects* a set of observables $\mathcal{A}$ if for any $Y \in \mathcal{A}$, there exists $Z \in \mathcal{A}$ such that $Y \circ \sigma = Z$. It means

that σ establishes an ordered bijection between the pieces $Y(i)$ and the pieces $Z(i)$, *i.e.*, $x \in Z(i)$ if and only if $\sigma(x) \in Y(i)$. In other words the permutation σ respects $\mathcal{A}$ when the map σ^* which associates the partition $Y \circ \sigma$ to any partition Y, sends $\mathcal{A}$ into $\mathcal{A}$.

In the same way, we say that σ respects a family of probabilities $\mathcal{Q}$ if the associated map σ_* sends an element of $\mathcal{Q}$ to an element of $\mathcal{Q}$.

In the quantum case, with E, h_0 and $U\mathbf{B}$, we do the same by asking in addition that σ is a linear unitary automorphism of E.

Definition 9. If $\mathbf{X}$, $\mathbf{S}$, $\mathbf{Q}$, $\mathbf{B}$ and A are given, the *Galois group* G_0 is the set of permutations of X (resp. linear maps) that respect $\mathbf{S}$, $\mathbf{Q}$, $\mathbf{B}$ and A.

Example 6. Consider the system $\mathbf{X}$ associated to the simple classical weighting problem: states are parameterized by points with coordinates $0, 1$ or -1 in the sphere S^{n-1} of radius 1 in $\mathbb{R}^n$, according to their weights, either normal, heavier or lighter. Thus in this case $\Omega = \mathbf{X}$ possesses $2n$ points. The set A of elementary observables is given by the weighting operations $F_{I,J}$, Equation (132). For $\mathcal{S}$ we take the set $\mathcal{S}(A)$ of all ordered partitions π_k obtained by applications of discrimination trees labelled by A. And we consider only the uniform probability P_0 on X; in $\mathcal{Q}$ this gives the images of this law by the elements of $\mathcal{S}$, and the conditioning by all the events associated to $\mathcal{S}$.

Then the Galois group G_0 is the subgroup $\mathfrak{S}_n \times C_2$ of $\mathfrak{S}_{2n}$ made by the product of the permutation group of n symbols by the group changing the signs of all the x_i for i in $[n]$.

Proof: the elements of $\mathfrak{S}_n$ respect A, and the uniform law. Moreover if σ changes the sign of all the x_i, one can compensate the effect of σ on $F_{I,J}$ by taking $G_{I,J} = F_{J,I}$, *i.e.*, by exchanging the two sides of the balance.

To finish we have to show that permutations of X outside $\mathfrak{S}_n \times C_2$ do not respect A. First, consider a permutation σ that does not respect the indices i. In this case there exists an index $i \in [n]$ such that $\sigma(i^+)$ and $\sigma(i^-)$ are states associated to different coins, for instance $\sigma(i^+) = j^+$ and $\sigma(i^-) = k^+$, with $j \neq k$, or $\sigma(i^+) = j^+$ and $\sigma(i^-) = k^-$, with $j \neq k$. Two cases are possible: these states have the same mass, or they have opposite mass. In both cases let us consider a weighting $F_{j,h}(x) = x_j - x_h$, where $h \neq k$; by applying $\sigma^* F_{j,h}$ to $x = \sigma(i^+)$ we find $+1$ (or -1), and by applying $\sigma^* F_{j,h}$ to $x = \sigma(i^-)$ we find 0. However, this cannot happen for a weighting, because for a weighting, either the change of i^+ into i^- has no effect, either it exchanges the results $+1$ and -1. Finally, consider a permutation σ that respects the indices but exchanges the signs of a subset $I = \{i_1, ..., i_k\}$, with $0 < k < n$. In this case let us consider a weighting $F_{i,j}(x) = x_i - x_j$ with $i \in I$ and $j \in [n] \backslash I$, the function $F_{i,j} \circ \sigma$ takes the value $+1$ for the states i^-, j^-, the value -1 for i^+, j^+ and the value 0 for the other states, which cannot happen for any weighting, because this weighting must involve both i and j, but it cannot be $F_{j,i}(x) = x_j - x_i$, which takes the value -1 for j^-, and it cannot be $F_{i,j}$ which takes the value $+1$ for i^+.

The probability laws we are considering express the beliefs in initial knowledge on the system, in this case it is legitimate to consider that they constrain the initial Galois group G_0. This corresponds to the *Jaynes principle* [28, 29].

We define in this framework the notion of *observation tree* adapted to a given subset A of $\mathbf{S}$: it is a finite oriented rooted tree Γ where each vertex s is labelled

by an observable F_s belonging to A and each arrow α beginning at s is labelled by an element $F_s(i)$ of F_s. *A priori* we introduce as many branches as there exist elements in F_s. The disposition of the arrows in the trigonometric circular order makes that the tree Γ is imbedded in the Euclidian plane up to homotopy.

A *branch* γ in the tree Γ is a sequence $\alpha_1, ..., \alpha_k$ of oriented edges, such that, for each i the initial extremity of α_{i+1} is the terminal extremity of α_i. Then α_{i+1} starts with the label F_i and ends with the label F_{i+1}. We will say that γ starts with the root if the initial extremity of α_1 is the root s_0, with a label F_0.

For any edge α in Γ, there exists a unique branch $\gamma(\alpha)$ starting from the root, and abutting in α. Along this branch, the vertices are decorated with the variables $F_i; i = 0, ..., F_k$ and the edges are decorated with values v_i of these functions; we note

(133) $$S(\alpha) = (F_0, v_0; F_1, v_1; ...; F_{k-1}, v_{k-1}; F_k)$$

By definition, the set $X(\alpha)$ of states which are *compatible with* α is the subset of elements of X such that $F_0(x) = v_0, ..., F_{k-1}(x) = v_{k-1}$.

At any level k the sets $X(\alpha)$ form a partition π_k de X.

Definition 10. We say that an observation tree Γ labelled by A is *allowed by* **S**, if all joint observable along each branch belongs to **S**.
We say simply *allowed* if their is no risk of confusion.

In what follows this restriction is imposed on all considered tree. Of course if we start with the algebra of all ordered partitions this gives no restriction, but this would exclude the quantum case, where the best we can do is to take maximal commutative families.

Definition 11. Let α be an edge of Γ, we note $\mathcal{Q}(\alpha)$ the set of probability laws on $X(\alpha)$ which are obtained by conditioning by the values $v_0, v_1..., v_{k-1}$ of the observables $F_0, F_1, ..., F_{k-1}$ along the branch $\gamma(\alpha)$ starting in the root and ending with α.

Definition 12. The Galois group $G(\alpha)$ is the set of permutations of elements of $X(\alpha)$ that belongs to G_0, preserve all the equations $F_i(x) = v_i$ (resp. all the summands of the orthogonal decomposition F_i labelling the edges) and preserve the sets of probability $Q(\alpha)$ (resp. quantum probabilities).
We consider $G(\alpha)$ as embedded in G_0 by fixing point by point all the elements of X outside $X(\alpha)$.

Remark 10. Let P be a probability law (either classical or quantum) on X, $\Phi = (F_i; i \in I)$ a collection of observables, and $\varphi = (v_i; i \in I)$ a vector of possible values of Φ; the law $P|(\Phi = \varphi)$ obtained by conditioning P by the equations $\Phi(x) = \varphi$, is defined only if the set X_φ of all solutions of the system of equations $\Phi(x) = \varphi$ has a non-zero probability $p_\varphi = P(X_\varphi)$. It can be viewed either as a law on X_φ, or as a law on the whole X by taking the image by the inclusion of X_φ in X.

Definition 13. The edge α is said *Galoisian* if the set of equations and probabilities that are invariant by $G(\alpha)$ coincide respectively with $X(\alpha)$ and $\mathcal{Q}(\alpha)$.
A tree Γ is said *Galoisian* when all its edges are Galoisian.

At each level k we define the group G_k which is the product of the groups $G(\alpha)$ for the free edges at level k; it is a subgroup of G_0 preserving elements by elements the pieces of the partition π_k.

Along the path γ the partition (or decomposition) π_l, $l \leq k$ of X is increasing (finer and finer) and the sequence of groups G_l, $l \leq k$ is decreasing.

Along a branch the sets $X(\alpha)$ are decreasing and the sequence of groups $G_0, G(\alpha_1), ..., G(\alpha_k)$ is decreasing. We propose that the quotient $G(\alpha_{i+1})/G(\alpha_i)$ gives a measure of the *Galoisian information* gained by applying F_i and obtaining the value v_i.

On each set $X(\alpha)$ the images of the elements of the probability family $\mathcal{Q}$ form sets $\mathcal{Q}(\alpha)$ of probabilities on $X(\alpha)$.

Thus also imposed in the group $G(\alpha)$ to preserve the set $\mathcal{Q}(\alpha)$.

Remark 11. In terms of coding, introducing probabilities on the $X(\alpha)$ permits to formulate the principle, that it is more efficient to choose, after the edge α, the observation having the largest conditional entropy in $Q(\alpha)$. In what circumstances it gives the optimal discrimination tree is a difficult problem, even if the folklore admit that as a theorem. It is the problem of optimal coding.

In virtue of a Shannon's theorem, the minimal length is bounded below by entropy of the law on X if this law is unique. We found it works in a simple example of weighting (*cf.* paper 3 [7]).

Note however important differences between our approach and the traditional one for coding: for us A is given and $\mathcal{Q}$ is given; they correspond respectively to an *a priori* limitation of possible codes for use (like a natural language), and to a set of possible *a priori* knowledges, for instance taking in account the Galois ambiguity in the system (Jaynes principle). All that is Bayesian in spirit.

Definition 14. We say that an observation tree Γ labelled by A is *allowed by* **S** *and by* $X \in \mathbf{S}$, if it is allowed by $\mathbf{S}_X$, which means that all joint observable along each branch is divided by X.

Definition 15. $\mathbf{S}(A)$ is the set of (ordered) observables π_k which can be obtained by allowed observation trees. For $X \in \mathbf{S}$ we note $\mathbf{S}_X(A)$ the set of (ordered) observables π_k which can be obtained by observation trees that are allowed by **S** and X.

Lemma 9. The joint product defines a structure of monoid on the set $\mathbf{S}_X(A)$.

Proof. Let Γ, Γ' be two observation trees allowed by A, **S** and $X \in \mathbf{S}$, of respective lengths k, k', giving final decompositions S, S'. To establish the lemma we must show that the joint SS' is obtained by a tree associated with A, allowed by **S** and X.

For that we just graft one exemplar of Γ' on each free edge of Γ. This new tree $\Gamma\Gamma'$ is associated with A, and its final partition is clearly finer than S. It is also finer than S', because at the end of any branch of $\Gamma\Gamma'$ we have an $X(\beta)$ which is contained in the corresponding element of the final partition $\pi_{k'}(\Gamma')$. To finish the proof we have to show that each element of $\pi_{k+k'}(\Gamma\Gamma')$ is the intersection of element of $\pi_k(\Gamma)$ with one element of $\pi_{k'}(\Gamma')$, because we know these observables are in $\mathcal{S}_X$, which is a monoid, by the definition of information structure. But a complete branch $\gamma.\gamma'$ in $\Gamma\Gamma'$, going from the root to a terminal edge at level $k + k'$, corresponds to a word $(F_0, v_0, F_1, v_1, ..., F_{k-1}, v_{k-1}, F'_0, v'_0, ..., F'_{k'-1}, v'_{k'-1})$, thus the final set of the branch $\gamma.\gamma'$ is defined by the equations $F_i = v_i; i = 0, ..., k - 1$ et $F'_j = v'_j; j = 0, ..., k' - 1$, and is the intersection of the sets respectively defined by the first and second groups of equations, that belong respectively to $\pi_k(\Gamma)$ and $\pi_{k'}(\Gamma')$.

Then $\mathbf{S}(A)$ form an information structure. In particular there is a unique maximal partition, initial element for each subcategory $\mathbf{S}_X(A)$ in the information structure $\mathbf{S}(A)$.

But on $\mathbf{S}(A)$ the operation of *grafting*, that we will describe now, is much richer than what we used in the above Lemma 9: we can graft an allowed tree on each free edge of an allowed tree, and this introduces to a theory of operads and monads for information theory.

6.3. **Co-Homology of Observation Strategies.**

Remember that the elements of the partitions or decompositions Y we are considering, are now numbered by the ordered set $\{1, ..., L(Y)\}$, where $L(Y)$ is the number of elements in the partition, or the decomposition, also called its length. In particular we consider as different two partitions which are labelled differently by the integers. This was already taken into account in the definition of the Galois groups.

We define the *multi-products* $\mu(m; n_1, ..., n_m)$ on the set of ordered partitions:

They are defined between a partition equipped with an ordering (π, ω) with m pieces and m ordered partitions $(\pi_1, \omega_1), ..., (\pi_m, \omega_m)$ of respective lengths $n_1, ..., n_m$; the results is the ordered partition obtained by cutting each piece X_i of π by the corresponding decomposition π_i and renumbering the non-empty pieces by integers in the unique way compatible with the orderings $\omega, \omega_1, ..., \omega_m$. Observe the important fact that the result has in general less than $n = n_1 + ... + n_m$ pieces. This introduces a strong departure from usual multi-products (*cf.* P. May [16, 42], Loday-Vallette [37]). We do not have an *operad*, when introducing vector spaces $V(m)$ generated by decompositions of length m, we get filtered but not graded structures. However a form of associativity and neutral element are preserved, hence we propose to name this structure a *filtered operads*.

There exists an evident unit to the right which is the unique decomposition of length 1.

The action of the symmetric group $\mathfrak{S}_m$ on the products is evident, and does not respect the length of the result. We will designate by μ_m the collection of products for the same length m.

The numbers m_i between 1 and n_i that counts the pieces of the decomposition of the element X_i of π are functions $m_i(\pi, \omega, \pi_i, \omega_i)$. There exists a growing injection $\eta_i : [m_i] \to [n_i]$, which depends only on $(\pi, \omega, \pi_i, \omega_i)$ telling what indices of (π_i, ω_i) survive in the product. These injections are integral parts of the structure of filtered operad. In particular, if we apply a permutation σ_i to $[n_i]$, *i.e.*, if we replace ω_i by $\omega_i \circ \sigma_i$, the number can change.

The axioms of operadic unity and associativity, conveniently modified are easy to verify (*cf.* [7]). The reference we follow here is Fresse "Basic concepts of operads" [20]. For unity nothing has to be modified. For associativity (Figure 1.3 in Fresse [20]), we modify by saying that if the (π_i, ω_i) of lengths n_i, for i between 1 et k, are composed from $\mu(n_i; n_i^1, ..., n_i^{n_i})$ with the n_i-uples $(..., (\pi_i^j, \omega_i^j), ...)$ whose respective lengths are n_i^j, and if the result μ_i for each i has length $(m_i^1 + ... + m_i^{n_i})$ where m_i^j is function of (π_i, ω_i) and (π_i^j, ω_i^j), then the product of (π, ω) of length k with the μ_i is the same as the one we would have obtained by composing $\mu(k; n_1, ..., n_k)((\pi, \omega); (\pi_1, \omega_1), ...))$ with the $m = m_1 + ... + m_k$ ordered decompositions (π_i^j, ω_i^j) for j belonging to the image of $\eta_i : [m_i] \to [n_i]$. This result is

more complicate to write than to prove, because it only expresses the associativity of the ordinary join of three partitions; from which ordering follows.

Moreover, the first axiom concerning permutations (Figure 1.1 in Fresse [20]), can be modified, by considering only permutations of n_i letters which preserve the images of the maps η_i.

The second axiom, which concerns a permutation σ of k elements in π, and the inverse permutation of the partitions π_i can be reformulated by telling the effect of σ on the multiple product μ is the same as the effect of σ on the indices of the (π_i, ω_i). In other terms, the effect of σ on ω is compensated by the action of σ^{-1} on the indices of the (π_i, ω_i). One has to be careful, because the result of μ applied to $(\pi, \omega \circ \sigma)$ has in general not the same length as μ applied to (π, ω). However the compensation implies that μ_k is well defined on the quotient of the set of sequences $((\pi, \omega), (\pi_1, \omega_1), ...)$ by the diagonal action of $\mathfrak{S}_k$, which permutes the k pieces of π and which permutes the indices i of the n_i in the other factors.

Geometrically, if the partition (π, ω) in $\mathbf{S}(A)$ is generated by an observation tree Γ with m ending edges and the partitions $(\pi_i, \omega_i); i = 1, ..., m$ are generated by a collection of observation trees Γ_i; then the result of the application of $\mu(m; n_1, ..., n_m)$ to (π, ω) and $(\pi_i, \omega_i); i = 1, ..., m$ is generated by the observation tree that is obtained by grafting each Γ_i on the vertex number i. Drawing the planar trees associated to three successive sets of decompositions for two successive grafting operations helps to understand the associativity property.

The fact that in general this does not give a tree with $n_1 + ... + n_m$ free edges, where n_i denotes the number of free edges of Γ_i comes from the possibility to find an empty set $X(\beta)$ at some moment along a branch of the grafted tree; this we call a *dead branch*. It expresses the fact that the empty set is excluded from the elements of a partition in the classical context, and the zero space excluded from the orthogonal decomposition in the quantum context. When computing conditioned probabilities we encounter the same problem if a set $X(\beta)$ at some place in a branch has measure zero.

The dead branches and the lack of graduation cause a lot of difficulties for studying algebraically the operations μ_m, thus we introduce more flexible objects, which are the *ordered partitions with empty parts* of Ω, resp. *ordered orthogonal decompositions with zero summands* of E: such a partition π^* (resp. decomposition) is a family $(E_1, ..., E_m)$ of disjoint subsets of Ω (resp. orthogonal subspaces of E), such that their union (resp. sum) is Ω (resp. E). The only difference with respect to ordered partitions, resp. decompositions, is that we accept to repeat $\emptyset$ (resp. 0) an arbitrary high number of times. For shortening we will name *generalized decompositions* these new objects. The number m is named the degree of π^*. These objects are the natural results of applying rooted observation trees embedded in an oriented half plane.

The notions of adaptation to A, $\mathbf{S}$ and X in $\mathbf{S}$ concerning the trees, apply to the generated generalized decompositions. The corresponding sets of generalized objets are written $\mathbf{S}^*(A)$ and $\mathbf{S}^*_X(A)$.

The multi-product $\mu(m; n_1, ..., n_m)$ extends naturally to generalized decompositions, and in this case the degrees are respected, *i.e.*, the result of this operation is a generalized decomposition of degree $n_1 + n_2 + ... + n_m$.

Remark that we could write $\mu^*(m; n_1, ..., n_m)$ for the multi-products extended to generalized decompositions, however we prefer to keep the same notation $\mu(m; n_1, ..., n_m)$;

this is justified by the following observation: to a generalized decomposition π^* is associated a unique ordered decomposition (π, ω), by forgetting the empty sets (resp. zero spaces) in the family, and the multi-product is compatible with this forgetting application. The gain of the extension is the easy construction of a monad we expose now.

The definition of operad was introduced by P. May [42] as the right tool for studying the homology of infinite loop spaces; then it was recognized as a fundamental tool for algebraic topology, and many other topics, see Loday and Valette, Fresse.

We will encounter only "symmetric" operads.

The multiple products μ_m on generalized decompositions can be assembled in a structure of *monad* by using the standard Schur construction (*cf.* Loday et Valette [37], or Fresse, "on partitions" [20]): For each $X \in \mathbf{S}$, we introduce the real vector space $V_X = V_X(A)$ freely generated by the set $\mathbf{S}_X^*(A)$ of generalized decompositions obtained by observation trees that are allowed by A, $\mathbf{S}$ and X; the length m define a graduation $V_X(m)$ of V_X. We put $V_X(0) = 0$.

The maps μ_m generate m-linear applications from products of these spaces to themselves which respect the graduation; these applications, also denoted by μ_m, are parameterized by the sets $\mathbf{S}_X^*(m)$, whose elements are the generalized decompositions of degree m which are divided by X:

$$(134) \qquad \mu_m : V_X(m) \otimes_{\mathfrak{S}_m} V_X^{\otimes m} \to V_X$$

The linear *Schur functor* from the category of real vector spaces to itself, is defined by the direct sum of symmetric co-invariants:

$$(135) \qquad \mathcal{V}_X(W) = \bigoplus_{m \geq 0} V_X(m) \otimes_{\mathfrak{S}_m} W^{\otimes m}$$

The composition of Schur functors is defined by

$$(136) \qquad \mathcal{V}_X \circ \mathcal{V}_X = \bigoplus_{m \geq 0} V_X(m) \otimes_{\mathfrak{S}_m} \mathcal{V}_X^{\otimes m}.$$

i.e., for each real vector space W:

$$(137) \quad \mathcal{V}_X \circ \mathcal{V}_X(W) = \bigoplus_{m \geq 0} \bigoplus_{l \geq m} \bigoplus_{n_1, \ldots, n_m; \sum_i n_i = l} V_X(m) \otimes_{\mathfrak{S}_m} \bigotimes_i V_X(n_i) \otimes_{\mathfrak{S}_{n_i}} W^{\otimes n_i}$$

$$(138) \qquad = \bigoplus_{l \geq 0} \bigoplus_{m \geq 0} \bigoplus_{n_1, \ldots, n_m; \sum_i n_i = l} V_X(m) \otimes_{\mathfrak{S}_m} \bigotimes_i V_X(n_i) \otimes_{\mathfrak{S}_{n_1, \ldots, n_k}} W^{\otimes l};$$

where $\mathfrak{S}_{n_1, \ldots, n_m}$ denotes the groups of permutations by blocs.

Proposition 5. For each X in $\mathbf{S}$, the collection of operations μ_m defines a linear natural transformation of functors $\mu_X : \mathcal{V}_X \circ \mathcal{V}_X \to \mathcal{V}_X$; and the trivial partition defines a linear natural transformation of functors $\eta_X : \mathbb{R} \to \mathcal{V}_X$, which satisfy the axioms of a *monad* (cf. MacLane "Categories for Working Mathematician" 2nd ed. [39], and Alain Proute, Introduction a la Logique Categorique, 2013, Prepublications [47]):

$$(139) \qquad \mu_X \circ (\mathcal{V}_X \mu_X) = \mu_X \circ (\mu_X \mathcal{V}_X), \quad \mu_X \circ (\mathcal{V}_X \eta_X) = Id = \mu_X \circ (\eta_X \mathcal{V}_X)$$

Proof. The argument is the same as the argument given in Fresse (partitions ...). The fact that the natural transformation μ_X is well defined on the quotient by the diagonal action of the symmetric group $\mathfrak{S}_m$ on $V_X(m) \otimes \bigotimes_i V_X(n_i) \otimes_{\mathfrak{S}_{n_1, \ldots, n_m}}$

$W^{\otimes s}$ comes from the verification of the symmetry axiom and the properties of associativity and neutral element comes from the verification of the corresponding axiom.

Moreover all these operations are natural for the functor of inclusion from the category $\mathbf{S}_Y$ to the category $\mathbf{S}_X$ of observables divided by Y and X respectively when X divides Y; therefore we have the following result:

Proposition 6. To each arrow $X \to Y$ in the category $\mathbf{S}$ is associated a natural transformation of functors $\rho_{X,Y} : \mathcal{V}_Y \to \mathcal{V}_X$, making a morphism of monads; this defines a contravariant functor $\mathcal{V}$ from the category $\mathbf{S}$ to the category of monads, that we name the arborescent structural sheaf of $\mathbf{S}$ and A.

Considering the discrete topology on $\mathbf{S}$, we introduce the topos of sheaves of modules over the functor in monads $\mathcal{V}$, which we call the *arborescent information topos* associated to $\mathbf{S}$ and A.

As explained in Proute *loc.cit.* [47] a monad in a category $\mathcal{C}$ becomes a monoid in the category of endo-functors of $\mathcal{C}$, thus the topos we introduce is equivalent to an ordinary ringed topos.

The monad $\mathcal{V}_X$, and the contravariant monadic functor $\mathcal{V}$ on $\mathbf{S}$, are better understood by considering trees, *cf.* Getzler-Jones [22], Ginzburg-Kapranov [23] and Fresse [20]; in our context we consider all observation trees labelled by elements of $\mathbf{S}_X^* A$:
if Γ is an oriented rooted tree of level k, each vertex v of Γ gives birth to m_v edges; we define

$$(140) \qquad V_X(\Gamma)(W) = \bigotimes_{v \in \Gamma} V_X(m_v) \otimes_{\mathfrak{S}_{m_v}} W^{\otimes m_v}.$$

The space $V(\Gamma)(W)$ is the direct sum of spaces $V_X(\Gamma_Y)(W)$ associated to trees which are decorated by a subset Y in $\mathbf{S}_X^*(A)$, with one element Y_v of $\mathbf{S}_X(m)$ for each vertex v which gives birth to m_v edges. Then the iterated functors $\mathcal{V}^{\circ k} = \mathcal{V} \circ ... \circ \mathcal{V}$ for $k \geq 1$ are the direct sums of the functors $V(\Gamma)$ of level k.
Remark that we could have worked directly with observation trees labelled by elements of A in spite of working with generalized partitions; this would have given a strictly larger monad but equivalent results.

Associated to probability families we define now a right $\mathcal{V}_X$-module (in the terms of Fresse, Partitions, the term $\mathcal{V}_X$-algebra being reserved to a structure of left module on a constant functor).

For that we introduce the notion of *divided probability*.

Definition 16. A divided probability law of degree m is a sequence of triplets $(p, P, U) = (p_1, P_1, U_1; ...; p_m, P_m, U_m)$, where $p_i; i = 1, ..., m$ are positive numbers of sum one, *i.e.*, $p_1 + ... + p_m = 1$, where each $P_i; i = 1, ..., m$ is a classical (resp. quantum) probability law when the corresponding p_i is strictly positive, and a probability law or the empty set when the corresponding p_i is equal to 0, and where each $U_i; i = 1, ..., m$ is the support in $\mathbf{X}$ of P_i; moreover the U_i are assumed to be orthogonal (resp. disjoint in the classical case). The letter P will designate the probability $p_1 P_1 + ... + p_m P_m$, where $0.\emptyset = 0$ when it happens.

The symbol $\mathcal{D}(m)$ designates the set of divided probabilities of degree m on $\mathbf{X}$, and $\mathcal{D}_X(m)$ denotes the subset made with probability laws in $\mathbf{Q}_X$ adapted to a

variable X.

The vector space generated by $\mathcal{D}_X(m)$ will be written $\mathcal{L}_X(m)$. We put $\mathcal{L}_X(0) = 0$.

We also introduce the subspace $\mathcal{K}(m)$ of $\mathcal{L}_X(m)$ which is generated by two families of vectors in $\mathcal{L}_X(m)$:

First the vectors

$$(141) \quad L(\lambda, p', p'', P, U) = \lambda(p', P, U) + (1 - \lambda)(p'', P, U) - (\lambda p' + (1 - \lambda)p'', P, U),$$

where λ is any real number between 0 and 1, and (p', P, U), (p'', P, U) two divided probabilities associated to the same sequence of probability laws $(P_1, ..., P_m)$ and the same supports $(U_1, ..., U_m)$;

Second the vectors

$$(142) \qquad\qquad D(p, P, U, Q, V) = (p, P, U) - (p, P', U'),$$

where for each index i between 1 and m, such that $p_i > 0$ we have $P_i = P'_i$, and consequently $U_i = U'_i$.

The we define the space of classes of divided probabilities as the quotient real vector space $\mathcal{M}_X(m) = \mathcal{L}_X(m)/\mathcal{K}(m)$. In particular $M_X(0) = 0$, $M_X(1)$ is freely generated over $\mathbb{R}$ by the elements of $\mathbf{Q}_X$.

Lemma 10. The space $\mathcal{M}_X(m)$ is freely generated over $\mathbb{R}$ by the vectors $(\emptyset, ..., \emptyset, P_i, \emptyset, ..., \emptyset)$ of length m, where at the rank i, P_i is an element of $\mathbf{Q}_X$.

Proof. Let $D = (p_1, P_1, U_1), ..., (p_m, P_m, U_m)$ be a divided probability; we consider for each i between 1 and m the divided probability

$$D_i = (0, P_1, U_1), ..., (0, P_{i-1}, U_{i-1}), (1, P_i, U_i), (0, P_{i+1}, U_{i+1}), ..., (0, P_m, U_m),$$

then the vector $D - \sum_i p_i D_i$ is a sum of vectors of type L in $\mathcal{K}_X(m)$. However, for each i, the vector $D_i - (\emptyset, ..., \emptyset, P_i, \emptyset, ..., \emptyset)$ is of type D, thus the particular vectors of the Lemma 10 generate $\mathcal{M}_X(m)$.

Now, we prove that, if a linear combination of r of these vectors belongs to $\mathcal{K}_X$, the coefficients of this combination must all be equal to 0. We proceed by recurrence on r, the result being evident for $r = 1$. We also can suppose that at least two involved vectors have a non-empty element at the same place, which we can suppose to be $i = 1$. All vectors with $p_1 = 0$ can be replaced by a vector where $P_1 = \emptyset$ using an element of type D in $\mathcal{K}_X(m)$, then we can assume that at least one of the vectors has a p_1 strictly positive, *i.e.*, equals to 1. Let us consider all these vectors $D_1, ..., D_s$, for $2 \leq s \leq r$, their other numbers p_i for $i > 1$ are zero. The other vectors D_j, for $j > s$ having the coordinate p_1 equal to zero. Let $\sum_j \lambda_j D_j$ be the linear combination of length r belonging to $\mathcal{K}_X(m)$; this vector is a linear combination of vectors of type L and D. We can suppose that every λ_j is non-zero. Let us consider an element Q of $\mathbf{Q}_X$ which appears in at least one of the D_j, $j \leq s$; this Q cannot appear in only one D_j, because the sum of coefficients λ multiplied by the first p_1 in front of any given Q in a vector L or D is zero. Thus we have at least two D_j with the same P_1. We can replace the sum of them with λ_j positive (resp. negative) by only one special vector of the Lemma 10 using a sum of multiples of vectors of type L. Then we are left with the case of two vectors, D_1, D_2 having $P_1 = Q$ such that $\lambda_1 + \lambda_2 = 0$, which means that $\lambda_1 D_1 + \lambda_2 D_2$ is multiple of a vector of type D. Subtracting it we can apply the recurrence hypothesis and conclude that the considered linear relation is trivial.

As a corollary an equivalent definition of the spaces $\mathcal{M}_X(m)$ would be the real vector space freely generated by pairs (P, i) where $P \in \mathbf{Q}_X$ and $i \in [m]$. Such a vector, identified with $(\emptyset, .., P, ..., \emptyset)$ in $\mathcal{L}_X(m)$, where only the place i is non-empty, will be named a *simple vector* of degree m.

Let $S = (S_1, ..., S_m)$ be a sequence of generalized decompositions in $\mathbf{S}_X^*(A)$, of respective degrees $n_1, ..., n_m$, with $n = n_1 + ... + n_m$, and let (p, P, U) be an element of $\mathcal{D}_X(m)$, we define $\theta((p, P, U), S)$ as the following divided probability of degree n: if, for $i = 1, ..., m$ the decomposition S_i is made of pieces $E_i^{j_i}$ where j_i varies between 1 and n_i, we take for $p_i^{j_i}$ is the classical probability $\mathbb{P}(E_i^{j_i} \cap U_i)$; we take for $P_i^{j_i}$ the law P_i conditioned by the event $S_i = j_i$ which corresponds to $E_i^{j_i}$; and we take for $U_i^{j_i}$ the support of $P_i^{j_i}$. Then we order the obtained family of triples $(p_i^{j_i}, P_i^{j_i}, U_i^{j_i})_{i=1,...,m;j_i=1,...,n_i}$ by the lexicographic ordering. It is easy to verify that the resulting sequence is a divided probability.

Extending by linearity we get a linear map,

(143) $$\lambda_m : \mathcal{L}_X(m) \otimes V_X(n_1) \otimes ... \otimes V_X(n_m) \to \mathcal{L}_X(n_1 + ... n_m),$$

By linearity a vector of type L in $\mathcal{L}_X(m)$, tensorized with $S_1 \otimes ... \otimes S_m$ goes to a linear combination of vectors of type L in $\mathcal{L}_X(n)$. Moreover, if $p_i = 0$ for an index i in $[m]$, all the $p_i^{j_i}$ are zero, thus a vector of type D goes to a vector of type D. Then the map λ_m sends the subspace $\mathcal{K}_X(m) \otimes V_X(n_1) \otimes ... \otimes V_X(n_m)$ into the subspace $\mathcal{K}_X(n_1 + ... n_m)$, thus it defines a linear map

(144) $$\theta_m : \mathcal{M}_X(m) \otimes V_X(n_1) \otimes ... \otimes V_X(n_m) \to \mathcal{M}_X(n_1 + ... n_m),$$

On a simple vector (P, i), the operation θ_m is independent of the S_j for $i \neq i$.

Now we introduce the Schur functor $\mathcal{M}_X$ of symmetric co-invariant spaces $\mathcal{M}_X(W) = \bigoplus_m \mathcal{M}_X(m) \otimes_{\mathfrak{S}_m} W^{\otimes m}$ from the category of real vector space to itself, associated to the $\mathfrak{S}$-module $\mathcal{M}_X^*$ (*cf.* Loday and Valette [37], Fresse [20]), formed by the graded family $\mathcal{M}_X(m); m \in \mathbb{N}$.

Then the maps θ_m define a natural transformation of functors:

(145) $$\theta_X : \mathcal{M}_X \circ \mathcal{V} \to \mathcal{M}_X.$$

In addition, this set of transformations behaves naturally with respect to X in the information category $\mathbf{S}$. Note that it defines a co-variant functor, not a presheaf.

For simplicity, we will note in general $\theta, \mu, \mathcal{F}, \mathcal{V}, ...$ and not $\theta_X, \mu_X, \mathcal{F}_X, \mathcal{V}_X, ...,$ but we memorize this is an abuse of language.

Then the composite functor $\mathcal{M} \circ \mathcal{V}(W)$ is given by

$$\mathcal{M}_X \circ \mathcal{V}_X(W) = \bigoplus_{m \geq 0} \mathcal{M}_X(m) \otimes_{\mathfrak{S}_m} \bigotimes_i (V_X(n_i) \otimes_{\mathfrak{S}_{n_i}} W^{\otimes n_i})$$

$$= \bigoplus_{n \geq 0} \bigoplus_{m \geq 0} \bigoplus_{n_1, ..., n_m; \sum_i n_i = n} \mathcal{M}_X(m) \otimes_{\mathfrak{S}_m} \bigotimes_i V_X(n_i) \otimes_{\mathfrak{S}_{n_1, ..., n_k}} W^{\otimes n};$$

where $\mathfrak{S}_{n_1, ..., n_m}$ denotes the groups of permutations by blocs.

Proposition 7. The natural transformation θ defines a right action in the sense of monads, i.e., we have

(146) $$\theta \circ (\mathcal{F}\mu) = \theta \circ (\theta\mathcal{V}); \quad \theta \circ (\mathcal{F}\eta) = Id.$$

Proof. The proof is the same as for proposition 5, by using the associativity of conditioning, and the Bayes identity $P(A \cap B) = P(A|B)P(B)$.

Ginzburg and Kapranov [23] gave a construction of the (co)bar complex of an operad based on decorated trees. It is a graded complex of operads, with a differential operator of degree -1. The dual construction can be found in Getzler et Jones [22]; it gives a graded complex of co-operads with a differential operator of degree $+1$. The link with quasi-free co-operads and operads (Quillen's construction) is developed by Fresse (in "partitions" [20]); in this article Fresse also shows that these constructions correspond to the simplicial bar construction for the monads (Maclane) and to the natural notions of derived functors in this context.

In our case, with two right modules, the easiest way is to use the bar construction of Beck (1967) [9], further explicited by Fresse with decorated trees in the case of monads coming from operads.

A *morphism* from a right module $\mathcal{M}$ over $\mathcal{V}$ to a right module $\mathcal{R}$ over $\mathcal{V}$ is a natural transformation f of the first functor in the second such that $f \circ \theta_M = \theta_R \circ f \mathcal{V}$.

In what follows we will use the module $\mathcal{R}$ which comes from the functor of symmetric powers:

$$(147) \qquad \mathcal{R}(W) = \bigoplus_m S^m(W);$$

it is the Schur functor associated to the trivial $\mathfrak{S}_*$-module, $\mathcal{R}(m) = \mathbb{R}$, *i.e.*, the action of $\mathfrak{S}_m$ on $\mathcal{R}(m)$ is trivial. We put $\mathcal{R}(0) = \mathbb{R}$.

The right action of $\mathcal{V}_X$ is given by the map

$$(148) \qquad \rho_m : \mathcal{R}_X(m) \otimes V_X(n_1) \otimes ... \otimes V_X(n_m) \to \mathcal{R}_X(n_1 + ...n_m),$$

which send each generator $(1, S_1, ..., S_m)$ to 1 in $\mathcal{R}(n) = \mathbb{R}$.

The axioms of a right module are easy to verify.

This $\mathcal{V}$-module $\mathcal{R}$ will play the dual role of the trivial module in the case of information structure co-homology.

Following Beck (Triples, Algebras, Cohomology, 1967, 2002 [9]), we consider the simplicial bar complex $\mathcal{M}_X \circ \mathcal{V}_X^*$ extending the right module $\mathcal{M}$ on $\mathcal{V}$ by the sequence of modules $.... \to \mathcal{M}_X \circ \mathcal{V}_X^{\circ(k+1)} \to \mathcal{M}_X \circ \mathcal{V}_X^{\circ k} \to ....$ Then we introduce the growing complex $C^*(\mathcal{M}_X)$ of *measurable* morphisms from $\mathcal{M}_X \circ \mathcal{V}_X^*$ to the symmetric right module $\mathbf{R}$.

For a given $k \geq 0$, a morphism F from $\mathcal{M}_X \circ \mathcal{V}_X^{\circ k}$ to $\mathcal{R}$ is defined by a family of maps $F(N) : \mathcal{M}_X \circ \mathcal{V}_X^{\circ k}(N) \to \mathcal{R}(N) = \mathbb{R}$, for $N \in \mathbb{N}$.

This gives a family of measurable numerical functions of a divided probability law (p, P, U), of degree $m \leq N$, indexed by forests having m components trees of height k and having total number of ending branches N.

We denote such a family of functions by the symbol $F_X(S_1; S_2; ...; S_k; (p, P, U))$, indexed by X in $\mathbf{S}$, where $S_1; ...; S_k$ here designates the sets of decompositions present in the trees at each level from 1 to k.

First we remark that the compatibility with the action of $\mathcal{V}_X$ to the right imposes that for any allowed set of variables S_{k+1} we must have

$$(149) \qquad F_X(S_1; S_2; ...; \mu(S_k, S_{k+1}); (p, P, U)) = F_X(S_1; S_2; ...; S_k; (p, P, U)).$$

By taking for S_k the collection $(\pi_0, ..., \pi_0)$, we deduce that F_X is independent of the last variable.

This has the effect of decreasing the degree in k by one, for respecting the preceding conventions on information cochains; *i.e.*, we pose $\mathcal{C}^k(M_X) = Hom(\mathcal{M}_X \circ \mathcal{V}^{\circ(k+1)}, \mathcal{R})$.

Secondly, as we are working with the quotient of the space generated by divided probabilities (p, P, U) by the space generated by linearity relations on the external law p, for (p, P, U) of degree m, we have

$$(150) \qquad F_X(S_1; S_2; ...; S_k; (p, P, U)) = \sum_{i=1}^{m} p_i F_X(S_1; S_2; ...; S_k; (P_i; i, m));$$

where $(Q; i, m)$ designates the divided probability of degree m where all the laws in the sequence are empty except for the number i where it is equal to Q.

Moreover, from the definition of θ and the rule of composition of functors, for any $m \geq 1$ and $i \in [m]$, and any simple vector (Q, i, m), the value of F on any forest depends only on the tree component of index i; that we can summarize by the following identity:

$$(151) \qquad F_X(S_1; S_2; ...; S_k; (Q; i, m)) = F_X(T(S_1^i; S_2^i; ...; S_k^i); (Q; i, m));$$

where $T(S_1^i; S_2^i; ...; S_k^i)$ designates the tree numbered by i, prolonged in any manner at all the places $j \neq i$.

Definition 17. An element of $\mathcal{C}^k(M_X)$ is said *regular* when for each degree m and each index i between 1 and m, we have, for each ordered forest $S_1; S_2; ...; S_k$ of m trees, and each probability Q,

$$(152) \qquad F_X(S_1; S_2; ...; S_k; (Q; i, m)) = F_X(S_1^i; S_2^i; ...; S_k^i; Q);$$

where $S_1^i; S_2^i; ...; S_k^i$ designates the tree number i.

Due to Equation (150), this makes that regular elements are defined by their values on trees and ordinary, not divided probabilities.

The adjective regular can be better interpreted as "local in the sense of observation trees".

The vector space $\mathcal{C}_X^k(N)$ is generated by families of functions of divided probabilities $F_X(S_1; S_2; ...; S_k; (p, P, U))$, indexed by X in $\mathbf{S}$ and forests $S_1; ...; S_k$ of level k. These families are supposed *local* with respect to X, which means that it is compatible with direct image of probabilities under observables in $\mathbf{S}^*$.

Remark 12. As we showed in the static case, in the classical context, locality is equivalent to the fact that the values of the functions depend on $\mathbb{P}$ through the direct images of $\mathbb{P}$ by the joint of all the ordered observables which decorate the tree (the joint of the joints along branches); but this is not necessarily true in the quantum context, where it depends on $\mathbf{Q}$. However it is true for $\mathbf{Q}^{min}$, in particular $\mathbf{Q}^{can}$ which is the most natural choice.

The spaces $\mathcal{C}^k(\mathcal{M}_X)$ form a natural degree one complex:

The faces $\delta_i^{(k)}; 1 \leq i \geq k$ are given by applying μ on $\mathcal{V} \circ \mathcal{V}$ at the places $(i, i+1)$; the last face $\delta_{k+1}^{(k)}; 1 \leq i \geq k$ consists in forgetting the last functor, the operation denoted by ϵ; and the zero face is given by the action θ. Then the boundary $\delta^{(k)}$ is the alternate sum of the operators $\delta_i^{(k)}; 0 \leq i \geq k+1$: if F is measurable morphism from $\mathcal{M} \circ \mathcal{V}^{\circ k}$ to $\mathbb{R}$, then

$$(153) \quad \delta F = F \circ (\theta \mathcal{V}^{\circ k}) - \sum_{i=0,...,k-1} (-1)^i F \circ \mathcal{M} \mathcal{V}^{\circ i} \mu \mathcal{V}^{\circ k-i-1} - (-1)^k F \circ \mathcal{M} \mathcal{V}^{\circ k} \epsilon.$$

The zero face in the complex $\mathcal{C}_X^*$ corresponds to the right action of the monad V_X on divided probabilities; on regular cochains it is expressed by a generalization

of the formula (20): if (P, i, m) is a simple vector of degree m and $S_0; S_1; ...; S_k$ a forest of level $k + 1$, with m component trees, then

$$
\begin{aligned}
F_{S_0}(S_1; ...; S_k; (P, i, m)) &= F(S_1; ...; S_k; \theta((P, i, m)S_0)) \\
(154) &= \sum_{j_i=1,...,n_i} \mathbb{P}(S_0^i = j_i) F((S_1^{j_i}; S_2^{j_i}; ...; S_2^{j_i}; (P|(S_0^i = j_i)),
\end{aligned}
$$

where $S_1^{j_i}; S_2^{j_i}; ...; S_k^{j_i}$ designates the tree number j_i grafted on the branch j_i of the variable $S_{0,i}$ at the place i in the collection S_0.

The formula (154) is compatible with the existence of dead branches.

Note that natural integers come into the play under two different aspects: m is for the internal monadic degree and counts the number of components, or the length of partitions, k is for the height of the trees in the forest. The number k gives the degree in co-homology.

The *coboundary* δ of $\mathcal{C}^*$ is of degree $+1$ with respect to k and degree 0 with respect to m. For any $m \in \mathbb{N}$, the operator δ has the formula of the coboundary given by the simplicial structure associated to θ and μ:

$$
\begin{aligned}
\delta F(S_0; S_1; ...; S_k; (p, P, U)) &= F_{S_0}(S_1; ...; S_k; (p, P, U)) \\
&\quad + \sum_{i=1}^{i=k} (-1)^i F(S_0; ...; \mu(S_{i-1} \otimes S_i); S_{i+1}; ...; S_k; (p, P, U)) \\
&\quad + (-1)^{k+1} F(S_0; ...; S_{k-1}; (p, P, U))
\end{aligned}
$$
(155)

We constat that locality is preserved by δ.

Lemma 11. If the transformation F is regular, then δF is regular; in other terms, the regular elements form a sub-complex $\mathcal{C}_r^k(\mathcal{M}_X)$.

Proof. Let (P, i, m) be a simple vector and $S_0; ...; S_k$ a forest with m components; let us denote by S_0^j the variable number j having degree n_j, and $n = n_1 + ... + n_m$; we have

$$
\begin{aligned}
&\delta F(S_0; ...; S_k; (P, i, m)) \\
(156) \quad &F(S_1; ...; S_k; \theta((P, i, m)S_0^i)) - F(\mu(S_0, S_1); ...; S_k; (P, i, m)) - ... \\
&+ (-1)^k F(S_0; ...; \mu(S_{k-1}, S_k); (P, i, m)) + (-1)^{k+1} F(S_0; ...; S_{k-1}; (P, i, m)).
\end{aligned}
$$

The first term on the right is a combination of the image of F for the n_i simple vectors $P.S_0^{i,j_i}$ of degree $n = n_1 + ... + n_m$ which result from the division of (P, i, m) by S_0^i. If F is regular, this combination is the same as the combination of the simple vectors of degree n_i constituting the division of (P, i, m) by S_0^i, which gives the same result as the first term on the right in the formula

$$
\begin{aligned}
\delta F(S_0^i; ...; S_k^i; (P, 1, 1)) &= F(S_1^i; ...; S_k^i; \theta(P, S_0^i)) - F(\mu(S_0^i, S_1^i); ...; S_k^i; P) - ... \\
&\quad + (-1)^k F(S_0^i; ...; \mu(S_{k-1}^i, S_k^i); P) + (-1)^{k+1} F(S_0^i; ...; S_{k-1}^i; P).
\end{aligned}
$$
(157)

If F is regular the term number $l > 1$ on the right of the equation (156) coincides with the corresponding term on the right of the Equation (157).

Therefore the terms on the left in Equation (156) coincides with the left term in (157); which establishes the lemma.

We define $\mathcal{C}_r^*(\mathcal{M}_X)$ as the sub-complex of regular vectors in $\mathcal{C}^*(\mathcal{M}_X)$. Its elements are named *tree information cochains* or *arborescent information cochains*.

By definition, the *tree information co-homology* is the homology of this regular complex, considered as a sheaf of complexes over the category $\mathbf{S}(A)$, *i.e.*, a contravariant functor. This corresponds to the topos information co-homology in the monadic context.

To recover the case of the ordinary algebra of partitions, and the formulas of the bar construction in the first sections of this article, we have to take the special case where all the decompositions of the same level coincide at every level of the forests. In this case, we can replace the quotient $\mathcal{M}_X$ by the modules of conditioning by a redefinition of the action on functions $\mathcal{F}_X$. However the notion of divided probabilities for observation trees and the definition of co-homology in the monadic context can be seen as the natural basis of information co-homology.

When $k = 0$, in the classical case, a cochain is a function $f(\mathbb{P})$, the locality condition tells that it is a constant; and in this case it is a cocycle because the sum of probabilities equals one implies $f(\mathbb{P}) = f_S(\mathbb{P})$. Then H_τ^0 has dimension one.

When $k = 0$, in the quantum case, the spectral functions of ρ in the $\mathbf{Q}_X$ gives invariant information co-chains. Among them the Von Neumann entropy is specially relevant because its co-boundary gives the classical entropy. However, only the constant function is an invariant zero degree co-cycle. Thus again H_U^0 has dimension one.

For $k = 1$, a cochain is given by a function $F_X(S; P)$, such that, each time we have $X \to Y \to S$ and elements of Y refines S, we have $F_X(S; P) = F_Y(S; Y_*P)$. It is a cocycle when for every collection $S_1, ..., S_m$ of m observables, where m is the length of S, we have

$$(158) \qquad F(\mu_m(S, (S_1, ..., S_m)); P) = F(S; P) + \sum_i \mathbb{P}(S = i) F(S_i; P | S = i).$$

Note that the partition $\mu_m(S, (S_1, ..., S_m))$ is not the joint of S and the S_i for $i \geq 1$, except when all the S_i coincide. Thus it is amazing that the ordinary entropy also satisfies this functional equation, finer than the Shannon's identity:

Proposition 8. The usual entropy $H(S_*\mathbb{P}) = H(S; \mathbb{P})$ is an arborescent co-cycle.

Proof. By linearity on the module of divided probabilities $\mathcal{M}_X$, we can decompose the probability $\mathbb{P}$ in the conditional probabilities $\mathbb{P} | (S = s)$, thus we can restrict the proof of the lemma to the case where $S = \pi_0$ is the trivial partition, *i.e.*, $m = 1$.

Let $X_i; i = 1, ..., m$ denote the elements of the partition associated to S_0 and $X_i^j; j = 1, ..., n_i$ the pieces of the intersection of X_i with the elements of the partition associate to S_i; note p_i the probability of the event X_i and p_i^j the probability of the event X_i^j; we have

$$(159) \qquad H(\mu_m(S_0; (S_1, ..., S_m)); \mathbb{P}) = -\sum_{i=1}^{i=m} \sum_{j=1}^{j=n_i} p_i^j \log_2 p_i^j,$$

and

$$(160) \qquad H_{S_0}(S_1; ...; S_m; \mathbb{P}) = -\sum_{i=1}^{i=m} p_i \sum_{j=1}^{j=n_i} \frac{p_i^j}{p_i} \log_2 \frac{p_i^j}{p_i}$$

$$(161) \qquad = -\sum_{i=1}^{i=m} \sum_{j=1}^{j=n_i} p_i^j (\log_2 p_i^j - \log_2 p_i)$$

$$(162) \qquad = -\sum_{i=1}^{i=m} \sum_{j=1}^{j=n_i} p_i^j \log_2 p_i^j + \sum_{i=1}^{i=m} \log_2 p_i \sum_{j=1}^{j=n_i} p_i^j$$

$$(163) \qquad = -\sum_{i=1}^{i=m} \sum_{j=1}^{j=n_i} p_i^j \log_2 p_i^j + \sum_{i=1}^{i=m} p_i \log_2 p_i,$$

then

$$(164) \qquad H(\mu_m(S_0; (S_1, ..., S_m)); \mathbb{P}) - H_{S_0}(S_1; ...; S_m; \mathbb{P}) = H(S_0; \mathbb{P}).$$

Q.E.D.

This identity was discovered by Faddeev, Baez, Fritz, Leinster see [3]. However, we propose that information homology explains its significance.

When the category of quantum information $\mathbf{S}$, the set A and the probability functor $\mathbf{Q}$ are invariant under the unitary group, and if we choose a classical full subcategory $\mathcal{S}$, there is trace map from $\mathbf{Q}$ to $\mathcal{Q}$, induces a morphism from the classical arborescent co-homology of $\mathcal{S}$, A and $\mathcal{Q}$ to the invariant quantum arborescent co-homology of $\mathbf{S}$, A and $\mathbf{Q}$.

As a corollary of the Lemma 10 and the Theorems 1 and 3, we obtain the following result:

Theorem 4. (i) both in the classical and the invariant quantum context, if $\mathbf{S}(A)$ is connected, sufficiently rich, and if $\mathbf{Q}$ is canonical, every 1-co-cycle is co-homologous to the entropy of Shannon; (ii) in the classical case $H^1(\mathcal{S}, A, \mathcal{Q})$ is the vector space of dimension 1 generated by the entropy; (iii) in the quantum case $H_U^1(\mathbf{S}, A, \mathbf{Q}) = 0$, and the only invariant 0-cochain which has for co-boundary the Shannon entropy is (minus) the Von-Neumann entropy.

6.4. **Arborescent Mutual Information.**

For $k = 2$, a cochain is given by a local function of a probability and a rooted decorated tree of level 2. It is a cocycle when the following functional equation is satisfied

$$\sum_i \mathbb{P}(S = i) F(T_i; U_i; P | S = i) - F(S; T; P)$$

$$(165) \qquad = \quad F(\mu_m(S \circ T); U; P) - F(S; (\mu_{n_i}(T_i \circ U_i); i \in [m]); P),$$

where S denotes a variable of length m, T a collection of m variables $T_1, ..., T_m$ of respective lengths $n_1, ..., n_m$ and U a collection of variables $U_{i,j}^k$ of respective lengths $n_{i,j}$, with i going from 1 to m, j going from 1 to n_i and k going from 1 to $n_{i,j}$; the notation U_i denoting the collection of variables $U_{i,j}^k$ of index i.

Our aim is to extend in the monadic context the topological action of the ordinary information structure on functions of probability used in the discussion of mutual information.

For that, we define another structure of $\mathcal{V}_X$-right module on the functor $\mathcal{M}_X$ associated to probabilities, by defining the following map $\theta_t(m)$ from $\mathcal{M}_X(m)$ tensorized with $V_X(n_1) \otimes ... \otimes V_X(n_m)$ to $\mathcal{M}_X(n)$, for $n = n_1 + ... + n_m$:

(166) $$\theta_t((P, i, m) \otimes S_1 \otimes ... \otimes S_m) = \sum_{j=1,...,n_i} (P, (i, j), n).$$

Remark that the generalized decompositions S_j are used only through the orders on their elements.

As for $\mathcal{R}$, it is easy to verify that the collection of maps $\theta_t(m)$ defines a right action of the monad $\mathcal{V}_X$ on the Schur functor $\mathcal{M}_X$.

Then we consider as before, the graded vector space $C^*(\mathcal{M}_X)$ of homomorphisms of $\mathcal{V}$-modules from the functors $\mathcal{M} \circ \mathcal{V}^{\circ k}; k \geq 0$ to the functor $\mathcal{R}$ which are measurable in the probabilities P. As before, on $C^*(\mathcal{M}_X)$, we shift the degree by one, because of the independency with respect to the last stage of the forest, which follows from the trivial action on $\mathcal{R}$.

The topological coboundary operator δ_t is defined in every degree by the formula of the simplicial bar construction, as in Equation (153) for δ, but with θ_t replacing θ. It corresponds to the usual simplicial complex of the family $\mathcal{V}^{\circ k}$. A cochain is represented by a family of functions of probability laws $F_X(S_1; ...; S_k; (P, i, m))$, where $S_1; ...; S_k$ denotes a forest with m trees of level k. The operator δ_t is given by

(167) $$\delta_t F(S_0; ...; S_k; (P, i, m)) = F(S_1; ...; S_k; \theta_t((P, i, m), S_0))$$
$$- F(\mu(S_0, S_1); ...; S_k; (P, i, m)) - ... + (-1)^k F(S_0; ...; \mu(S_{k-1}, S_k); (P, i, m))$$
$$+ (-1)^{k+1} F(S_0; ...; S_{k-1}; (P, i, m)).$$

where $n = n_1 + ... + n_m$ is the sum of numbers of branches of the generalized decompositions S_0^i for $i = 1, ..., m$.

As for δ, a value $F(S_1; ...; S_k; (P, j, n)$ depends only on the tree $S_1^j; ...; S_k^j$ rooted at the place numbered by j in the forest $S_1; ...; S_k$.

Lemma 12. The coboundary δ_t sends a regular cochain to a regular cochain.

Proof. Consider a simple vector (P, i, m) in $\mathcal{M}_X(m)$ and a forest $S_0; ...; S_k$ with m components; we denote by S_0^j the variable number j having degree n_j, and $n = n_1 + ... + n_m$, and we consider the formula (167).

If F is regular the first term on the right is the sum of the images by F for P and the n_i trees S_1^{i,j_i} which result from the forgetting of the first branches S_0^i, and the other terms on the right are equal to the value of F for P and the tree rooted at i in S_0. On the other side for the tree $S_0^i; ...; S_k^i$, if F is regular, we have

(168)
$$\delta F(S_0^i; ...; S_k^i; (P, 1)) = \sum_j F(S_1^{i,j}; ...; S_k^{i,j}; (P, 1)) - F(\mu(S_0^i, S_1^i); ...; S_k^i; (P, 1)) - ...$$
$$+ (-1)^k F(S_0^i; ...; \mu(S_{k-1}^i, S_k^i); (P, 1)) + (-1)^{k+1} F(S_0^i; ...; S_{k-1}^i; (P, 1)).$$

Thus δF is topologically regular.

Consequently we can restrict δ_t to the subcomplex $\mathcal{C}_r^*(N_X)$, and name its homology the *arborescent*, or *tree*, *topological information co-homology*, written $H_{\tau,t}^*(\mathbf{S}^*, A, \mathbf{Q})$.

Now we suggest to extend the notion of mutual information $I(X;Y;\mathbb{P})$ in the way it will be a cocycle for this co-homology as it was the case for the Shannon mutual information in the ordinary topological information complex. We suggest to adopt the formulas using δ and δ_t, as in the standard case:

Definition 18. Let $H(T;(P,i,m))$ denotes the regular extension to forests of the usual entropy; then the *mutual arborescent information* between a partition S of length m and a collection T of m partitions $T_1, ..., T_m$ is defined by

$$(169) \qquad I_\alpha(S;T;\mathbb{P}) = \delta_t H(S;T;\mathbb{P}).$$

The identity $\delta H = 0$ implies

$$(170) \qquad I_\alpha(S;T;\mathbb{P}) = \sum_{i=1}^{i=m} H(T_i;\mathbb{P}) - \mathbb{P}(S=i)H(T_i;\mathbb{P}|S=i)).$$

In the particular case were all the T_i are equal to a variable T, it gives

$$I_\alpha(S;T;\mathbb{P}) = \sum_{i=1}^{i=m} \mathbb{P}(S=i)(H(T;\mathbb{P}) - H(T;\mathbb{P}|S=i)) + (m-1)H(T;\mathbb{P})$$

$$= H(T;P) - \sum_{i=1}^{i=m} \mathbb{P}(S=i)H(T;\mathbb{P}|S=i)) + (m-1)H(T;\mathbb{P})$$

$$= H(T;\mathbb{P}) - H_S(T;\mathbb{P}) + (m-1)H(T;\mathbb{P}),$$

then

$$(171) \qquad I_\alpha(S;T;\mathbb{P}) = I(S;T;\mathbb{P}) + (m-1)H(T;\mathbb{P}).$$

For $\mathcal{S}(A)$, the function I_α is an arborescent topological 2-cocycle.

It satisfies the Equation (165) were $\mathbb{P}$ replaces conditional probabilities $\mathbb{P}|(S=i)$ and where the factors $\mathbb{P}(S=i)$ disappear. Remark that, in this manner, maximization of $I_\alpha(S;T;\mathbb{P})$ comports maximization of usual mutual information $I(S;T;\mathbb{P})$ and unconditioned entropies $H(T_i;\mathbb{P})$.

Pursuing the homological interpretation of higher mutual information quantities given by the Formulas (55) and (56), we suggest the following definition:

Definition 19. The mutual arborescent informations of higher orders are given by $I_{\alpha,N} = -(\delta\delta_t)^M H$ for $N = 2M+1$ odd and by $I_{\alpha,N} = \delta_t(\delta\delta_t)^M H$ for $N = 2M+2$ even.

REFERENCES

[1] M. Artin, A. Grothendieck, and J.L. Verdier, *Thorie des topos et cohomologie tale des schmas - (sga 4) vol i,ii,iii*, Berlin; New York, Springer-Verlag, coll. Lecture notes in mathematics ,? 1972, 1964.

[2] C.G. Bachet de Mziriac, *Problmes plaisans et dlectables, qui se font par les nombres*, rdition : A. Blanchard, Paris 1993, 1612.

[3] J.C. Baez, T. Fritz, and T. Leinster, *A characterization of entropy in terms of information loss*, Entropy **13** (2011), 1945–1957.

[4] R. Balian, *Emergences in quantum measurement processes*, KronoScope **13:1** (2013), 85–95.

[5] P. Baudot and D. Bennequin, *Information topology i*, Preprint I (2015).

[6] _____, *Information topology ii*, Preprint II (2015).

[7] _____, *Information topology iii*, Preprint III (2015).

[8] P. Baudot and D. Bennequin, *Topological forms of information.*, AIP Conf. Proc. **1641** (2015), 213–221.

[9] J. Beck, *Triples, algebras and cohomology*, Dissertation, Columbia University., 1967.

[10] A. Borel and L. Ji, *Compactifications of symmetric and locally symmetric spaces*, Mathematics: Theory & Applications, Boston, MA: Birkhuser Boston., 2006.

[11] N. Bourbaki, *Algbre, chapitre 10, algbre homologique*, Masson, paris., 1980.

[12] N. Brenner, S. Strong, R. Koberle, and W. Bialek, *Synergy in a neural code.*, Neural Computation. **12** (2000), 1531–1552.

[13] S. Cartan, H.and Eilenberg, *Homological algebra*, The Princeton University Press, Princeton, 1956.

[14] J.L. Cathelineau, *Sur l'homologie de sl2 a coefficients dans l'action adjointe*, Math. Scand. **63** (1988), 51–86.

[15] G. Catren, *Towards a group-theoretical interpretation of mechanics.*, Phil. Sci arch. http://philsci-archive.pitt.edu/10116/ (2013).

[16] F.R. Cohen, T.J. Lada, and J.P. May, *The homology of iterated loop spaces*, Lecture Notes in Mathematics 533, Springer, Berlin, 1976.

[17] J.A. Cover, T.M.; Thomas, *Elements of information theory.*, Wiley Series in Telecommunication, 1991.

[18] A. Doering and C.J. Isham, *Classical and quantum probabilities as truth values.*, Journal of Mathematical Physics **Vol: 53** (2012).

[19] Ph. Elbaz-Vincent and H. Gangl, *On poly(ana)logs i.*, Compositio Mathematica **130(2)** (2002), 161–214.

[20] B. Fresse, *Koszul duality of operads and homology of partitionn posets.*, Contemp. Math. Amer. Math. Soc. **346** (2004), pp. 115–215.

[21] P. Gabriel, *Objets injectifs dans les catgories abliennes*, Sminaire Dubreil. Algbre et thorie des nombres **tome 12, no2 , exp. n17** (1958-1959), p. 1–32.

[22] E. Getzler and J. D. S. Jones, *Operads, homotopy algebra and iterated integrals for double loop spaces*, preprint. arXiv:hep-th/9403055v1 (1994).

[23] V. Ginzburg and M. M. Kapranov, *Koszul duality for operads*, Duke Math. J. **76** (1994), pp. 203–272.

[24] M. Gromov, *In a search for a structure, part 1: On entropy.*, unpublished manuscript http://www.ihes.fr/ gromov/PDF/structre-serch-entropy-july5-2012.pdf (2013).

[25] A. Grothendieck, *Sur quelques points d'algbre homologique, i*, Tohoku Math. J. **Vol 9, Nb 2** (1957), 119–221.

[26] T. S. Han, *Linear dependence structure of the entropy space*, Information and Control. **vol. 29** (1975), p. 337–368.

[27] Kuo Ting Hu, *On the amount of information.*, Theory Probab. Appl. **7(4)** (1962), 439–447.

[28] E. T. Jaynes, *Information theory and statistical mechanics*, Statistical Physics. New York: Benjamin., 1963.

[29] _____, *Prior probabilities.*, IEEE Transactions on Systems Science and Cybernetics **4 (3)** (1968), pp.227–241.

[30] D.G. Kendall, *Functional equations in information theory*, Z. Wahrscheinlichkeitstheorie **2** (1964), p. 225–229.

[31] A.I. Khinchin, *Mathematical foundations of information theory*, New York: Dover. Translated by R. A. Silverman and M.D. Friedman from two Russian articles in Uspekhi Matematicheskikh Nauk, 7 (1953): 320 and 9 (1956): 1775, 1957.

[32] Hoon Kim, John Watkinson, Vinay Varadan, and Dimitris Anastassiou, *Multi-cancer computational analysis reveals invasion-associated variant of desmoplastic reaction involving inhba, thbs2 and col11a1*, BMC Medical Genomics , **3:51** (2010).

[33] A. Kolmogorov, *Combinatorial foundations of information theory and the calculus of probabilities*, Russ. Math. Surv. **38 29** (1983).

[34] A. N. Kolmogorov, *Grundbegriffe der wahrscheinlichkeitsrechnung.(english translation (1950): Foundations of the theory of probability.)*, Springer, Berlin (Chelsea, New York)., 1933.

[35] M. Kontsevitch, *The 11/2 logarithm*, Unpublished note. Reproduced in Elbaz-Vincent & Gangl, 2002 On poly(ana)logs I. Compositio Mathematica (1995).

[36] P.M. Lee, *On the axioms of information theory*, The Annals of Mathematical Statistics **Vol. 35, No. 1** (1964), pp. 415–418.

[37] J-L. Loday and B. Valette, *Algebraic operads.*, Springer, 2012.

[38] S. Mac Lane, *Homology.*, Classic in Mathematics, Springer, Reprint of the 1975 edition, 1975.

[39] _____, *Categories for the working mathematician.*, Graduate Texts in Mathematics. Springer, 1998.

[40] M. Marcolli and R. Thorngren, *Thermodynamic semirings*, arXiv 10.4171/JNCG/159 **Vol. abs/1108.2874** (2011).

[41] H. Matsuda, *Information theoretic characterization of frustrated systems.*, Physica A: Statistical Mechanics and its Applications. **294 (1-2)** (2001), 180–190.

[42] J. P. May, *The geometry of iterated loop spaces*, Volume 271 of Springer Lecture Notes in Mathematics. Springer., 1972.

[43] _____, *Einfinite ring spaces and einfinite ring spectra.*, volume 577 of Springer Lecture Notes in Mathematics. Springer, 1977.

[44] W.J. McGill, *Psychometrika*, Multivariate information transmission **19** (1954), p. 97–116.

[45] P.A. Meyer, *Quantum probability for probabilists*, Lecture notes in mathematic. Springer. Berlin, 1993.

[46] M.A. Nielsen and I.L. Chuang, *Quantum computation and quantum information*, Cambridge University Press, 2000.

[47] A. Prout, *Introduction   la logique catgorique*, Prpublications, Paris 7 Logique www.logique.jussieu.fr/ alp/, 2013.

[48] W. Rindler and R. Penrose, *Spinors and spacetime*, Cambridge University Press, 1986, 2e d., 1984.

[49] C. E. Shannon, *A mathematical theory of communication.*, The Bell System Technical Journal **27** (1948), 379–423.

[50] J.M. Souriau, *Structure des systemes dynamiques*, ditions Dunod, 1970.

[51] R. Thom, *Stabilit struturelle et morphognse.*, deuxime dition, InterEdition, Paris, 1977.

[52] H. Tverberg, *A new derivation of information function*, Math. Scand. **6** (1958), 297–298.

[53] Shinsuke Uda, Takeshi H Saito, Takamasa Kudo, Toshiya Kokaji, Takaho Tsuchiya, Hiroyuki Kubota, Yasunori Komori, Yu ichi Ozaki, and Shinya Kuroda, *Robustness and compensation of information transmission of signaling pathways*, Science **Vol. 341 no. 6145** (2013), pp. 558–561.

[54] J. Watkinson, K.C. Liang, X. Wang, T. Zheng, and D. Anastassiou, *Inference of regulatory gene interactions from expression data using three-way mutual information.*, The Challenges of Systems Biology: Ann. N.Y. Acad. Sci. **1158** (2009), 302–313.

[55] R.W. Yeung, *Information theory and network coding.*, Springer, 2007.

Max Planck Institute for Mathematics in the Sciences, Inselstrasse 22, 04103 Leipzig, Germany

E-mail address: `baudot@mis.mpg.de`

Universite Paris Diderot-Paris 7, UFR de Mathematiques, Equipe Geometrie et Dynamique, Batiment Sophie Germain, 5 rue Thomas Mann, 75205 Paris Cedex 13, Franc