Topos and Stacks of Deep Neural Networks

Jean-Claude Belfiore

Huawei Advanced Wireless Technology Lab.
Paris Research Center

Daniel Bennequin

Huawei Advanced Wireless Technology Lab.
Paris Research Center
University of Paris Diderot, Faculty of Mathematics

Abstract

Every known artificial Deep Neural Network (DNN) corresponds to an object in a canonical Grothendieck's topos; its learning dynamic corresponds to a flow of morphisms in this topos. Invariance structures in the layers (like CNNs or LSTMs) correspond to Giraud's stacks. This invariance is supposed to be responsible of the generalization property, that is extrapolation from learning data under constraints. The fibers represent pre-semantic categories (Culioli [CLS95], Thom [Tho72]), over which artificial languages are defined, with internal logics, intuitionist, classical or linear (Girard [Gir87]). Semantic functioning of a network is its ability to express theories in such a language for answering questions in output about input data. Quantities and spaces of semantic information are defined by analogy with the homological interpretation of Shannon's entropy (Baudot & Bennequin [BB15]). They generalize the measures found by Carnap and Bar-Hillel [CBH52]. Amazingly, the above semantical structures are classified by geometric fibrant objects in a closed model category of Quillen [Qui67], then they give rise to homotopical invariants of DNNs and of their semantic functioning. Intentional type theories (Martin-Löf [ML80]) organize these objects and fibrations between them. Information contents and exchanges are analyzed by Grothendieck's derivators [Gro90].

Contents

Preface

Introduction

This text presents a general theory of semantic functioning of deep neural networks, DNNs, based on topology, more precisely, Grothendieck's topos, Quillen's homotopy theory, Thom's singularity theory and the pre-semantic of Culioli in enunciative linguistic.

The theory is based on the existing networks, transforming data, as images, movies or written texts, to answer questions, achieve actions or take decisions. Experiments, recent and past, show that the deep neural networks, which have learned under constrained methods, can achieve surprising semantic performances [XQLJ20], [BBD+11], [BBDH14], [BBG21a], [DHSB20], [KL14], [MXY+15], [ZRS+18], [ZCZ+19], [GLH+20]. However, the exploitation of more explicit invariance structures and adapted languages, are in great part a task for the future. Thus the present text is a mixture of an analysis of the functioning networks, and of a conjectural frame to make them able to approach more ideal semantic functioning.

Note that categories, homology and homotopy were recently applied in several manners to semantic information. An example is the application of category theory to the design of networks, by Fong and Spivak [FS18]. For a recent review on many applications of category theory to Machine Learning, see [SGW21]. Other examples are given by the general notion of Information Networks based on Segal spaces by Yuri Manin and Matilde Marcolli, [MM20] and the Čech homology reconstruction of the environment by place fields of Curto and collaborators, [Cur17]. Let us also mention the characterization of entropy, by Baez, Fritz, Leinster, [BFL11], and the use of sheaves and cosheaves for studying information networks, Ghrist, Hiraoka 2011 [GH11], Curry 2013 [Cur13], Robinson and Joslyn [Rob17], and Abramsky et al. specially for Quantum Information [AB11]. Persistent homology for detecting structures in data must also be cited in this context, for instance Port, Karidi, Marcolli 2019, [PKM19] on syntactic structures, and Carlsson et al. on shape recognition [CZCG05]. More in relation with Bayes networks, there are the three recent PhD theses of Juan-Pablo Vigneaux [Vig19], Olivier Peltre [Pel20] and Grégoire Sergeant-Perthuis [SP21].

With respect to these works, we look at a notion of information which is a (toposic) topological invariant of the situation which involves three dimensions of dynamics:

1) a logical flow along the network;

2) in the layers, the action of categories;

3) the evocations of meaning in languages.

The resulting notion of information generalizes the suggestion of Carnap and Bar-Hillel 1952 in these three dynamical directions. Our inspiration came from the toposic interpretation of Shannon's entropy in [BB15] and [Vig20]. A new fundamental ingredient is the interpretation of internal implication (exponential) as a *conditioning* on theories, analogous to the conditioning in probabilities. We distinguish between the theoretically accessible information, concerning all the theories in a fibred languages, and the practically accessible information, that corresponds to the semantic functioning of concrete neural networks, associated to a feed-forward dynamics which depends on a learning process.

The main results in this text are,

✔ theorems 1.1 and 1.2 characterizing the topos associated to a DNN

✔ theorem 2.1 giving a geometric sufficient condition for a fluid circulation of semantics in this topos

✔ theorems 2.2 and 2.3, characterizing the fibrations (in particular the fibrant objects) in a closed model category made by the stacks of the DNNs having a given network architecture

✔ the tentative definition of Semantic Information quantities and spaces in sections 3.4 and 3.5

✔ theorem 4.1 on the generic structures and dynamics of LSTMs.

Specific examples, showing the nature of the semantic information that we present here, are at the end of section 3.5 extracted from the exemplar toy language of Carnap and Bar-Hillel and the mathematical interpretation of the pre-semantic of Culioli in relation with the artificial memory cells of sections 4.4 and 4.5.

Chapter 1 describes the nature of the sites and the topos associated to deep neural networks, said $DNNs$, with their dynamics, feedforward and backward (backpropagation) learning.

Chapter 2 presents the different stacks of a DNN, which are fibred categories over the site of the DNN, incorporating symmetries and logics for approaching the wanted semantics in functioning. Usual examples are $CNNs$ for translation symmetries, but also other ones regarding logic and semantics (see experiments in *Logical Information Cells I* [BBG21a]). Thus the logical structure of the classifying topos of such a stack is described. We introduce hypotheses on the stack and the language objects that allow a transmission of theories downstream and of propositions upstream in the network. The 2-category of the stacks over a given architecture is shown to constitute a closed model theory of injective type, in the sense of Quillen (also Cisinski and Lurie). The fibrant objects, which are difficult to characterize in general, are determined in the case of the Grothendieck sites of $DNNs$. Interestingly, they correspond to the hypothesis guarantying the transmission of theories. Using the work of Arndt and Kapulkin [AK11] we show that the above model theory gives rise to a Martin-Löf type theory associated to every DNN. Semantics in the sense of topos (Lambek) is added by considering objects in the classifying topos of the stack.

In **chapter 3**, we start exploring the notion of semantic information and semantic functioning in $DNNs$, by using homology and homotopy theory. Then we define semantic conditioning of the theories by the propositions, and compute the corresponding ringed cohomology of the functions of these theories; this gives a numerical notion of semantic ambiguity, of semantic mutual information and of semantic Kullback-Leibler divergence. Then we generalize the homogeneous bar-complex to define a bi-simplicial set $I_\star^\bullet$ of classes of theories and propositions histories over the network, by taking homotopy colimits. We introduce a class of increasing and concave functions from $I_\star^\bullet$ to an external model category $\mathcal{M}$; and with them, we obtain natural homotopy types of semantic information, associated to coherent semantic functioning of a network with respect to a semantic problem; they satisfy properties conjectured by Carnap and Bar-Hillel in 1952 [CBH52] for the sets of semantic information. On the simple example they studied we show the interest of considering spaces of information, in particular groupoids, in addition to the more usual combinatorial dimension of logical content of propositions.

Chapter 4 describes examples of memory cells, as the long and short terms memory cells (LSTM), and shows that the natural groupoids for their stack have as fundamental group the group of Artin's braids with three strands $\mathfrak{B}_3$. Generalizations are proposed, for semantics closer to the semantic of natural languages, in appendix E.

Finally **chapter 5** introduces possible applications of topos, stacks and models to the relations between several $DNNs$: understanding the modular structures of networks, defining and studying the obstructions to integrate some semantics or to solve problems in some contexts. Examples could be taken from the above mentioned experiments on logical information cells, and from recent attempts of several teams in artificial intelligence: Hudson & Manning [HM18], Santoro, Raposo et al. [SRB+17], Bengio and Hinton, using memory modules, linguistic analysis modules, attention modules and relation modules, in addition to convolution $CNNs$, for answering questions about images and movies (also see [RSB+17], [ZCZ+19], [HB20]).

Most of the figures mentioned in the text can be found in the chapter by Bennequin and Belfiore *On new mathematical concepts for Artificial Intelligence*, in the Huawei volume on *Mathematics for Future Computing and Communication*, edited by Liao Heng and Bill McColl [HM21]. We also refer to this chapter for the elements of category theory that are necessary to understand this text, the definitions and first properties of topos and Grothendieck topos, and the presentation of elementary type theories.

Chapter 9 in [HM21], by Ge Yiqun and Tong Wen, *Mathematics, Information and Learning*, explains the large place of topology in the notions of semantic information.

In a forthcoming preprint, entitled *A search of semantic spaces*, we will compute spaces of semantic information for several elementary languages, along the lines indicated in section 3.5, and develop further the Galois point of view on the information flow in a network. The notions of intentional signification, meaning and knowledge are discussed from a philosophical point of view, and adapted to artificial semantic and its intelligibility.

In another following preprint, *A mathematical theory of semantic communication*, we plan to present the application of the above stacks of functioning DNNs and their information spaces, to the problem of semantic communication. In particular we show how the invariance structures in the fibers, made by categories acting on artificial languages, give a way to understand generalization properties of DNNs, for extrapolation, not only interpolation.

Analytical aspects, as equivariant standard DNNs approximation of functions, or gradient descent respecting the invariance, are developed in this context.

Acknowledgements

The two authors deeply thank Olivia Caramello and Laurent Lafforgue for the impulsion they gave to this research, for their constant encouragements and many helpful suggestions. They also warmly thank Merouane Debbah for his deep interest, the help and the support he gave, Xavier Giraud, for the concrete experiments he realized with us, allowing to connect the theory with the lively spontaneous behavior of artificial neural networks, and Zhenrong Liu (Louise) for her constant and very kind help at work.

D.B. gives special thanks to Alain Berthoz, with whom he has had the chance to work and dream since many years on a conjectural topos geometry (properly speaking stacks) for the generation and control of the variety of humans voluntary movements. He also does not forget that the presence of natural invariants of topos in Information theory was discovered during a common work with Pierre Baudot, that he heartily thanks. D.B. had many inspiring discussions on closely related subjects with his former students, in particular Alireza Bahraini, Alexandre Afgoustidis, Juan-Pablo Vigneaux, Olivier Peltre and Grégoire Sergeant-Perthuis, that he friendly thanks, with gratitude.

1 Architectures

Let us show how every (known) artificial deep neural network (DNN) can be described by a family of objects in a well defined topos.

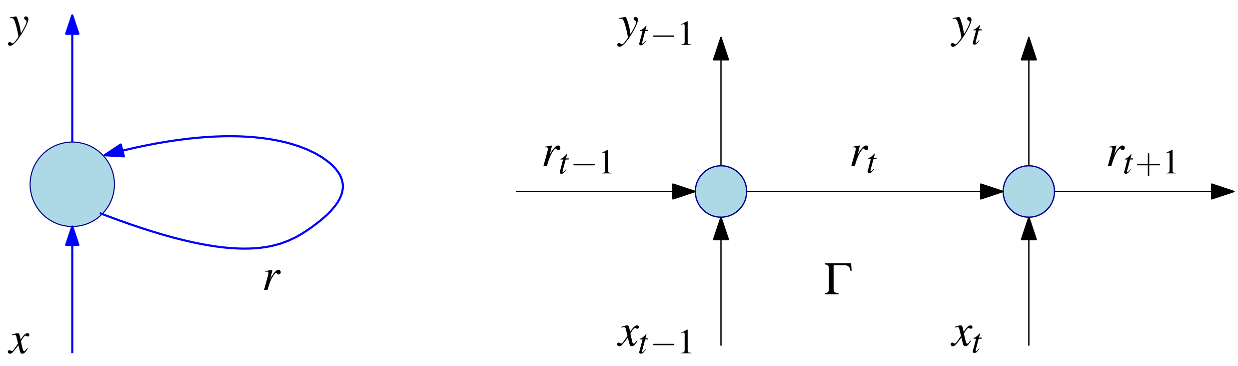
1.1 Underlying graph

Definition. *An oriented graph* Γ *is* directed *when the relation* $a \leq b$ *between vertices, defined by the existence of an oriented path, made by concatenation of oriented edges, is a partial ordering on the set* $V(\Gamma) = \Gamma_{(0)}$ *of vertices. A graph is said* classical *if there exists at most one edge between two vertices, and no loop at one vertex (also named tadpole). A classical directed graph can have non-oriented cycles, but no oriented cycles.*

The layers and the direct connections between layers in an artificial neural network constitute a finite oriented graph Γ, which is directed, and classical.

The minimal elements correspond to the initial layers, or input layers, and the maximal elements to the final layers, or output layers, all the other correspond to hidden layers, or inner layers. In the case of $RNNs$ (as when we look at feedback connections in the brain) we apparently see loops, however they are not loops in space-time, the graph which represents the functioning of the network must be seen in the space-time (not necessary Galilean but causal), then the loops disappear and the graph appears directed and classical (see figure 1.1). Apparently there is no exception to these rules in the world of $DNNs$.

Remark. Bayesian networks are frequently associated to oriented or non-oriented graphs, which can be non-directed, and have oriented loops. However, the underlying random variables are associated to vertices and to edges, the variable of an edge ab being the joint variable of the variables of a and b. More generally, an hypergaph is considered, made by a subset $\mathcal{A}$ of the set $\mathcal{P}(I)$ of subsets of a given set I. In this situation, we have a poset, where the natural partial ordering relation is the opposite of the inclusion, i.e. it goes from the finer variable to the coarser one.



(a) Original RNN

(b) Unfolded RNN in Space-Time

Figure 1.1: RNN with space-time unfolding

1.2 Dynamical objects of the chains

The simplest architecture of a network is a chain, and the feed-forward functioning of the network, when it has learned, corresponds to a covariant functor X from the category $C^o(\Gamma)$ freely generated by the graph to the category of sets, Set: to a layer $L_k; k \in \Gamma$ is associated the set X_k of possible activities of the population of neurons in L_k, to the edge $L_k \mapsto L_{k+1}$ is associated the map $X^w_{k+1,k} : X_k \longrightarrow X_{k+1}$ which corresponds to the learned weights $w_{k+1,k}$; then to each arrow in $C^o(\Gamma)$, we associate the composed map.

But also the weights can be encoded in a covariant functor Π from $C^o(\Gamma)$ to Set: for L_k we define Π_k as the product of all the sets $W_{l+1,l}$ of weights for $l \geq k$, and to the edge $k \mapsto k+1$ we associate the natural forgetting projection $\Pi_{k+1,k} : \Pi_k \longrightarrow \Pi_{k+1}$. (The product over an empty set is the singleton $\star$ in Set, then for the output layer L_n the last projection is the unique possible map from Π_{n-1} to $\star$.) In what follows, we will note $\mathbb{W} = \Pi$, for remembering that it describes the functor of weights, but the notation Π is less confusing for denoting the morphisms in this functor.

The cartesian products $X_k \times \Pi_k$ together with the maps

$$X_{k+1,k} \times \Pi_{k+1,k}\left(x_k, (w_{k+1,k}, w'_{k+1})\right) = \left(X^w_{k+1,k}(x_k), w'_{k+1}\right) \tag{1.1}$$

also defines a covariant functor $\mathbb{X}$; it represents all the possible feed-forward functioning of the network, for every potential weights. The natural projection from $\mathbb{X}$ to $\mathbb{W} = \Pi$ is a natural transformation of functors. It is remarkable that, in supervised learning, the Backpropagation algorithm is represented by a flow of natural transformations of the functor $\mathbb{W}$ to itself. We give a proof below in the general case, not only for a chain, where it is easier.

Remark a difference with Spivak et al. [FST19], where backpropagation is a functor, not a natural transformation.

In fact, the weights represent mappings between two layers, individually they correspond to morphisms in a functor X^w, then it should have been more intuitive if they had been coded by morphisms, however globally they are better encoded by the objects in the functor $\mathbb{W}$, and the morphisms in this functor are the erasure of the weights along the arrows that correspond to them. This appears as a kind of dual representation of the mappings X^w.

As we want to respect the convention of Topos theory, [AGV63], we introduce the category $C = C(\Gamma)$ which is opposed to $C^0(\Gamma)$; then X^w, $\mathbb{W} = \Pi$ and $\mathbb{X}$ become contravariant functors from this category C to Sets, i.e. presheaves over C, i.e. objects in the topos $C^\wedge$ [HM21]. This is this topos which is associated to the neural network which has the shape of a chain (multi-layer perceptron). Observe that the arrows between sets continue to follow the natural dynamical ordering, from the initial layer to the final layer, but the arrows in the category (the site) C are going now in the opposite direction.

The object X^w can be naturally identified with a subobject of $\mathbb{X}$, we call this singleton the fiber of $pr_2 : \mathbb{X} \to \mathbb{W}$ over the singleton w in $\mathbb{W}$, (that is a morphism in $C^\wedge$ from the final object $\mathbf{1}$ (the constant functor equal to the point $\star$ at each layer) to the object $\mathbb{W}$), which is a system of weights for each edge of the graph Γ.

In this simple case of a chain, the classifying object of subobjects Ω, which is responsible of the logic in the topos [Pro19], is given by the subobjects of $\mathbf{1}$; more precisely, for every $k \in C$, $\Omega(k)$ is the set of subobjects of the localization $\mathbf{1}|k$, made by the arrows in C going to k. All these subobjects are increasing sequences $(\emptyset, ..., \emptyset, \star, ..., \star)$. This can be interpreted as the fact that a proposition in the language (and internal semantic theory) of the topos is more and more determined when we approach the last layer. Which corresponds well to what happens in the internal world of the network, and also, in most cases, to the information about the output that an external observer can deduce from the activity in the inner layers [BBG21a].

1.3 Dynamical objects of the general DNNs

However, many networks, and most today's networks, are far from being simple chains. The topology of Γ is very complex, with many paths going from a layer to a deeper one, and many inputs and outputs at a same vertex. In these cases, the functioning and the weights are not defined by functors on $C(\Gamma)$ (the category opposite to the category freely generated by Γ). But a canonical modification of this category allows to solve the problem: at each layer a where more than one layer sends information, say $a', a", ...$, i.e. where there exist irreducible arrows $aa', aa", ...$ in $C(\Gamma)$ (edges in Γ^{op}), we perform a surgery: between a and a' (*resp.* a and $a"$, a.s.o.) introduce two new objects $A^\star$ and A, with arrows $a' \to A^\star$, $a" \to A^\star$, ..., and $A^\star \to A$, $a \to A$, forming a fork, with tips in $a', a", ...$ and handle $A^\star A a$ (more precisely if not too pedantically, the arrows $a'A^\star, a"A^\star, ...$ are the tines, the arrow $A^\star A$ is the tang, or socket, and the arrow aA is the handle) (see figure 1.2). By reversing arrows, this gives a new oriented graph $\mathbf{\Gamma}$, also without oriented cycles, and the category C which replaces $C(\Gamma)$ is the category $C(\mathbf{\Gamma})$, opposite of the category which is freely generated by $\mathbf{\Gamma}$.

Remark. In $\mathbf{\Gamma}$, the complement of the unions of the tangs is a forest. Only the convergent multiplicity

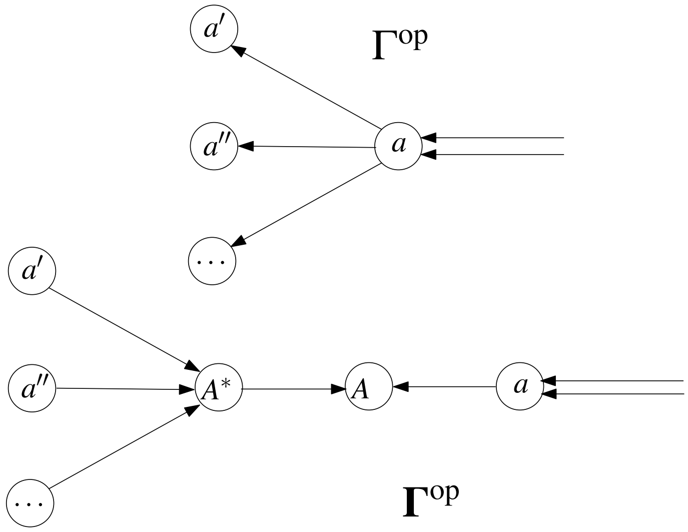


Figure 1.2: From the initial graph to the Fork

in Γ gives rise to forks, not the divergent one. In the category C, this convergence (*resp.* divergence) corresponds to a divergence (*resp.* convergence) of the arrows.

When describing concrete networks (see for instance RNN, and $LSTM$ or GRU memory cells that we will study in chapter 4), ambiguity can appear with the input layers: they can be considered as input or as tips when several inputs join for connecting a deeper layer a. The better attitude is to duplicate them; for instance two input layers x_t, h_{t-1} going to h_t, y_t, we introduce X_t, x'_t, H_{t-1}, h'_{t-1}, then a fork $A^\star, A$, and in C, arrows $x'_t \to X_t$, $h'_{t-1} \to H_{t-1}$ for representing the input data, arrows of fork $x'_t \to A^\star$, $h'_{t-1} \to A^\star$, $A^\star \to A$, and arrows of information transmissions $h_t \to A$ and $y_t \to A$, representing the output of the memory cell.

With this category C, it is possible to define the analog of the presheaves X^w, $\mathbb{W} = \Pi$ and $\mathbb{X}$ in general.

First X^w: at each old vertex, the set X^w_a is as before the set of activities of the neurons of the corresponding layer; over a point like $A^\star$ and A we put the product of all the incoming sets $X^w_{a'} \times X^w_{a"}, \dots$. The map from X_A to X_a is the dynamical transmission in the network, joining the information coming from all the inputs layers $a', a", \dots$ at a, all the other maps are given by the structure: the projection on its factors from $X^w_{A^\star}$, and the identity over the arrow $A^\star A$. It is easy to show, that given a collection of activities ε^0_{in} in all the initial layers of the network, it results a unique section of the presheaf X^w, a singleton, or an element of $\lim_C X^w$, which induces $\varepsilon^0_{\mathrm{in}}$. Thus, dynamically, each arrow of type $a \to A$ has replaced the set of arrows from a to $a', a", \dots$.

It is remarkable that the main structural part (which is the projection from a product to its components) can be interpreted by the fact that the presheaf is a sheaf for a natural Grothendieck topology J on the category C: in every object x of C the only covering is the full category $C|x$, except when x is of the type of $A^\star$, where we add the covering made by the arrows of the type $a' \to A^\star$ [AGV63].

The sheafification process, associating a sheaf $X^\star$ over (C,J) to any presheaf X over C is easy to describe: no value is changed except at a place $A^\star$, where $X_{A^\star}$ is replaced by the product $X^\star_{A^\star}$ of the $X_{a'}$, and the map from $X^\star_A = X_A$ to $X^\star_{A^\star}$ is replaced by the product of the maps from X_A to the $X_{a'}$ given by the functor X. In particular, important for us, the sheaf $C^\star$ associated to a constant presheaf C replaces C in $A^\star$ by a product C^n and the identity $C \to C$ by the diagonal map $C \to C^n$ over the arrow $A^\star A$.

Let us now describe the sheaf $\mathbb{W}$ over (C,J) which represents the set of possible weights of the DNN (or RNN a.s.o.). First consider at each vertex a of the initial graph Γ, the set W_a of weights describing the allowed maps from the product $X_A = \prod_{a' \leftarrow a} X_{a'}$ to X_a, over the projecting layers $a', a", ...$ to a. Then consider at each layer x the (necessarily connected) subgraph Γ_x (or $x|\Gamma$) which is the union of the connected oriented paths in Γ from x to some output layer (i.e. the maximal branches issued from x in Γ); take for $\mathbb{W}(x)$ the product of the W_y over all the vertices in Γ_x. (For the functioning, it is useful to consider the part $\mathbf{\Gamma}_x$ (or $x|\mathbf{\Gamma}$) which is formed from Γ_x, by adding the collections of points $A^\star, A$ when necessary, and the arrows containing them in $\mathbf{\Gamma}$.) At every vertex of type $A^\star$ or A of $\mathbf{\Gamma}$, we put the product $\mathbb{W}_A$ of the sets $\mathbb{W}_{a'}$ for the afferent $a', a", ...$ to a. If $x'x$ is an oriented edge of $\mathbf{\Gamma}$, there exists a natural projection $\Pi_{xx'} : \mathbb{W}(x') \to \mathbb{W}(x)$. This defines a sheaf over $C = C(\mathbf{\Gamma})$.

The crossed product $\mathbb{X}$ of the X^w over $\mathbb{W}$ is defined as for the simple chains. It is an object of the topos of sheaves over C that represents all the possible functioning of the neural network.

1.4 Backpropagation as a natural (stochastic) flow in the topos

Nothing is loosed in generality if we put together the inputs (*resp.* the output) in a product space X_0 (*resp.* X_n); this corresponds to the introduction of an initial vertex x_0 and a final vertex x_n in Γ, respectively connected to all the existing initial or final vertices.

We also assume that the spaces of states of activity X_a and the spaces of weights W_{aA} are smooth manifolds, and that the maps $(x,w) \mapsto X^w(x)$ defines smooth maps on the corresponding product manifolds.

In particular it is possible to define tangent objects in the topos of the network $T(\mathbb{X})$ and $T(\mathbb{W})$, and smooth natural transformations between them.

Supervised learning consists in the choice of an energy function

$$(\xi_0, w) \mapsto F(\xi_0; \xi_n(w, \xi_0)); \tag{1.2}$$

then in the search of the absolute minimum of the mean $\Phi = \mathbb{E}(F)$ of this energy over a measure on the inputs ξ_0; it is a real function on the whole set of weights $W = \mathbb{W}_0$. For simplicity, we assume that F

is smooth, and we do not enter the difficult point of effective numerical gradient descent algorithms, we just want to develop the formula of the linear form dF on $T_{w_0}W$, for a fixed input ξ_0 and a fixed system of weights w_0. The gradient will depend on the choices of a Riemannian metric on W. And the gradient of Φ is the mean of the individual gradients.
We have

$$dF(\delta w) = F^{\star} d\xi_n(\delta w), \tag{1.3}$$

then it is sufficient to compute $d\xi_n$.
The product formula is

$$\mathbb{W}_0 = \prod_{a \in \Gamma} W_{aA}, \tag{1.4}$$

where a describes all the vertices of Γ, Aa is the corresponding edge in $\mathbf{\Gamma}$. Then it is sufficient to compute $d\xi_n(\delta w_a)$ for $\delta w_a \in T_{w_0}W_{aA}$, assuming that all the other vectors δw_{bB} are zero, except δw_a which denotes the weight over the edge Aa.

For that, we consider the set Ω_a of directed paths γ_a in Γ going from a to the output layer x_n. Each such path gives rise to a zigzag in $\mathbf{\Gamma}$:

$$\ldots \leftarrow B' \rightarrow b' \leftarrow B \rightarrow b \leftarrow \ldots \tag{1.5}$$

which gives a feed-forward composed map, by taking over each $B \rightarrow b$ the map $X^{w_{bB}}$ from the product X_B to the manifold X_b, where everything is fixed by ξ_0 and w_0 except on the branch coming from b', where w_a varies, and by taking over each $b' \leftarrow B$ the injection $\rho_{Bb'}$ defined by the other factors $X_{b"}, X_{b'''}, \ldots$ of X_B. This composition is written

$$\phi_{\gamma_a} = \prod_{b_k \in \gamma_a} X^{w_0}_{b_k B_k} \circ \rho_{B_k b_{k-1}} \circ X^{w}_{aA}; \tag{1.6}$$

going from the manifold $W_a \times X_A$ to the manifold X_n. In the above formula, k starts with 1, and $b_0 = a$.

Two different elements γ'_a, $\gamma"_a$ of Ω_a must coincide after a given vertex c, where they join from different branches $c'c$, $c"c$ in Γ; they pass through B in $\mathbf{\Gamma}$; then we can define the sum $\phi_{\gamma'_a} \oplus \phi_{\gamma"_a}$, as a map from $W_{aA}^{\oplus 2} \times X_A$ to X_n, by composing the maps between the $X's$ after b, from b to x_n, with the two maps $\phi_{\gamma'_a}$ and $\phi_{\gamma"_a}$ truncated at B. We name this operation the cooperation, or cooperative sum, of $\phi_{\gamma'_a}$ and $\phi_{\gamma"_a}$.
Cooperation can be iterated in associative and commutating manner to any subset of Ω_a, representing a tree issued from x_n, embedded in Γ, made by all the common branches between the pairs of paths from a to x_n. The full cooperative sum is the map

$$\bigoplus \phi_{\gamma_a} : X_A \times \bigoplus_{\gamma_a \in \Omega_a} W_{aA} \rightarrow X_n. \tag{1.7}$$

For a fixed ξ_0, and all w_{bB} fixed except w_{aA}, the point $\xi_n(w)$ can be described as the composition of the diagonal map with the total cooperative sum

$$w_a \mapsto (w_a, ...w_a) \in \bigoplus_{\gamma_a \in \Omega_a} W_{aA} \to X_n. \tag{1.8}$$

This gives

$$d\xi_n(\delta w_a) = \sum_{\gamma_a \in \Omega_a} d\phi_{\gamma_a} \delta w_a; \tag{1.9}$$

which implies the backpropagation formula:

Lemma 1.1.

$$d\xi_n(\delta w_a) = \sum_{\gamma_a \in \Omega_a} \prod_{b_k \in \gamma_a} DX^{w_0}_{b_k B_k} \circ D\rho_{B_k b_{k-1}} \circ \partial_w X^w_{aA}.\delta w_a \tag{1.10}$$

going from the tangent space $T_{w_a^0}(W_a)$ to the tangent space $T_{\xi_n^0}(X_n)$. In this expression, k starts with 1*, and $b_0 = a$.*

To get the backpropagation flow, we compose to the left with $F^\star = dF$, which gives a linear form, then apply the chosen metric on the manifold W, which gives a vector field $\beta(w_0|\xi_0)$. Let us assume that the function F is bounded from below on $X_0 \times W$ and coercive (at least proper). Then the flow of β is globally defined on W. From it we define a one parameter group of natural transformations of the object $\mathbb{W}$.

In practice, a sequence $\Xi_m; m \in [M]$ of finite set of inputs ξ_0 (benchmarks) is chosen randomly, according to the chosen measure on the initial data, and the gradient is taken for the sum

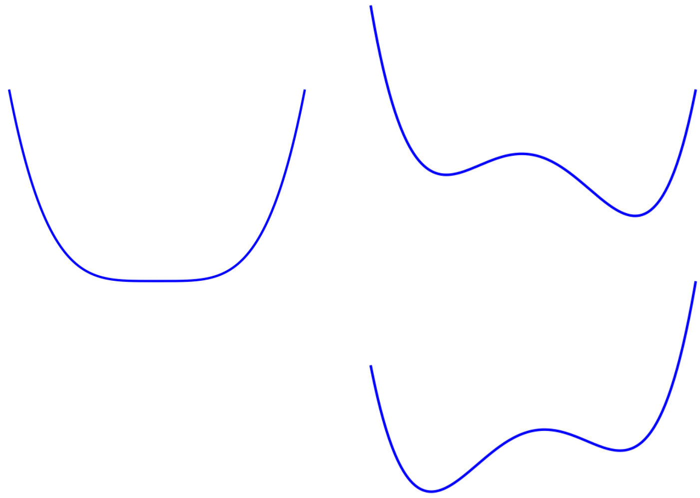
$$F_m = \sum_{\Xi_m} F_{\xi_0}, \tag{1.11}$$

then the flow is integrated (with some important cooking) for a given time, before the next integration with F_{m+1}.

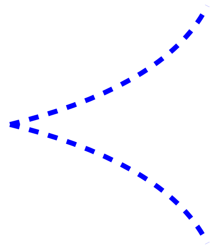
This changes nothing to the result:

Theorem 1.1. *Backpropagation is a flow of natural transformations of $\mathbb{W}$, computed from collections of singletons in $\mathbb{X}$.*

Figure 1.3 shows a bifurcation Σ in $\mathbb{W}$, $\mathbb{X} \to \mathbb{W}$. Subfigure 1.3a shows three forms of potentials for dynamics of X^w on the left part when, in the upper-right part, we can see the regions of a planar projection of $\mathbb{W}$, where the learned dynamics has the corresponding shape.



(a) Dynamics of X^w



$$\begin{tikzcd}
\mathbb{X} \arrow[d] & X^w \arrow[l, hook'] \arrow[r] \arrow[d] & X^{w_t} \arrow[d] \\
\mathbb{W} & \{w\} \arrow[l, hook'] \arrow[r] & \{w_t\}
\end{tikzcd}$$

(b) Illustration of theorem 1.1

Figure 1.3: Examples of bifurcations

Remark. Frequently, the function F takes the form of a Kullback-Leibler divergence

$$D_{KL}(P(\xi_n)|P_n)$$

and can be rewritten as a free energy, which can itself be replaced by a Bethe free energy over inner variables, which are probabilistic laws on the weights. This is where information quantities could enter [Pel20].

1.5 The specific nature of the topos of DNNs

We wonder now to what species the topos $C^\sim$ of a DNN belongs.

Definitions. *Let* $\mathbf{X}$ *denotes the set of vertices of* $\boldsymbol{\Gamma}$ *of type a or of type A (see figure 1.2). We introduce the full subcategory* $C_\mathbf{X}$ *of* C *generated by* $\mathbf{X}$.

There only exists one arrow from a vertex of type a' to a vertex of type A through $A^\star$ (but a given a' can join different $A^\star$ then different A), only one arrow from a vertex of type a to its preceding A (but A can belong to several vertices a). Moreover there exists only one arrow from a vertex c to a vertex b when b and c are on a chain in C which does not contain a fork. And no other arrows exist in $C_\mathbf{X}$. By definition of the forks, a point a (i.e. a handle) cannot join another point than its tang A, and an input or a tang A is the center of a convergent star.

Any maximal chain in $C_\mathbf{X}^{\mathrm{op}}$ joins an input entry or a A-point (i.e. a tang), to a vertex of type a' (i.e. a tip) or to an output layer. Issued from a tang A it can pass through a handle a or a tip a', because nothing forbids a tip to join a vertex b.

If x, y belong to $\mathbf{X}$, we note $x \leq y$ when there exists a morphism from x to y; then it is equivalent to write $x \rightarrow y$ in the category $\mathcal{C}_{\mathbf{X}}$.

Proposition 1.1. *(i) $\mathcal{C}_{\mathbf{X}}$ is a poset.*

(ii) Every presheaf on C induces a presheaf on $\mathcal{C}_{\mathbf{X}}$.

(iii) For every presheaf on $\mathcal{C}_{\mathbf{X}}$, there exists a unique sheaf on C which induces it.

Proof. (i) let γ_1, γ_2 be two different simple directed paths in $\mathcal{C}_{\mathbf{X}}$ going from a point z in $\mathbf{X}$ to a point x in $\mathbf{X}$, there must exists a first point y where the two paths disjoin, going to two different points y_1, y_2. This point y cannot be a handle (type a), nor an input, nor a tang (type A), then it is an output or a tip. It cannot be an output, because a fork would have been introduced here to manage the divergence. If the two points y_1, y_2 were tangs, they were the ending points of the paths, which is impossible. But at least one of them is a tang, say A_2, because a tip cannot diverge to two ordinary vertices, if not, there should be a fork here. Then one of them, say y_1, is an ordinary vertex and begins a chain, without divergence until it attains an input or a tang A_1. Therefore $A_1 = A_2$, but this gives an oriented loop in the initial graph Γ, which was excluded from the beginning for a DNN. This final argument directly forbids the existence of $x, \neq y$ with $x \leq y$ and $y \leq x$. Then $\mathcal{C}_{\mathbf{X}}$ is a poset.

(ii) is obvious.

(iii) remark that the vertices of $\mathbf{\Gamma}$ which are eliminated in $\mathbf{X}$ are the $A^{\star}$. Then consider a presheaf F on $\mathbf{X}$, the sheaf condition over C tells that $F(A^{\star})$ must be the product of the entrant $F(a'), \ldots$, then the product map $F(A) \rightarrow F(A^{\star})$ of the maps $F(A) \rightarrow F(a')$ gives a sheaf.

■

Corollary. *$C^{\sim}$ is naturally equivalent to the category of presheaves $\mathcal{C}_{\mathbf{X}}^{\wedge}$.*

Remark. In Friedman [Fri05], it was shown that every topos defined by a finite site, where objects do not possess non unit endomorphisms, has this property to be equivalent to a topos of presheaves over a finite full subcategory of the site: this is the category generated by the objects that have only the trivial full covering. Then we are in a particular case of this theorem. The special fact, that we get a site which is a poset, implies many good properties for the topos [Bel08], [Car09].

In what follows, $\mathbf{X}$ will often denote the poset $\mathcal{C}_{\mathbf{X}}$.

Definitions 1. *The (lower) Alexandrov topology on $\mathbf{X}$, is made by the subsets U of $\mathbf{X}$ such that ($y \in U$ and $x \leq y$) imply $x \in U$.*
A basis for this topology is made by the collections U_α of the β such that $\beta \leq \alpha$. In fact, consider the intersection $U_x \cap U_{x'}$; if $y \leq x$ and $y \leq x'$, we have $U_y \subseteq U_x \cap U_{x'}$, then $U_x \cap U_{x'} = \bigcup_{y \in U_x \cap U_{x'}} U_y$.

In our examples the poset $\mathbf{X}$ *is in general not stable by intersections or unions of subsets of* $\mathbf{X}$*, but the intersection and union of the sets* U_x*,* U_y *for* $x, y \in \mathbf{X}$ *plays this role.*
We note Ω *or* $\Omega(\mathbf{X})$ *when there exists a possibility of confusion, the set of (lower) open sets on* $\mathbf{X}$*.*
A sheaf in the topological sense over the Alexandrov space $\mathbf{X}$ *is a sheaf in the sense of topos over the category* $\Omega(\mathbf{X})$*, where arrows are the inclusions, equipped with the Grothendieck topology, generated by the open coverings of open sets.*

Proposition 1.2. *(see [Car18, Theorem 1.1.8, the comparison lemma] and [Bel08, p. 210]): every presheaf of sets over the category* $\mathcal{C}_{\mathbf{X}}$ *can be extended to a sheaf on* $\mathbf{X}$ *for the Alexandrov topology, and this extension is unique up to a unique isomorphism.*

Proof. Let F be a presheaf on $\mathcal{C}_{\mathbf{X}}$; for every $x \in \mathbf{X}$, $F(U_x)$ is equal to $F(x)$. For any open set $U = \bigcup_{x \in U} U_x$ we define $F(U)$ as the limit over $x \in U$ of the sets $F(x)$ (that is the set of families $s_x; x \in U$ in the sets $F(x); x \in U$, such that for any pair x, x' in U and any element y in $U_x \cap U_{x'}$, the images of s_x and $s_{x'}$ in $F(y)$ coincide. This defines a presheaf for the lower topological topology.
This presheaf is a sheaf:

1) if $\mathcal{U}$ is a covering of U, and if s, s' are two elements of $F(U)$ which give the same elements over V for all $V \in \mathcal{U}$, the elements s_x, s'_x that are defined by s and s' respectively in every $F(x)$ for $x \in U$ are the same, then by definition, $s = s'$.

2) To verify the second axiom of a sheaf, suppose that a collection s_V is defined for V in the covering $\mathcal{U}$ of U, and that for any intersection $V \cap W, V, W \in \mathcal{U}$ the restrictions of s_V and s_W coincide, then by restriction to any U_x for $x \in U$ we get a coherent section over U.

3) For the uniqueness, take a sheaf F' which extends F, and consider the open set $U = \bigcup_{x \in U} U_x$, any element s' of $F'(U)$ induces a collection $s'_x \in F(U_x) = F(x)$ which is coherent, then defines a unique element $s = f_U(s') \in F(U)$. These maps $f_U; U \in \Omega$ define the required isomorphism.

∎

Corollary. *The category* $C^{\sim}$ *is equivalent to the category* $\mathrm{Sh}(\mathbf{X})$ *of sheaves of* $\mathbf{X}$*, in the ordinary topological sense, for the (lower) Alexandrov topology.*

Consequences from [Bel08, pp.408-410]: the topos $\mathcal{E} = C^{\sim}$ of a neural network is coherent. It possesses sufficiently many points, i.e. geometric functors $\mathsf{Set} \to C^{\sim}$, such that equality of morphisms in $C^{\sim}$ can be tested on these points.
In fact, such an equality can be tested on sub-singletons, i.e. the topos is generated by the subobjects of the final object $\mathbf{1}$. This property is called *sub-extensionality* of the topos $\mathcal{E}$.
Moreover $\mathcal{E}$ (as any Grothendieck topos) is defined over the category of sets, i.e. there exists a unique geometric functor $\mu : \mathcal{E} \to \mathsf{Set}$. This functor is given by the global sections of the sheaves over $\mathbf{X}$. In this case, as shown in [Bel08], the equality of subobjects (i.e. propositions) in every object of the form

$\mu^\star(S)$ (named sub-constant objects) is decidable.
The two above properties characterize the so-called *localic topos* [Bel08], [MLM92].

The points of $\mathcal{E}$ correspond to the ordinary points of the topological space $\mathbf{X}$; they are also the points of the poset $C_\mathbf{X}$. For each such point $x \in \mathbf{X}$, the functor $\epsilon_x : \mathsf{Set} \to \mathcal{E}$ is the right adjoint of the functor sending any sheaf F to its fiber $F(x)$.

In the neural network, the minimal elements for the ordering in $\mathbf{X}$ are the output layers plus some points a' (tips), the maximal ones are the input layers, and the points of type A (tangs). However, for the standard functioning and for supervised learning, in the objects $\mathbb{X}$, $\mathbb{W}$, the fibers in A are identified with the products of the fibers in the tips a', a", ..., and play the role of transmission to the branches of type a. Therefore the feed-forward functioning does not reflect the complexity of the set Ω. The backpropagation learning algorithm also escapes this complexity.

Remarks. If A were not present in the fork, we should have added the empty covering of a in order to satisfy the axioms of a Grothendieck topology, and this would have been disastrous, implying that every sheaf must have in a the value $\star$ (singleton). A consequence is the existence of more general sheaves than the ones that correspond to usual feed-forward dynamics, because they can have a value X_A different from the product of the $X_{a'}$ appearing in $A^\star$, equipped with a map $X_{A^\star A} : X_A \to \prod X_{a'}$ and $X_{aA} : X_A \to X_a$. Then, depending on the value of $\varepsilon^0_{\text{in}}$ and of the other objects and morphisms, a propagation can happen or not. This opens the door to new types of networks, having a part of spontaneous activities (see chapter 3).

Remark. Several evidences show that the natural neuronal networks in the brain of the animals are working in this manner, with spontaneous activities, internal modulations and complex variants of supervised and unsupervised learning, involving memories, spontaneous activities, genetically and epigenetically programmed activations and desactivations, which optimize the survival at the level of the evolution of species.

Remark. Appendix A gives an interpretation due to Bell of the class of topos we encounter here, named *localic topos*, in terms of a categorical version of fuzzy sets, called *sets with fuzzy identities* taking values in a given Heyting algebra.

For the topos of a DNN, the Heyting algebra Ω is the algebra of open subsets of the poset $\mathbf{X}$. However, we can go further in the characterization of this topos by using the particular properties of the poset $\mathbf{X}$, and of the algebra Ω.

Theorem 1.2. *The poset* $\mathbf{X}$ *of a DNN is made by a finite number of trees, rooted in the maximal points and which are joined in the minimal points.*

More precisely, the *minimal* elements are of two types: the outputs layers $x_{n,j}$ and the tips of the forks, i.e. the points of type a'; the *maximal* elements are also of two types: the input layers $x_{0,i}$ and the tangs of the forks (i.e. the points A). Moreover, the tips and the tanks are joined by an irreducible arrow, but a tip can join several tanks and some ordinary point (of type a but not being an input $x_{0,i}$), and a tank can be joined by several tips and other ordinary points (but not being an output $x_{n,j}$) as it is illustrated in figure 1.4.

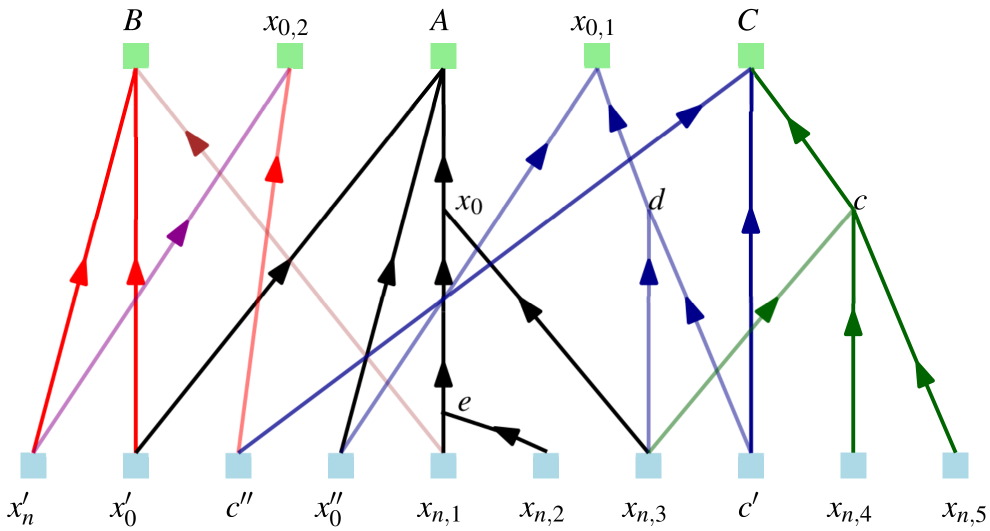


Figure 1.4: Poset of a DNN

Remark. The only possible divergences happen at tips, because they can joint several tanks and additional ordinary points in $\mathbf{X}$.

Remark. Appendix B gives an interpretation of the type of toposes we may obtain for $DNNs$ in terms of spectrum of commutative rings.

Any object in the category $C^\wedge_{\mathbf{X}}$ can be interpreted as a dynamical network, because it describes a flow of maps between sets $\{F_x; x \in \mathbf{X}\}$, along the arrows between the layers, and each of these sets can be interpreted as a space of states, not necessarily made by vectors. However what matters for the functioning of the network is the correspondence between input data, that are elements in the product F_{in} of spaces over input layers, and output states, that are elements in the product F_{out} of the spaces over output layers. This correspondence is described by the limit of F over $C_{\mathbf{X}}$, i.e. $H^0(F) = H^0(C_{\mathbf{X}}; F)$, [Mac71]. This contains the graphs of ordinary applications from F_{in} to F_{out}, when taking the products at the forks, but in general, except for chains, that are models of simple reflexes, this limit is much wider, and a source of innovation (see the above remarks on spontaneius activity and section 3.2 below).

2 Stacks of DNNs

2.1 Groupoids, general categorical invariance and logic

In many interesting cases, a restriction on the structure of the functioning X^w, or the learning in $\mathbb{W}$, comes from a geometrical or semantic invariance, which is extracted (or expected) from the input data and/or the problems that the network has to solve as output.

The most celebrate example is given by the convolutional networks $CNNs$. These networks are made for analyzing images; it can be for finding something precise in an image in a given class of images, or it can be for classifying special forms. The images are assumed to be by nature invariant by planar translation, then it is imposed to a large number of layers to accept a non trivial action of the group G of $2D$-translations and to a large number of connections between two layers to be compatible with the actions, which implies that the underlying linear part when it exists is made by convolutions with a numerical function on the plane. This does not forbid that in several layers, the action of G is trivial, to get invariant characteristics under translations, and here, the layers can be fully connected. The *Resnets* today have such a structure, with non-trivial architectures, as described in the preceding chapter.

Other Lie groups and their associated convolutions were recently used for DNNs, (see Cohen et al. [CWKW19], [CGW20], [BBCV21]). Diverse case of equivariant deep learning are presented in [SGW21]. For example, studies of graph networks involve invariance and equivariance under groupoids of isomorphisms between graphs [MBHSL19].
Cohen et al. [CWKW19] underline the analogy with Gauge theory in Physics. In the same spirit, Bondesan and Welling [BW21] give an interpretation of the excited states in DNNs in terms of particles in Quantum Field Theory.

DNNs that analyze images today, for instance in object detection, have several channels of convolutional maps, max pooling and fully connected maps, that are joint together to take a decision. It looks as a structure for localizing the translation invariance, as it happens in the successive visual areas in the brains of animals. Experiments show that in the first layers, kinds of wavelet kernels are formed spontaneously to translate contrasts, and color opposition kernels are formed to construct color invariance.

A toposic manner to encode such a situation consists to consider contravariant functors from the category C of the network with values in the topos $G^\wedge$ of G-sets, in place of taking values in the category Set of sets. Here the group G is identified with the category with one object and whose arrows are given by the elements of G, then a G-set, that is a set with a left action of G, is viewed as a set valued sheaf over G. The collection of these functors, with morphisms given by the equivariant natural transformations, form a category $C^\sim_G$, which was shown to be itself a topos by Giraud [Gir72]. We will prove this fact in the following section 2.2: there exists a category $\mathcal{F}$, which is fibred in groups isomorphic to G over C, $\pi : \mathcal{F} \to C$, and satisfies the axioms of a stack, equipped with a canonical topology J (the least fine such that π is cocontinuous [Sta, 7.20], i.e. a comorphism of site [CZ21]), in such a manner that the topos $\mathcal{E} = \mathcal{F}^\sim$ of sheaves of sets over the site $(\mathcal{F}, J)$, is naturally equivalent to the category $C^\sim_G$. This topos is named the *classifying topos* of the stack.

The construction of Giraud is more general; it extends to any stack over C, not necessarily in groups or in groupoids. In this chapter, we will consider this more general situation, given by a functor F from C^{op} to the category Cat of small categories, then corresponding to a fibred category $\mathcal{F} \to C$. But we will not consider the issue of non-trivial topologies, because, as we have shown in chapter 1, the topos of $DNNs$ are topos of presheaves. Then we determine the inner logic of the classifying topos, from fibers to fibers, to describe later the possible (optimal) flows of information in functioning networks.

The case of groupoids has the interest that the presheaves on a groupoid form a Boolean topos, then ordinary logic is automatically incorporated.

Remarks. 1) The logic in the topos of a groupoid consists of simple Boolean algebras; however, things appear more interesting when we remember the meaning of the atoms $Z_i; i \in K$, because they are made of irreducible G_a-sets. We interpret that as a part of the semantic point of view, in the languages of topos and stacks.

2) In the experiments reported in [BBG21a] as in $CNNs$, the irreducible linear representations of groups appear spontaneously among the dynamical objects.

3) In every language we can talk of the future, the uncertain past, and introduce hypotheses, this does not mean that we are leaving the world of usual Boolean logic, we are just considering externally some intuitionist Heyting algebra, this can be done within ordinary set theory, as is done topos theory in Mathematics, in the fibers, defined by groupoids.

Appendix C gives a description of the classifying object of a groupoid, that is well known by specialists of category theory.

However, other logics, intuitionist, can also have an interest. In more recent experiments done with Xavier Giraud on data representing time evolution, we used simple posets in the fibers.

The notion of invariance goes further than groupoids.

Invariance is synonymous of action (like group action), and is understood here in the categorical sense: a category $\mathcal{G}$ acts on another category $\mathcal{V}$ when a (contravariant) functor from $\mathcal{G}$ to $\mathcal{V}$ is given. The example that justifies this terminology is when $\mathcal{G}$ is a group G, and $\mathcal{V}$ the Abelian category of vector spaces and linear maps over a commutative field $\mathbb{K}$. In the latter case, we obtain a linear representation of the group G.

In any category $\mathcal{V}$, there exists a notion which generalizes the notion of *element* of a set. Any morphism $\varphi : u \to v$ in $\mathcal{V}$ can be viewed as an *element* of the object v of $\mathcal{V}$.

Definition. *Suppose that $\mathcal{G}$ acts through the functor $f : \mathcal{G} \to \mathcal{V}$ and that $v = f(a)$, then the* orbit *of φ under $\mathcal{G}|a$ is the functor from the left slice category $\mathcal{G}|a$ to the right slice category $u|\mathcal{V}$, that associates to any morphism $a' \to a$ the element $u \to f(a) \to f(a')$ of $f(a')$ in $\mathcal{V}$ and to an arrow $a" \to a'$ over a the corresponding morphism $f(a') \to f(a")$, from $u \to f(a')$ to $u \to f(a")$.*

In the classical example of a group representation, $u = \mathbb{K}$ and the morphism φ defines a vector x in the space V_e. The group G is identified with $G|e$ and the vector space V_e, identified with $\mathsf{Hom}(K, V_e)$, contains the whole orbit of x.

In a stack, the notion of action of categories is extended to the notion of fibred action of a fibred category $\mathcal{F}$ to a fibred category $\mathcal{N}$:

Definition. *Suppose we are given a sheaf of categories $F : C \to \mathsf{Cat}$, that we consider as a general structure of invariance, and another sheaf $M : C \to \mathsf{Cat}$. An action of F on M is a family of contravariant functors $f_U : \mathcal{F}_U \to \mathcal{M}_U$ such that, for any morphism $\alpha : U \to U'$ of C, we have*

$$f_U \circ F_\alpha = M_\alpha \circ f_{U'}. \tag{2.1}$$

This is the equivariance formula *generalizing group equivariance as it can be found in [Kon18] for instance. It is equivalent to morphisms of stacks, and allows to define the orbits of sections $u_U \to f_U(\xi_U)$ in the sheaf $u|\mathcal{M}$ under the action of the relative stack $\mathcal{F}|\xi$.*

Remark that Eilenberg and MacLane, when they invented categories and functors in [EM45], were conscious to generalize the Klein's program in Geometry (Erlangen program).

In the next sections, we will introduce languages with types taken from presheaves over the fibers of the stack, where we define the terms of theories and propositions of interest for the functioning of the DNN. Then the above notion of invariance will concern the action of a kind of pre-semantic categories on the languages and the possible sets of theories, that the network could use and express in functioning.

This view is a crucial point for our applications of topos theory to DNNs, because it is in this framework that logical reasoning, and more generally semantics, in the neural network, can be set: in a stack, the different layers interpret the logical propositions and the sentences of the output layers. As we will see, the interpretations are expected to become more and more faithful when approaching the output, however the information flow in the whole networks is interesting by itself.

This shift from groups to groupoids, then to categories, then to more general semantic, by taking presheaves in groupoids or categories, is a fundamental addition to the site C. The true topos associated to a network is the classifying topos $\mathcal{E}$ over $\mathcal{F}$; it incorporates much more structure than the visible architecture of layers, it takes into account invariance (which appears here to be part of the semantic, or better pre-semantic). More generally, it can concern the domain of natural human semantics that the network has to understand in his own artificial world.

Moreover, as we will show below, working in this setting gives access to more flexible type theories, like the Martin-Löf intensional types, and goes into the direction of homotopy type theory according to Hofmann and Streicher [HS98], Hollander [Hol01], Arndt and Kapulkin [AK11], enlarged by objects and morphisms in classifying topos in the sense of Giraud.

2.2 Objects classifiers of the fibers of a classifying topos

In this section we study the propagation of logical theories through a stack (equipped with a scindage in the sense of Giraud). In particular we find a sufficient condition for free propagation downstream and upstream, that was apparently not described before; it asks that gluing functors are fibrations, plus a supplementary geometrical condition, always satisfied in the case of groupoids, (see theorem 2.1).

The application to the dynamics of functioning $DNNs$ is presented in the next section 2.3, with the notion of semantic functioning. It is developed in the next chapter 3. Examples are presented in the following chapters 4, with long and short term memory cells and variants of them, and their tentative relation with cognitive linguistic, then in 5, with more general networks moduli.

In the general case and in more canonical toposic terms, the logic in the stack $\mathcal{F}$ over C is studied by Olivia Caramello and Riccardo Zanfa [CZ21]; see also the available notes written for "Topos Online", 24-30 june 2021.

Also the contributions of Shulman, [Shu10], [Shu19], and the slides of his talk "Large categories and quantifiers in topos theory", January 26 2021, Cambridge Category Seminar are of interest.

Among the equivalent points of view on stacks and classifying topos [Gir64], [Gir71], and [Gir72]), the most concrete one starts with a contravariant functor F from the category C to the 2-category of small categories Cat. (This corresponds to an element of the category $\mathsf{Scind}(C)$ in the book of Giraud [Gir71].) To each object $U \in C$ is associated a small category $\mathcal{F}(U)$, and to each morphism $\alpha : U \to U'$ is associated a covariant functor $F_\alpha : F(U') \to F(U)$, also denoted $F(\alpha)$, satisfying the axioms of a presheaf over C. If $f_U : \xi \to \eta$ is a morphism in $F(U)$, the functor F_α sends it to a morphism $F_\alpha(f_U) : F_\alpha(\xi) \to F_\alpha(\eta)$ in $F(U')$.

The corresponding fibration $\pi : \mathcal{F} \to C$, written ∇F by Grothendieck, has for objects the pairs (U,ξ) where $U \in C$ and $\xi \in F(U)$, sometimes shortly written ξ_U, and for morphisms the elements of

$$\mathsf{Hom}_{\mathcal{F}}((U,\xi),(U',\xi')) = \bigcup_{\alpha \in \mathsf{Hom}_C(U,U')} \mathsf{Hom}_{F(U)}(\xi, F(\alpha)\xi'). \tag{2.2}$$

For every morphism $\alpha : U \to U'$ of C, the set $\mathsf{Hom}_{F(U)}(\xi, F(\alpha)\xi')$ is also denoted $\mathsf{Hom}_\alpha((U,\xi),(U',\xi'))$; it is the subset of morphisms in $\mathcal{F}$ that lift α.

The functor π sends (U,ξ) on U. We will write indifferently $F(U)$ or $\mathcal{F}_U$ the fiber $\pi^{-1}(U)$.

A section s of π corresponds to a family $s_U \in \mathcal{F}_U$ indexed by $U \in C$, and a family of morphisms $s_\alpha \in \mathsf{Hom}_{F(U)}(s_U, F(\alpha)s_{U'})$ indexed by $\alpha \in \mathsf{Hom}_C(U,U')$ such that, for any pair of compatible morphisms α,β, we have

$$s_{\alpha\circ\beta} = F_\beta(s_\alpha) \circ s_\beta. \tag{2.3}$$

As shown by Grothendieck and Giraud [Gir64], a presheaf A over $\mathcal{F}$ corresponds to a family of presheaves A_U on the categories $\mathcal{F}_U$ indexed by $U \in C$, and a family A_α indexed by $\alpha \in \mathsf{Hom}_C(U,U')$, of natural transformations from $A_{U'}$ to $F_\alpha^\star A_U$. (Here $F_\alpha^\star$ denotes the pullback of presheaf associated to the functor $F_\alpha : F(U') \to F(U)$, that is, for $A_U : F(U) \to \mathsf{Set}$, the composed functor $A_U \circ F_\alpha$.)
Moreover, for any compatible morphisms $\beta : V \to U$, $\alpha : U \to U'$, we must have

$$A_{\alpha\circ\beta} = F_\alpha^\star(A_\beta) \circ A_\alpha. \tag{2.4}$$

If ξ is an object of $\mathcal{F}_U$, we define $A(U,\xi) = A_U(\xi)$, and if $f : \xi_U \to F_\alpha \xi'_{U'}$ is a morphism of $\mathcal{F}$ between $\xi_U \in \mathcal{F}_U$ and $\xi'_{U'} \in \mathcal{F}_{U'}$ lifting α, we take

$$A(f) = A_U(f) \circ A_\alpha : A_{U'}(\xi') \to A_U(F_\alpha(\xi')) \to A_U(\xi). \tag{2.5}$$

The relation $A(f \circ g) = A(g) \circ A(f)$ follows from (2.4).

A natural transformation $\varphi : A \to A'$ corresponds to a family of natural transformations

$$\varphi_U : A_U \to A'_U,$$

such that, for any arrow $\alpha : U \to U'$ in C,

$$F_\alpha^\star \varphi_U \circ A_\alpha = A'_\alpha \circ \varphi_{U'} : A_{U'} \to F_\alpha^\star A'_U. \tag{2.6}$$

This describes the category $\mathcal{E}$ of presheaves over $\mathcal{F}$ from the family of categories $\mathcal{E}_U$ of presheaves over the fibers $\mathcal{F}_U$ and the family of functors $F_\alpha^\star : \mathcal{E}_U \to \mathcal{E}_{U'}$.
Note that for two consecutive morphisms $\beta : V \to U$, $\alpha : U \to U'$, we have $F_{\alpha\beta}^\star = F_\alpha^\star \circ F_\beta^\star$.

The category $\mathcal{E}$ is fibred over the category C, it corresponds to the functor E from C to Cat, which associates to $U \in C$ the category $\mathcal{E}_U$ and to an arrow $\alpha : U \to U'$, the functor $F_!^\alpha : \mathcal{E}_{U'} \to \mathcal{E}_U$, which is the *left adjoint* of $F_\alpha^\star$. This functor extends F_α through the Yoneda embedding, [AGV63, Chap. I, Presheaves].
For two consecutive morphisms $\beta : V \to U$, $\alpha : U \to U'$, we have $F_!^{\alpha\beta} = F_!^\beta \circ F_!^\alpha$.

Let $\eta_\alpha : F_!^\alpha \circ F_\alpha^\star \to Id_{\mathcal{E}_U}$ the counit of the adjunction; a natural transformation $A_\alpha : A_{U'} \to F_\alpha^\star A_U$ gives a natural transformation $A_\alpha^\star : F_!^\alpha A_{U'} \to A_U$, by taking $A_\alpha^\star = (\eta_\alpha \otimes Id) F_!^\alpha(A_\alpha)$. This gives another way to describe the elements of $\mathcal{E}$, through the presheaves over $\mathcal{F}$.

Remark. A section (s_U, s_α) defines a presheaf A, by taking

$$A_U(\xi) = \mathsf{Hom}_{\mathcal{F}_U}(\xi, s_U); \tag{2.7}$$

and $A_\alpha = s_\alpha^\star \circ F_\alpha$, according to the following sequence:

$$\mathsf{Hom}(\xi', s_{U'}) \to \mathsf{Hom}(F_\alpha \xi', F_\alpha(s_{U'})) \to \mathsf{Hom}(F_\alpha \xi', s_U). \tag{2.8}$$

The identity (2.4) follows from the identity (2.3).
This construction generalizes in the fibered situation the Yoneda objects in the absolute situation.
A morphism of sections gives a morphism of presheaves.

In each topos $\mathcal{E}_U$ there exists a classifying object $\mathbf{\Omega}_U$, such that the natural transformations $\mathsf{Hom}_U(X_U, \mathbf{\Omega}_U)$ naturally correspond to the subobjects of X_U; the presheaf $\mathbf{\Omega}_U$ has for value in $\xi_U \in \mathcal{F}_U$ the set of subobjects in $\mathcal{E}_U$ of the Yoneda presheaf $\xi_U^\wedge$ defined by $\eta \mapsto \mathsf{Hom}(\eta, \xi_U)$, with morphisms given by composition to the right.
The set $\mathbf{\Omega}_U(\xi_U)$ can also be identified with the set of subobjects of the final sheaf $\mathbf{1}_{\xi_U}$ over the slice category $\mathcal{F}_U|\xi_U$.

Remark. In general, the object of parts $\mathbf{\Omega}^X$ of an object X in a presheaf topos $\mathcal{D}^\wedge$ over a category $\mathcal{D}$, is the presheaf given in $x \in \mathcal{D}_0$ by the set of subsets of the product set $(\mathcal{D}|x) \times X(x)$ and by the maps induced by $X(f)$ for $f \in \mathcal{D}_1$. Observe that $\mathbf{\Omega}^X$ realizes an equilibrium between the category of basis through $\mathcal{D}|x$ and the set theoretic nature of the value $X(x)$.
A special case is when $X = \mathbf{1}$, the final object, made by a singleton $\star$ at each $x \in \mathcal{D}_0$, and the unique possible maps for $f \in C_1$. The presheaf $\mathbf{\Omega}^\mathbf{1}$ is denoted by $\mathbf{\Omega}$. Its value in $x \in \mathcal{D}_0$, is the set $\mathbf{\Omega}_x$ of subsets of the Yoneda object $x^\wedge$.
It can be proved that a subobject Y of an object X of $\mathcal{D}^\wedge$ corresponds to a unique morphism $\chi_Y : X \to \mathbf{\Omega}$

such that at any $x \in \mathcal{D}_0$, we have $Y(x) = \chi_Y^{-1}(\top)$.
The exponential presheaf $\mathbf{\Omega}^X$ is characterized by the natural family of bijections

$$\mathsf{Hom}_{\mathcal{D}^\wedge}(Y \times X, \mathbf{\Omega}) \approx \mathsf{Hom}_{\mathcal{D}^\wedge}(Y, \mathbf{\Omega}^X), \tag{2.9}$$

which expresses the universal property of the classifier $\mathbf{\Omega}$.
We will also frequently consider the set of subobjects of $\mathbf{1}$ over the whole category $\mathcal{D}$, and we simply denote it by the letter Ω. It is named the Heyting algebra of the topos $C^\wedge$. See appendix A for more details.

As just said before, the functor $F_\alpha^\star : \mathcal{E}_U \to \mathcal{E}_{U'}$ which associates $A \circ F_\alpha$ to A, possesses a left adjoint $F_!^\alpha : \mathcal{E}_{U'} \to \mathcal{E}_U$ which extends the functor F_α on the Yoneda objects. For any object ξ' in $\mathcal{F}_{U'}$, note $\xi = F_\alpha(\xi')$; the functor $F_!^\alpha$ sends $(\xi')^\wedge$ to $\xi^\wedge$, and sends a subset of $(\xi')^\wedge$ to a subset of $\xi^\wedge$. This is not because $F_!^\alpha$ is necessarily left exact, but because we are working with Grothendieck topos, where subobjects are given by families of coherent subsets.
Moreover $F_!^\alpha$ respects the ordering between these subsets, then it induces a poset morphism between the posets of subobjects

$$\Omega_\alpha(\xi') : \Omega_{U'}(\xi') \to \Omega_U(F_\alpha(\xi')) = F_\alpha^\star \Omega_U(\xi'); \tag{2.10}$$

the functoriality of Ω_U, $\Omega_{U'}$ and F_α implies that these maps constitute a natural transformation between presheaves

$$\Omega_\alpha : \Omega_{U'} \to F_\alpha^\star \Omega_U. \tag{2.11}$$

The naturalness of the construction insures the formula (2.4) for the composition of morphisms. Consequently, we obtain a presheaf $\mathbf{\Omega}_{\mathcal{F}}$.

Moreover the final object $\mathbf{1}_{\mathcal{F}}$ of the classifying topos $\mathcal{E} = \mathcal{F}^\wedge$ corresponds to the collection of final objects $\mathbf{1}_U; U \in C$ and to the collection of morphisms $\mathbf{1}_{U'} \to F_\alpha^\star \mathbf{1}_U; \alpha \in \mathsf{Hom}_C(U, U')$, then we have:

Proposition 2.1. *The classifier of the classifying topos is the sheaf $\mathbf{\Omega}_{\mathcal{F}}$ given by the classifiers Ω_U and the pullback morphisms Ω_α, which can be summarized by the formula*

$$\mathbf{\Omega}_{\mathcal{F}} = \nabla_{U \in C} \Omega_U d\Omega_\alpha. \tag{2.12}$$

In general the functor $F_\alpha^\star$ is not *geometric*; by definition, it is so if and only if its left adjoint

$$(F_\alpha)_! = (F^\alpha)_!$$

, which is right exact (i.e. commutes with the finite colimits), is also *left exact* (i.e. commutes with the finite limits). Also by definition, this is the case if and only if the morphism F_α is a *morphism of sites* from $\mathcal{F}_U$ to $\mathcal{F}_{U'}$, [AGV63, IV 4.9.1.1.], not to be confused with a comorphism, [AGV63, III.2], [CZ21].
Important for us: it results from the work of Giraud in [Gir72], that $F_\alpha^\star$ is geometric when F_α is itself a stack, and when finite limits exist in the sites $\mathcal{F}_U$ and $\mathcal{F}_{U'}$ and are preserved by F_α. (We will see in the next section, that these stacks $\mathcal{F} \to C$, made by stacks between fibers, correspond to some admissible

contexts in a dependent type theory, when C is the site of a DNN.)
When $F_\alpha^\star$ is geometric, a great part of the logic in $\mathcal{E}_{U'}$ can be transported to $\mathcal{E}_U$:

Let us write $f = F_\alpha^\star$ and $f^\star = (F_\alpha)_!$ its left adjoint, supposed to be left exact, therefore exact, as just mentionned. This functor $f^\star$ preserves the monomorphisms, and the final elements of the slices categories. Then it induces a map between the sets of subsets, called the inverse image or pullback by f, for any object $X' \in \mathcal{E}_{U'}$:

$$f^\star : \mathrm{Sub}(X') \to \mathrm{Sub}(f^\star X'). \tag{2.13}$$

When X' describes the Yoneda objects $(\xi')^\wedge$, this gives the morphism $\Omega_\alpha : \Omega_{U'} \to F_\alpha^\star \Omega_U$.
As it is shown in MacLane-Moerdijk [MLM92, p. 496], this map is a morphism of lattices, it preserves the ordering and the operations $\wedge$ and $\vee$. If $h : Y' \to X'$ is a morphism in $\mathcal{E}_{U'}$, the reciprocal image $h^\star$ between the sets of subsets has a left adjoint $\exists_h$ and a right adjoint $\forall_h$. The morphism $f^\star$ commutes with $\exists_h$, but in general not with $\forall_h$, for which there is only an inclusion:

$$f^\star(\forall_h P') \leq \forall_{f^\star h}(f^\star P'). \tag{2.14}$$

To have an equality, the morphism f must be geometric and *open*. This is equivalent to the existence of a left adjoint, in the sense of posets morphisms, for Ω_α, [MLM92, Theorem 3, p. 498].
In MacLaneMoerdijk1992, this natural transformation Ω_α is denoted λ_α, and its left adjoint when it exists is denoted μ_α.

When this left adjoint in the sense of Heyting algebras exists, we have, by adjunction, the counit and unit morphisms:

$$\mu \circ \lambda \leq \mathsf{Id} : \Omega_{U'} \to \Omega_{U'}; \tag{2.15}$$

$$\lambda \circ \mu \geq \mathsf{Id} : F^\star \Omega_U \to F^\star \Omega_U. \tag{2.16}$$

If f is geometric and open, the map $f^\star$ also commutes with the negation $\neg$ and with the (internal) implication $\Rightarrow$.
If openness fails, only inequality (external implication) holds for the universal quantifier.

Remark. When $\mathcal{F}_{U'}$ and $\mathcal{F}_U$ are the posets of open sets of (sober) topological spaces $\mathcal{X}'$ and $\mathcal{X}$, and when F_α is given by the direct image of a continuous *open* map $\varphi_\alpha : \mathcal{X}_{U'} \to \mathcal{X}_U$, the functor $F_\alpha^\star$ is geometric and open. This extends to locale, [MLM92].

When $F_\alpha^\star$ is geometric and open, it transports the predicate calculus of formal theories from $\mathcal{E}_{U'}$ to $\mathcal{E}_U$, as exposed in the book of Mac Lane and Moerdijk, [MLM92]. This is expressed by the following result,

Proposition 2.2. *Suppose that all the $F_\alpha; \alpha : U \to U'$ are open morphisms of sites (in the direction from $F(U)$ to $F(U')$), then,*

(i) the pullback Ω_α commutes with all the operations of predicate calculus;

(ii) *any theory at a layer U', i.e. in $\mathcal{E}_{U'}$, can be read and translated in a deeper layer U, in $\mathcal{E}_U$, in particular at the output layers.*

In the sequence we will be particularly interested by the case where all the $\mathcal{F}_U$ are groupoids and the F_α are morphisms of groupoids, in this case, the algebras of subobjects $\mathrm{Sub}_{\mathcal{E}}(X)$ are boolean, then, in this case, the following lemma implies that, as soon as $F_\alpha^\star$ is geometric, it is open:

Lemma 2.1. *In the boolean case the morphism of lattices $f^\star : \mathrm{Sub}(X') \to \mathrm{Sub}(f^\star X')$ is a morphism of algebras which commutes with the universal quantifiers $\forall_h$.*

Proof. Since $f^\star$ is right and left exact, it sends $0 = \perp$ to $0 = \perp$ and $X' = \top$ to $X = \top$. Therefore, for every $A \in \mathrm{Sub}_{\mathcal{E}'}(X')$, $f^\star(X' \setminus A') = X \setminus f^\star(A')$, i.e. $f^\star$ commutes with the negation $\neg$. This negation establishes a duality between $\exists$ and $\forall$, then $f^\star$ commutes with the universal quantifier. More precisely:

$$f^\star(\neg(\forall x', P'(x'))) = f^\star(\exists a', \neg P'(a')) = \exists a, f^\star(\neg P')(a) = \neg(\forall x f^\star(P')(x)), \tag{2.17}$$

then by commutation with $\neg$, and $\neg\neg = \mathsf{Id}$, we have

$$f^\star(\forall x', P'(x')) = \forall x, f^\star(P')(x). \tag{2.18}$$

■

Let us mention here a difficulty: in the case of groups or groupoids, $F_\alpha^\star$ is geometric if and only if F_α is an equivalence of categories (then an isomorphism in the case of groups). This is because a morphism of group is flat if and only if it is an isomorphism, [AGV63, 4.5.1.]. The main problem is with the preservation of products.

However, it is remarkable that for any kind of group homomorphisms $F : G' \to G$, in every algebra of subobjects the map $f^\star$ induced by $F_!$ preserves "locally" and "naturally" all the logical operations:

Lemma 2.2. *For every object X' in $B_{G'}$, note $X = F_!(X')$, then $f^\star$ induces a map of lattices $f^\star$: $\mathrm{Sub}(X') \to \mathrm{Sub}(X)$, that is bijective. It preserves the order $\leq$, the elements $\top$ and $\perp$, and the operations $\wedge$ and $\vee$, therefore it is a morphism of Heyting algebras. Moreover, for any natural transformation $h : Y' \to X'$, it commutes with both the existential quantifier $\exists_h$ and the universal quantifiers $\forall_h$.*

Proof. As said in [AGV63, 4.5.1], if $F : G' \to G$ is a morphism of groups, the functor $F_!$ from $B_{G'}$ to B_G is given on X' by the *contracted product* $F_!(X') = G \times_{G'} X'$, that is the set of orbits of the action of G' on the G-set $G_d \times X'$.
The algebra $\mathrm{Sub}(X')$ is the boolean algebra generated by the primitive representations of G' on the orbits $G'x'$ of the elements of X'. But each orbit $G'x'$ is sent in X to an orbit of G, that is the product of $G/F(H'_{x'})$ with the singleton $\{G'x'\}$, where $H'_{x'}$ is the stabilizer of x'. These sets describe the orbits of the action of G on X, then the elements of $\mathrm{Sub}(X)$.
The commutativity with $h^\star$ for a G'-morphism $h : Y' \to X'$ is evident, the rest follows from the bijection property, orbitwise. ■

Therefore, even if $F^\star$ is not a geometric morphism, it is legitimate to say that in some sense, it is *open*, because all logical properties are preserved by the induced morphisms between the local Heyting algebras. We could say that $F^\star$ is "weakly geometric and open".

This can be easily extended to morphisms of groupoids. The left adjoint $F_!$ admits a description which is analogous to the contracted product of groups. Lemma 2.2 holds true. The only difference is that $f^\star$ is not a bijection, but it is a surjection when F is surjective on the objects. More details and the generalization of the above results to fibrations of categories that are themselves fibrations in groupoids over posets will be given in the text *Search of semantic spaces*.

In the reverse direction of the flow, it is important that a proposition in the fiber over U can be understood over U'.

Hopefully, this can always be done, at least in part: the functor $F_\alpha^\star$ is left exact and has a right adjoint $F_\star^\alpha : \mathcal{E}_{U'} \to \mathcal{E}_U$, which can be described as a right Kan extension [AGV63]: for a presheaf A' over $\mathcal{F}_{U'}$, the value of the presheaf $F_\star^\alpha(A'_{U'})$ at $\xi_U \in \mathcal{F}_U$ is the limit of $A'_{U'}$ over the slice category $F_\alpha|\xi_U$, whose objects are the pairs (η', φ) where $\eta' \in \mathcal{F}_{U'}$ and $\varphi : F_\alpha(\eta') \to \xi_U$ is a morphism in $\mathcal{F}_U$, and whose morphisms from (η', φ) to (ζ', ϕ) are the morphisms $u : \eta' \to \zeta'$ such that $\varphi = \phi \circ F_\alpha(u)$.
Therefore, if we denote ρ the forgetting functor from $F_\alpha|\xi_U$ to $\mathcal{F}_{U'}$, we have

$$F_\star^\alpha(A')(\xi_U) = H^0(F_\alpha|\xi_U; \rho^\star A'), \tag{2.19}$$

that is the set of sections of the presheaf $\rho^\star A'$ over the slice category.

Remark. In the case where $F_\alpha : \mathcal{F}_{U'} \to \mathcal{F}_U$ is a morphism of groupoids, this set is the set of sections of A' over the connected components of $F_\alpha^{-1}(\xi_U)$.

Therefore the functor $g = F_\star^\alpha$ is always geometric in our situation of presheaves. By definition, this proves that F_α is a comorphism of sites. Consequently, as shown in [MLM92], the pullback of subobjects defines a natural transformation of presheaves over $\mathcal{F}_{U'}$:

$$\lambda'_\alpha : \Omega_U \to F_\star^\alpha \Omega_{U'}; \tag{2.20}$$

which corresponds by the adjunction of functors $F_\alpha^\star \dashv F_\star^\alpha$, to a natural transformation of sheaves over $\mathcal{F}_U$:

$$\tau'_\alpha : F_\alpha^\star \Omega_U \to \Omega_{U'}. \tag{2.21}$$

Lemma 2.3. *If F_α is a fibration (not necessarily in groupoids), it is an open morphism of sites, and the functor $F_\star^\alpha$ is open [Gir72].*

Proof. This results directly from [MLM92, Proposition 1, pp. 509-513]. Precisely this proposition says that a morphism of sites $F : \mathcal{F}' \to \mathcal{F}$ induces an open geometric morphism $F_\star : \mathsf{Sh}(\mathcal{F}', J') \to \mathsf{Sh}(\mathcal{F}, J)$ between the categories of sheaves, as soon as the following three conditions are satisfied:

(i) F has the property of lifting of the coverings:

$$\forall \xi' \in \mathcal{F}', \forall S \in J(F(\xi')), \exists T' \in J'(\xi'), F(T') \subseteq S; \tag{2.22}$$

where $F(T')$ is the sieve generated by the images of the arrows in T';

(ii) F preserves the covers, i.e.

$$\forall \xi' \in \mathcal{F}', \forall S' \in J'(\xi'), F(S') \in J(F(\xi')); \tag{2.23}$$

(iii) for every $\xi' \in \mathcal{F}'$, the sliced morphism $F|\xi' : \mathcal{F}'|\xi' \to \mathcal{F}|F(\xi')$ is surjective on the objects.

The two first conditions are true for the canonical topology of a stack [Gir72]. They are obvious in our case of presheaves. Condition (iii) is part of the definition of fibration (pre-fibration). ■

If in addition F itself is surjective on the objects, as it will be the case in our applications, the maps of algebras $g_X^\star : \mathrm{Sub}(X) \to \mathrm{Sub}(f^\star X)$ are injective and the geometric open morphism $g = F_\star$ is surjective on the objects [MLM92, page 513].

Lemma 2.4. *When F_α is a fibration, the relation between $\lambda_\alpha = \Omega_\alpha : \Omega_{U'} \to F_\alpha^\star \Omega_U$ and $\lambda'_\alpha : \Omega_U \to F_\star^\alpha \Omega_{U'}$, is given by the adjunction of posets morphisms:*

$$\Omega_\alpha \dashv \tau'_\alpha; \tag{2.24}$$

where $\tau'_\alpha : F_\alpha^\star \Omega_U \to \Omega_{U'}$ is the dual of λ'_α.
The morphism Ω_α is the left adjoint of the morphism τ'_α. Moreover, τ'_α is an injective section of the surjective morphism Ω_α.

Proof. If F_α is a fibration, $F_\alpha^\star \Omega_U$ is isomorphic to Ω_U, it is the sub-algebra of $\Omega_{U'}$ formed by the subobjects of $\mathbf{1}_{U'}$ that are invariant by F_α, i.e. by $\lambda_\alpha : \Omega_{U'} \to F_\alpha^\star \Omega_U$.
The map τ'_α associates to an element P of Ω_U the element $P \circ F_\alpha$, seen as a sub-sheaf of $1_{U'}$, that is an element of $\Omega_{U'}$ saturated by F_α. Therefore, for every $P' \in \Omega_{U'}$, the element $\tau'_\alpha \circ \lambda_\alpha(P')$ of $\Omega_{U'}$ is the saturation of P', then it contains P'. This gives a natural transformation

$$\eta : \mathsf{Id}_{\Omega_{U'}} \to \tau'_\alpha \circ \Omega_\alpha. \tag{2.25}$$

In the other direction, τ'_α is a section over $\Omega_{U'}$ of the map λ_α, i.e. $\Omega_\alpha \circ \tau'_\alpha = Id_{F_\alpha^\star \Omega_U}$. Which gives a natural transformation

$$\epsilon : \Omega_\alpha \circ \tau'_\alpha \to \mathsf{Id}_{F_\alpha^\star \Omega_U}. \tag{2.26}$$

In the following lines, we forget the indices α everywhere, and show that η and ϵ are respectively the unit and counit of an adjunction of posets morphisms.
Let P' and Q, be respectively elements of $\Omega_{U'}$ and Ω_U, if we have a morphism from $\lambda P'$ to Q, by applying

τ', we obtain a morphism from $\tau' \circ \lambda P'$ to $\tau' Q$, then a morphism from P' to $\tau' Q$. All that is equivalent to the following implications:

$$(\lambda P' \leq Q) \Rrightarrow (P' \leq \tau' \lambda P' \leq \tau' Q). \tag{2.27}$$

In the other direction,

$$(P' \leq \tau' Q) \Rrightarrow (\lambda P' \leq \lambda \tau' P' \leq Q). \tag{2.28}$$

Therefore

$$(P' \leq \tau' Q) \Lleftarrow\Rrightarrow (\lambda P' \leq Q). \tag{2.29}$$

Which is the statement of lemma 2.4. ■

From the above lemmas, we conclude the following result (central for us):

Theorem 2.1. *When for each $\alpha : U \to U'$ in C, the functor F_α is a fibration, the logical formulas and their truth in the topos propagate from U to U' by λ'_α (feedback propagation in the DNN), and if in addition F_α is a morphism of groupoids (surjective on objects and morphisms), the logic in the topos also propagates from U' to U, by λ_α (feed-forward functioning in the DNN).*
Moreover, the map λ_α is the left adjoint of the transpose τ'_α of the map λ'_α. And we have, for any $\alpha : U \to U'$ in C,

$$\lambda_\alpha \circ {}^t\lambda'_\alpha = \mathsf{Id}_{\Omega_{U'}}. \tag{2.30}$$

Definition 2.1. *When the conclusion of the above theorem holds true, even if the F_α are not fibrations, we say that the stack $\pi : \mathcal{F} \to C$ satisfies the* strong standard hypothesis *(for logical propagation). Without the equation* (2.30)*, we simply say that the standard hypothesis is satisfied.*

In this case, the logic is richer in U' than in U, like a fibration of Heyting algebras of subobjects of objects.

To finish this section, let us describe the relation between the classifier $\Omega_{\mathcal{F}}$ and the classifier Ω_C of the basis category C of the fibration $\pi : \mathcal{F} \to C$.

As reminded above, proposition 2.1 in [Gir71], gives sufficient conditions for guarantying that the functor $\pi^\star$ is geometric. But, even in the non-geometric case, when the fibers are groupoids, the morphism has locally (at the level of subobjects) the logical properties of an open geometric morphism, (see lemmas 2.1 and 2.2) and lemma 2.3 says that the functor $\pi_\star$, which is its right adjoint, is geometric and open. We can then apply lemma 2.4, and get an adjunction $\lambda_\pi \dashv \tau'_\pi$, where

$$\lambda_\pi : \Omega_{\mathcal{F}} \to \pi^\star \Omega_C, \tag{2.31}$$

is a surjective morphism of lattices, and

$$\tau'_\pi : \pi^\star \Omega_C \to \Omega_F, \tag{2.32}$$

is the section by invariant objects.

When π is fibration of groupoids, $\pi^\star$ is open, and λ_π is a morphism of Heyting algebras. In this case, there exists a perfect lifting of the theories in C to the theories in $\mathcal{F}$.

2.3 Theories, interpretation, inference and deduction

Main references are Bell [Bel08], Lambek and Scott [LS81], [LS88] , MacLane and Mœrdijk [MLM92].

The formal languages, that we will mainly consider, are the typed languages of type theory, in the sense of Lambek and Scott [LS81]. In particular, in such a type theory we have a notion of deduction, conditioned by a set S of propositions, named axioms, which is denoted by $\vdash_S$. This is a relation between two propositions, $P \vdash_S Q$, which satisfies the usual axioms, structural, logical, and set theoretical, also named rules of inference, of the form

$$(P_1 \vdash_S Q_1, P_2 \vdash_S Q_2, ..., P_n \vdash_S Q_n)/P \vdash_S Q, \tag{2.33}$$

meaning that the truth (or validity) of the left (said upper) conjunction of deductions implies the truth of the right deduction (said lower).
The conditional validity of a proposition R is noted $\vdash_S R$.

A (valid) proof of $\vdash_S R$ is an oriented classical graph without oriented cycles, whose vertices are labelled by valid inferences, and whose oriented edges are identifying one of the upper terms of its final extremity to the lower term of its initial extremity, and having only one final vertex whose lower term is $\vdash_S R$. The initial vertices have left terms that are empty or belonging to the set S.
A theory $\mathbb{T}$ in a formal language $\mathbb{L}$ is the set of propositions that can be asserted to be true if some axioms are assumed to be true, this means that these propositions are deduced by valid proofs from the axioms.

A language $\mathbb{L}$ is interpreted in a topos $\mathcal{E}$ when some objects of $\mathcal{E}$ are associated to every type, the object $\Omega_{\mathcal{E}}$ corresponding to the logical type $\Omega_{\mathbb{L}}$, when some arrows $A \to B$ are associated to the variables (or terms) of B in the context A, all that being compatible with the respective definitions of products, subsets, exponentials, singleton, changes of contexts (substitutions), and logical rules, including the predicate calculus, which includes the two projections (existential and universal) on the side of topos [Bel08], [LS81].

A theory $\mathbb{T}$ is represented in $\mathcal{E}$ when all its axioms are true in $\mathcal{E}$. The fact that all the deductions are valid in $\mathcal{E}$ is the statement of the *soundness theorem* of $\mathbb{T}$ in $\mathcal{E}$.

Remark. The *completeness theorem* says that, for any language and any theory, there exists a minimal "elementary topos" $\mathcal{E}_\mathbb{T}$, which in general is not a Grothendieck topos, where the converse of the soundness theorem is true; validity in $\mathcal{E}_\mathbb{T}$ implies validity in $\mathbb{T}$. The different interpretations in a topos $\mathcal{E}$ of a theory

$\mathbb{T}$ form a category $\mathcal{M}(\mathbb{T},\mathcal{E})$, which is equivalent to the category of "logical functors" from $\mathcal{E}_{\mathbb{T}}$ to $\mathcal{E}$. This equivalence needs precisions given by Lambek and Scott, in particular to fix representant of subobjects, which is automatic in a Grothendieck topos.

As suggested by Lambek, an interpretation of a type theory in a topos constitutes a *semantic* of this theory.

If a formal language $\mathbb{L}$ can be interpreted in a topos $\mathcal{E}$, and if $F:\mathcal{E}\to\mathcal{F}$ is a left exact functor from $\mathcal{E}$ to a topos $\mathcal{F}$, the interpretation is transferred to $\mathcal{F}$. The condition for transporting any theory $\mathbb{T}$ by f is that it admits a right adjoint $f:\mathcal{F}\to\mathcal{E}$ which is geometric and open.
A geometric functor allows the transportation of the restricted family of geometric theories as in [Car09], [Car18] or [MLM92].

Remark. If $\mathbb{T}$ is a geometric theory, there is a Grothendieck topos $\mathcal{E}'_{\mathbb{T}}$ which classifies the interpretations of $\mathbb{T}$, i.e. for every Grothendieck topos $\mathcal{E}$ the category of geometric functors from $\mathcal{E}$ to $\mathcal{E}'_{\mathbb{T}}$ is equivalent to $\mathcal{M}(\mathbb{T},\mathcal{E})$ [Car09], [Car18],[MLM92]. A *logical functor* is the left adjoint of a *geometric functor*.

In many applications of $DNNs$, a network has to proceed to a semantic analysis of some data. Our aim now is to precise what this means, and how we, observers, can have access to the internal process of this analysis.

As before, the network is presented as a dynamic object $\mathbb{X}$ in a topos, with learning object of weights $\mathbb{W}$, and the considered topos $\mathcal{E}$ is the classifying topos of a fibration $\pi:\mathcal{F}\to C$.

In the applications, the logic is richer in U' than in U when there is a morphism $\alpha:U\to U'$ in C. We suppose given a family of typed language $\mathbb{L}_U;U\in C$, interpreted in the topos $\mathcal{E}_U;U\in C$ of the corresponding layers.
We say that the functors $f=g^\star=F_\alpha^\star$ propagate these languages backward, when for each morphism $\alpha:U\to U'$ in C, there exists a natural transformation

$$\mathbb{L}_\alpha:\mathbb{L}_{U'}\to F_\alpha^\star\mathbb{L}_U, \tag{2.34}$$

which extends $\Omega_\alpha=\lambda_\alpha$, implying that the types define objects or morphisms in $\mathcal{E}$, in particular 0_U, 1_U.
And we say that the left adjoint functor $f^\star$ propagates the languages feed-forward, when for each morphism $\alpha:U\to U'$ in C, there exists a natural transformation

$$\mathbb{L}'_\alpha:\mathbb{L}_U\to F_\star^\alpha\mathbb{L}_{U'}, \tag{2.35}$$

which extends λ'_α, implying that the types define objects or morphisms in the fibration $\mathcal{E}'$, defined by the right adjoint functors $F_\star^\alpha$.

We assume that the standard hypothesis 2.1 is satisfied for the extensions $\mathbb{L}_\alpha$ and $\mathbb{L}'_\alpha$.

Note that in the case of stacks of DNNs, there exist two kinds of functors $F_\alpha:\mathcal{F}_{U'}\to\mathcal{F}_U$ over C, the ordinary ones, flowing from the input to the output, and the added canonical projections from the fiber at a fork A to the fibers of their tines a', a”, ... The second kind of functors are canonically fibrations, but for the other functors, this is a condition we can require for a good semantic functioning (see theorem 2.1).

Let $\mathbb{L}$ denote the corresponding presheaf in languages over $\mathcal{C}$, $\Omega_{\mathbb{L}}$ its logical type, and for each $U \in \mathcal{C}$, we note $\Omega_{\mathbb{L}_U}$ the value of this logical type at U. For each $U \in \mathcal{C}$, we write Θ_U the set of possible sets of axioms in $\mathbb{L}_U$, that is $\Theta_U = \mathcal{P}(\Omega_{\mathbb{L}_U})$. This is also the set of theories.

We take as output (*resp.* input) the union of the output (*resp.* output) layers. In supervised and reinforcement learning, we can tell that, for every input $\xi_{\text{in}} \in \Xi_{\text{in}}$ in a set of inputs for learning, a theory $\mathbb{T}_{\text{out}}(\xi)$ in $\mathbb{L}_{\text{out}}$ is imposed at the output of the network., i.e. some propositions are asked to be true, other are asked to be false.
The set of theories in the language $\mathbb{L}_{\text{out}}$ is denoted Θ_{out}. Then the objectives of the functioning is a map $\mathbb{T}_{\text{out}} : \Xi_{\text{in}} \to \Theta_{\text{out}}$.

Definition. *A* semantic functioning *of the dynamic object X^w of possible activities in the network, with respect to the mapping $\mathbb{T}_{\text{out}}$, is a family of quotient sets D_U of X^w_U, $U \in C$, equipped with a map $S_U : D_U \to \Theta_U$, such that for every $\xi_{\text{in}} \in \Xi_{\text{in}}$ and every $U \in C$, the image $S_U(\xi_U)$ generates a theory which is coherent with $\mathbb{T}_{\text{out}}(\xi_{\text{in}})$, for the transport in both directions along any path.*

Remark. In the known applications, the richer logic relies on a richer language with more propositions and less axioms, present near the input layers, but the opposite happens to expressed theories; they are more constrained in the deepest layers, with more axioms in general.

In the examples we know [BBG21a], the quotient D_U (from *discretized cells*) is given by the activity of some special neurons in the layer L_U, which saturate at a finite number of values, associated to propositions in the Heyting algebras $\Omega_{\mathbb{L}_U}$. In this case, the definition of semantic functioning can be made more concrete: for each neuron $a \in L_U$, each quantized value of activity ϵ_a implies the validity of a proposition $P_a(\epsilon_a)$ in $\Omega_{\mathbb{L}_U}$; this defines the map S_U. Then the definition of semantic functioning asks that, for each input $\xi_{\text{in}} \in \Xi_{\text{in}}$, the generated activity defines values $\epsilon_a(\xi_{\text{in}})$ of the special neurons, such that the generated set of propositions $P_a(\epsilon_a)$, implies the validity of a given proposition in $\Omega_{\mathbb{L}_{\text{out}}}$, which is valid for $\mathbb{T}_{\text{out}}(\xi_{\text{in}})$.

In particular, we saw experimentally that the inner layers understand the language $\mathbb{L}_{\text{out}}$, which is an indication that the functors $f = g^\star = F^\star_\alpha$ propagate the languages backward.

This gives a crude notion of *logical information* of a given layer, or any subset E of neurons in the union of the sets D_U: it is the set of propositions predicted to hold true in $\mathbb{T}_{\text{out}}(\xi_{\text{in}})$ by the activities in E. If all the involved sets are finite, the amount of information given by the set E can be defined as the ratio of the number of predicted propositions over the number of wanted decisions, and a mean of this ratio can be taken over the entries ξ_{in}.

Remark. The above notion of semantic functioning and semantic information can be extended to sets of global activities Ξ, singletons sections of X^w, more general that the ones used for learning.

Our experiments in [BBG21a] have shown that the number of hidden layers, or the complexity of the architecture, strongly influences the nature of the semantic functioning. This implies that the semantic functioning, then the corresponding accessible semantic information, depend on the characteristics of the dynamic X^w, for instance the non-linearities for saturation and quantization, and of the characteristics of the learning, the influence of the non-linearities of the gradient of backpropagation on the optimal weights $w \in W$. Therefore, it appears a notion of *semantic learning*, which is a flow of natural transformations between dynamic objects X^{w_t}, increasing the semantic information.

In the mentioned experiments, the semantic behavior appears only for sufficiently deep networks, and for non-linear activities.

2.4 The model category of a DNN and its Martin-Löf type theory

In this section, we study the collection of stacks over a given layers architecture, with fibers in a given category, as groupoids, and we show that it possesses a natural structure of closed model category of Quillen, giving both a theory of homotopy and an intensional type theory, where the above stacks with free logical propagation, described by theorem 2.1, correspond respectively to fibrant objects and admissible contexts.

Consider two fibrations $(\mathcal{F}_U, F_\alpha)$ and $(\mathcal{F}'_U, F'_\alpha)$ over C; a morphism φ from the first to the second is given by a collection of functors $\varphi_U : \mathcal{F}_U \to \mathcal{F}'_U$ such that for any arrow $\alpha : U \to U'$ of C, $\varphi_U \circ F_\alpha = F'_\alpha \circ \varphi_{U'}$. With the fibrations in groupoids, this gives a category Grpd_C. Natural transformations between two morphisms give it a structure of strict 2-category.

We consider this category fibred over $C_\mathbf{X}$. Remind that the Grothendieck topology on $C_\mathbf{X}$ that we consider is chaotic [AGV63]. If we consider an equivalent site, with a non-trivial topology, homotopical constraints appear for defining stacks [Gir72], [Hol08]. However the category of stacks (*resp.* stacks in groupoids) is equivalent to the category obtained from $C_\mathbf{X}$.

Hofmann and Streicher [HS98], have proved that the category Grpd of groupoids gives rise to a Martin-Löf type theory [ML80], by taking for types the fibrations in groupoids, for terms their sections, for substitutions the pullbacks, and they have defined non-trivial (non-extensional) identity types in this theory.

Hollander [Hol01], [Hol08], using Giraud's work and homotopy limits, constructed a Quillen model theory on the category of fibrations (*resp.* stacks) in groupoids over any site C, where the fibrant objects are the stacks, the cofibrant objects are generators, and the weak equivalences are the homotopy equivalence in the fibers (see also Joyal-Tierney and Jardine cited in Hollander [Hol08]). These results were extended to the category of general stacks, not only in groupoids, over a site by Stanculescu [Sta14].

Awodey and Warren [AW09] observed that the construction of Hofmann-Streicher is based on the most natural closed model category structure in the sense of Quillen on Grpd, and proposed an extension of the construction to more general model categories. Thus they established the connection between Quillen's models and Martin-Löf intensional theories, which was soon extended to a connection between more elaborate Quillen's models and Voedvosky univalent theory.

Arndt and Kapulkin, in *Homotopy Theoretic Models of Type Theory* [AK11], have proposed additional axioms on a closed model theory that are sufficient to formally deduce a Martin-Löf theory. This was extended later by Kapulkin and Lumsdaine [KLV12], to obtain models of Voedvosky theory, by using more simplicial techniques. Here, we will follow their approach, without going to the special properties of HoTT, that are functions extensionality, Univalence axiom and Higher inductive type formations.

In what follows, we focus on the model structure of groupoids and stacks in groupoids, which are the most useful models for our applications. However, many things also work with Cat in place of Grpd, and some other model categories $\mathcal{M}$. The complication is due to the difference between fibrations (*resp.* stacks) in the sense of Giraud and Grothendieck and the fibrations in the sense of Quillen's models, which is not the case with groupoids. For Cat, there exists a unique closed model structure, defined by Joyal and Tierney, such that the weak equivalences are the equivalence of categories [SP12]. It is named for this reason the *canonical model structure on* Cat; in this structure, the cofibrations are the functors injective on objects and the fibrations are the *isofibrations*. An isofibration is a functor $F : \mathcal{A} \to \mathcal{B}$, such that every isomorphism of $\mathcal{B}$ can be lifted to an isomorphism of $\mathcal{A}$. Any fibration of category is an iso-fibration, but the converse is true only for groupoids. A different model theory was defined by Thomason [Tho80], which is better understandable in terms of ∞-groupoids and ∞-categories.

The axioms of Quillen [Qui67] concern three subsets of morphisms in a category $\mathcal{M}$, supposed to be (at least finitely) complete and cocomplete, the set Fib of fibrations, the set Cofib of cofibrations and the set WE of weak equivalences. An object A of $\mathcal{M}$ is said fibrant (*resp.* cofibrant) if $A \to \mathbf{1}$, the final object (*resp.* $\emptyset \to A$ from the initial object) is a fibration (*resp.* a cofibration).

Definitions. *Two morphisms $i : A \to B$ and $p : C \to D$ in a category are said* orthogonal*, written (non-traditionally) $i \times p$, if for any pair of morphisms $u : A \to C$ and $v : B \to D$, such that $p \circ u = v \circ i$, there exists a morphism $j : B \to C$ such that $j \circ i = u$ and $p \circ j = v$. The morphism j is named a lifting, left lifting of i and a right lifting of p.*

Two sets $\mathcal{L}$ and $\mathcal{R}$ are said be the orthogonal one of each other if $i \in \mathcal{L}$ is equivalent to $\forall p \in \mathcal{R}, i \times p$ and $p \in \mathcal{R}$ is equivalent to $\forall i \in \mathcal{L}, i \times p$.

The three axioms of Quillen for a closed category $\mathcal{M}$ of models are:

1) given two morphisms $f : A \to B$, $g : B \to C$, define $h = g \circ f$; if two of the morphisms f, g, h belong to WE, then the third one belongs to WE;

2) every morphism f is a composition $f = p \circ i$ of an element p of Fib and an element i of Cofib $\cap$ WE, and a composition $p' \circ i'$ of an element p' of Fib $\cap$ WE and an element i' of Cofib;

3) the sets Fib and $\mathsf{Cofib} \cap \mathsf{WE}$ are the orthogonal one of each other and the sets $\mathsf{Fib} \cap \mathsf{WE}$ and Cofib also.

An element of $\mathsf{Fib} \cap \mathsf{WE}$ is named a *trivial fibration*, and an element of $\mathsf{Cofib} \cap \mathsf{WE}$ is named a *trivial cofibration*.

These axioms (and some more general) allowed Quillen to develop a convenient homotopy theory in $\mathcal{M}$, and to define a homotopy category $Ho\mathcal{M}$ (see his book, *Homotopical Algebra*, [Qui67]). The objects of $Ho\mathcal{M}$ are the fibrant and cofibrant objects of $\mathcal{M}$, and its morphisms are the homotopy classes of morphisms in $\mathcal{M}$; two morphisms f, g from A to B are homotopic if there exists an object A', equipped with a weak equivalence $\sigma : A' \to A$ and two morhisms i_0, i_1 from A to A' such that $\sigma \circ i_0 = \sigma \circ i_1$, and a morphism $h : A' \to B$, such that $h \circ i_0 = f$ and $h \circ i_1 = g$. In the category $Ho\mathcal{M}$, the weak equivalences of $\mathcal{M}$ are inverted.

A particular example is the category of sets with surjections as fibrations, injections as cofibrations and all maps as equivalences. Another trivial structure, which exists for any category is no restriction for Fib and Cofib but isomorphisms for WE.
As we already said, an important example is the category of groupoids Grpd, with the usual fibrations in groupoids, with all the functors injective on the objects as cofibrations, and the usual homotopy equivalence (i.e. here category equivalence) as weak equivalences.
We also mentioned the canonical structure on Cat, that is the only one where weak homotopy corresponds to the usual equivalence of category.
Other fundamental examples are the topological spaces Top and the simplicial sets $\mathsf{SSet} = \Delta^\wedge$, with Serre and Kan fibrations for Fib respectively.
The closed model theory of Thomason 1980 [Tho80] on Cat is deduced by the above structure on SSet, by using the nerve construction and the square of the right adjoint functor f the barycentric subdivision. In this structure the weak equivalences are not reduced to the category equivalences and the cofibrant objects are constrained [Cis06]; this theory is weakly equivalent to the Kan structure on SSet. Then in this structure, a category is considered through its weak homotopy type (the weak homotopy type of its nerve).

We now call on a general result of Lurie 's book, [Lur09, appendix A.2.8, prop. A.2.8.2], which establishes the existence of two canonical closed model structures on the category of functors $\mathcal{M}_C = \mathsf{Fun}(C^{\mathsf{op}}, \mathcal{M})$ when $\mathcal{M}$ is a model category. (Caution, Lurie consider diagrams, i.e. C and not C^{op}.) An additional hypothesis is made on $\mathcal{M}$, that it is combinatorial in the sense of Smith (see Rosicky in [rR09]), i.e. locally presentable (i.e. accessible by a regular cardinal), and generated by cofibrant objects, which are both satisfied by Grpd and by Cat. Moreover $\mathcal{M}$ is supposed to have all small limits and small colimits, which is also the case for Grpd (or Cat); as Set, both are cartesian closed categories; every object is fibrant and cofibrant.

The two Lurie structures are respectively obtained by defining the sets Fib or Cofib in the fiberwise manner, as for the set WE, and by taking respectively the set Cofib or Fib of morphism which satisfy the required lifting properties, respectively on the left and on the right, i.e. the orthogonality of Quillen.
The structure obtained by fixing Fib (*resp.* Cofib) by the behavior in the fibers, is named the *projective* structure, or *right* one (*resp.* the *injective* one, or *left* one).

Caution: depending on the authors, the term right and left may be exchanged.
The model structure of Hollander on Grpd_C (or Stanculescu for Cat_C) is the right Lurie model. She called this model a left model.

A model category is said *right proper* when the pullback of any weak equivalence along an element of Fib is again a weak equivalence. Dually, *left proper* is when push-forward of weak equivalence along cofibrations is again in WE.
In the right proper case, the injective (left) structure of Lurie was defined before by D-C. Cisinski in "*Images directes cohomologiques dans les catégories de modèles*" [Cis03].

The cofibrations in the right model (*resp.* the fibrations in the left model) depend on the category C. They certainly deserve to be better understood.
See the discussion of Cisinski, in his book *Higher Categories and Homotopical Algebra*, [Cis19, section 2.3.10].

Proposition 2.3. *If C has sufficiently many points, the elements of* Fib *for the left Lurie structure are fibrations in the fibers (i.e. elements of* Fib *for the right structure) and the elements of* Cofib *for the right structure are injective on the objects in the fibers (i.e. elements of* Cofib *for the left structure.*

Proof. Suppose that a morphism φ is right orthogonal to any trivial cofibration ψ of the left Lurie structure; for every point x in C, this gives an orthogonality in the model Grpd, then over x, φ_x induces a fibration in groupoids. From the hypothesis, this implies that in every fiber over C, φ is a fibration, then an element of Fib for the right Lurie structure.
The other case is analog. ∎

However in general, even if C is a poset, not all fibrations in the fibers are in Fib for the left model structure, and not all the injective in fibers are in Cofib for the right model. This was apparent in Hollander [Hol01].

Trying to determine the obstruction for a local fibration (*resp.* local cofibration) to be orthogonal to functors that are locally injective on the objects (*resp.* local fibrations) and locally homotopy equivalence, we see that the intuitionistic structure of Ω_C enters the game, through the global constraints on the complement of presheaves:

Lemma 2.5. *The category C being the oriented segment $1 \to 0$ and the category $\mathcal{M}$ being* Set *(then $\mathcal{M}_C$ is the topos of the Shadoks [Pro08]); in the left Lurie model the fibrant objects are the (non-empty) surjective maps $f : F_0 \to F_1$.*

Proof. A trivial cofibration is a natural transformation

$$\eta : (h : H_0 \to H_1) \to (h' : H'_0 \to H'_1); \tag{2.36}$$

such that η_0 and η_1 are injective.

Suppose given a natural transformation $u = (u_0, u_1)$ from h to $f : F_0 \to F_1$; the lifting problem is the extension of u to u' from h' to f. If H_1 is empty, there is no problem. If not, we choose a point $\star_0$ in H_0 and note $\star_1 = h(\star_0)$. If $x'_1 \in H'_1$ does'nt belong to H_1 we define $u'_1(x'_1) = u_1(\star_1)$, and for any x'_0 such that $h'(x'_0) = x'_1$, we define $u'_0(x'_0) = u_0(\star_0)$. Now the problem comes with the points $x"_0$ in $H'_0 \backslash H_0$ such that $h'(x"_0) \in H_1$ (a shadok with an egg); their image by u_1 is defined, then $u'_1(h'(x"_0))$ is forced to be in the image of F_0 by f. If f is not surjective there exists η such that the lifting is impossible. But, if f is surjective there is no obstruction: we define $u'_0(x"_0)$ to be any point y_0 in F_0 such that $f(y_0) = u_1(h'(x"_0))$ in F_1. ∎

Lemma 2.6. *Also $\mathcal{M} = \mathsf{Set}$, but C being the (confluence) category $\bigwedge$ with three objects $0, 1, 2$ and two non-trivial arrows $1 \to 0$ and $2 \to 0$. In the left Lurie model, the fibrant objects are the pairs $(f_1 : F_0 \to F_1, f_2 : F_0 \to F_2)$, such that the product map (f_1, f_2) is surjective.*

Proof. Following the path of the preceding proof, with an injective transformation η from a triple H_0, H_1, H_2 to a triple H'_0, H'_1, H'_2, we are in trouble with the elements $x"_0 \in H'_0$ that h'_1 or h'_2 sends into H_1 or H_2 respectively. Under the hypothesis of bi-surjectivity, we know where to define $u'_0(x"_0)$. But if this hypothesis is not satisfied, impossibility happen in general for η. ∎

Lemma 2.7. *Also $\mathcal{M} = \mathsf{Set}$, but C being the (divergence) category $\bigvee$ with three objects $0, 1, 2$ and two non-trivial arrows $0 \to 1$ and $0 \to 2$. In the left Lurie model, the fibrant objects are the pairs $(f_1 : F_1 \to F_0, f_2 : F_2 \to F_0)$, such that separately f_1 and f_2 are surjective.*

Proof. following the path of the preceding proof, with an injective transformation η from a triple H_0, H_1, H_2 to a triple H'_0, H'_1, H'_2, we are in trouble with the elements $x"_1 \in H'_1$ (*resp.* $x"_2 \in H'_2$) that h'_1 (*resp.* h'_2) sends into H_0. As in the proff of the lemma 1, the problem is solved under the hypothesis of surjectivity, but it cannot be solved without it. ∎

More generally, we can determine the fibrant objects of the left Lurie model (injective) for every closed model category $\mathcal{M}$, and a finite poset C which has the structure of a DNN, coming with a graph, with unique directed paths:

Theorem 2.2. *When C is the poset of a DNN, for any combinatorial category of model, the fibrations of $\mathcal{M}_C$ for the injective (left) model structure are made by the natural transformations $\mathcal{F} \to \mathcal{F}'$ between functors in C to $\mathcal{M}$, that induce fibrations in $\mathcal{M}$ at each object of C, such that the functor $\mathcal{F}$ is also a fibration in $\mathcal{M}$ along each arrow of C coming from an internal of minimal vertex (ordinary vertex, output or tip), and a fibration along each of the arrows issued from a minimal vertex (output and tip), and a multi-fibration at each confluence point, in particular at the maximal vertices (input or tank).*

By *multi-fibration* $f_i, i \in I$ from an object F_A of $\mathcal{M}$ to a family of objects $F_i, i \in I$ of $\mathcal{M}$, we mean a fibration (element of Fib) from F_A to the product $\prod_{i \in I} F_i$.

Proof. We proceed by recurrence on the number of vertices. For an isolated vertex, this is the definition of fibration in $\mathcal{M}$. Then consider an initial vertex (tank or input) A with incoming arrows $s_i : i \to A$ for $i \in I$ in the graph poset C, and note $C^\star$ the category with the star A, s_i deleted. A trivial cofibration in $\mathcal{M}_C$ is a natural transformation $\eta; \mathcal{H} \to \mathcal{H}'$ between contravariant functors in $C \to \mathcal{M}$, which is at each vertex injective on objects and an element of $\mathcal{WE}$. Let us consider a morphism (u, u') in in $\mathcal{M}_C$ from η to a morphism $\varphi : \mathcal{F} \to \mathcal{F}'$, where $\mathcal{F}$ belongs to $\mathcal{M}_C$.

Suppose that φ satisfies the hypotheses of the theorem. From the recurrence hypothesis, there exists a lifting $\theta^\star : (\mathcal{H}')^\star \to \mathcal{F}^\star$ between the restrictions of the functors to $C^\star$; it is in particular defined on the objects $H'_i, i \in I$ to the objects $F_i, i \in I$.
Consider the functor from H'_A to the product $\prod_i F_i$, which is obtained by composing the horizontal arrows of η, from H'_A to the product $H' = \prod_i H'_i$ with θ'. The fact that $F_A \to \prod_i F_i$ is a multi-fibration in $\mathcal{M}$ and the fact that $\eta_A : H_A \to H'_A$ is a trivial cofibration in $\mathcal{M}$ imply the existence of a lifting $\theta_A : H'_A \to F_A$, which is given on H_A.

Conversely, if the hypothesis of multi-fibration is not satisfied, there exists elements $\eta_A : H_A \to H'_A$ in $\mathsf{Cofib} \cap \mathsf{WE}$ of M, such that the lifting θ_A of H'_A to F_A does'nt exist, by the axiom (3) of closed models. To finish the proof, we note that the necessity to be a fibration at each vertex in C is given by proposition 2.3. ∎

Corollary. *Under the same hypotheses, the fibrant objects of $\mathcal{M}_C$ for the injective (left) model structure are made by the functors that are a fibration in $\mathcal{M}$ at each internal of minimal vertex (ordinary vertex, output or tip), and a fibrant object at the minimal (output and tip), and a multi-fibration at each confluence point (see lemma 2.7), in particular at the maximal vertices (input or tank).*

One interest of this result is that it will describe the allowed contexts in the associated Martin-Löf theory when it exists, as we will see just below.
Another interest is for the behavior of the classifying object $\mathbf{\Omega}_{\mathcal{F}}$: in the case of Grpd_C the fibrant objects are all good for the induction theory in logic over the network (see theorem 2.1). In the case of Cat_C, with the canonical structure, we will see below that it is not the case, only a subclass of fibrant objects are good, which are made by composition of Giraud-Grothendieck fibrations.
Last by not least, this corollary allows to enter the homotopy theory of the stacks, according to Quillen [Qui67], because it associates objects up to homotopy with the stacks that have a fluid semantic functioning as in theorem 2.1.

In Grpd_C the final object $\mathbf{1}$ (*resp.* the initial object $\emptyset$) is the constant functor on C with values a singleton, (*resp.* the empty set). It follows that any object is cofibrant.

The additional axioms of Arndt and Kapulkin for a *Logical Model Theory* are as follows:

(1) for any element $f \in \mathsf{Fib}$, $f : B \to A$, the pullback functor $f^\star : \mathcal{M}|A \to \mathcal{M}|B$, once restricted to the fibrations, possesses a right-adjoint, denoted Π_f.

(2) The pullback of a trivial cofibration, i.e. an element of $\mathsf{Cofib} \cap \mathsf{WE}$, along an element of Fib is again a trivial cofibration.

Remark. In Arndt and Kapulkin [AK11], the first axiom is written without the restriction of the adjunction to fibrations, however they remark later [AK11, section 4.1, acknowledging an anonymous reviewer] that this restricted axiom is sufficient for the application below.

The second axiom is satisfied if separately Cofib and WE are stable by pullback along a fibration. As we already said, a model category satisfying the second property for WE is called *right proper*.

When every object in $\mathcal{M}$ is fibrant (*resp.* cofibrant) the theory is right (*resp.* left) proper [Hir03]. This is the case for Grpd (and Cat). And Lurie proved that his two model structures on diagrams (or phe-sheaves) are right (reps. left) proper as soon as $\mathcal{M}$ is so. Then in our case, all the considered models are right proper and left proper. This was shown by Hollander [Hol01] for Grpd_C.

The injectivity on objects in the fibers and the equivalence of category in the fibers are preserved by every pullback, thus condition (2) is satisfied for the left injective structure. This is the structure we choose. What happens to the right structure?
Arndt and Kapulkin noticed the example of the injective structure [AK11, Prop. 27, p.12] and its Bousfield-Kan localizations; this gives in particular the injective model structures for the category of stacks over any site (see Hirschhorn, *Localization of Model Categories* [Hir03]).

The existence of a right adjoint and a left adjoint of the pullback of fibrations in categories, as it holds for presheaves of sets, was proved by Giraud in 1964 [Gir64, section I.2.].

Then, by proposition 2.3, for $\mathcal{M} = \mathsf{Grpd}$, both left and right structures satisfy the condition (1). For $\mathcal{M} = \mathsf{Cat}$ this is true only if f is a fibration in the geometric sense, not only an isofibration. What happens to other models categories $\mathcal{M}$?

As noticed by Arndt and Kapulkin, the left adjoint of $f^\star : \mathcal{M}|A \to \mathcal{M}|B$ always exists, it is written Σ_f, and the right properness implies that it respects WE.

If $\mathcal{M}$ satisfies the axioms (1) and (2), Arndt and Kapulkin generalized the constructions of Seely [See84], Hofmann and Streicher [HS98], and Awodey–Warren [AW09] to define a M-L theory:

A *context* is a fibration $\Gamma \to C$, that is a fibrant object. A *type* $\mathcal{A}$ in this context is a fibration $\mathcal{A} \to \Gamma$. The declaration (judgment) of a type is written $\Gamma \vdash \mathcal{A}$. A *term* $a : A$ is a section $\Gamma \to \mathcal{A}$. It is denoted

$\Gamma \vdash a : \mathcal{A}$.
A *substitution* x/a is given by a change of base $F^{\star}$ for a functor $F : \Delta \to \Gamma$ in $\mathcal{M}_C$, not necessarily a fibration.
The adjoint functor Σ_f and Π_f of $f^{\star}$, allows to define new types of objects: given Γ and $f : \mathcal{A} \to \Gamma$, and $g : \mathcal{B} \to \mathcal{A}$, we get $\Sigma_f(g) : \Sigma_{x:\mathcal{A}}\mathcal{B}(x) \to \Gamma$ and $\Pi_f(g) : \Pi_{x:\mathcal{A}}\mathcal{B}(x) \to \Gamma$. They respectively replace the union over $\mathcal{A}$ and the product over $\mathcal{A}$.

On the types, logical operations are applied, $\mathcal{A} \wedge \mathcal{B}$, $\mathcal{A} \vee \mathcal{B}$, $\mathcal{A} \Rightarrow \mathcal{B}$, $\perp$ is empty, $\exists x, B(x)$, $\forall x, B(x)$. The rules for these operations satisfy the usual axioms.
More types, like the integers or the real numbers or the well ordering can be added, with specific rules.

As remarked by Arndt and Kapulkin, it is not necessary to have a fully closed model theory to get a Martin-Löf type theory [AK11, remarks pp. 12-15]. They noticed that $M - L$ type theories are probably associated to fibration-categories (or categories with fibrant objects) in the sense of Brown [Bro73] (see also [Uem17]). In these categories, cofibrations are not considered, however a nice homotopy theory can be developed.

We have the following result concerning the weak factorization system made by cofibrations and trivial fibrations in the canonical model Cat:

Lemma 2.8. *A canonical trivial fibration in* Cat *is a geometric fibration.*

Proof. Consider an isofibration $f : \mathcal{A} \to \mathcal{B}$ that is also an equivalence of category. Take $a \in \mathcal{A}$ and $f(a) = b \in \mathcal{B}$ and a morphism $\varphi : b' \to b$ of $\mathcal{B}$; because f is surjective on the objects, there exists $a' \in A$ such that $f(a") = b'$, and because f is an equivalence the map from $\mathsf{Hom}(a', a)$ to $\mathsf{Hom}(b", b)$ is a bijection, then there exists a unique morphism $\psi : a' \to a$ such that $f(\psi) = \varphi$. In the same manner, every morphism $b" \to b'$ has a unique lift $a" \to a'$, and conversely any morphism $\psi' : a" \to a'$ defines a composed morphism $\chi : a" \to a$ and a morphism image $\varphi' : b" \to b'$ that define the same morphism $\varphi \circ \varphi"$ from $b"$ to b. As the morphisms from $a"$ to a are identified by f with the morphisms from $b"$ to b, this gives a natural bijection between the morphisms ψ' from $a"$ to a' and the pairs (χ, φ') in $\mathsf{Hom}(a", a) \times \mathsf{Hom}(b", b')$ over the same element in $\mathsf{Hom}(b", b)$. Therefore ψ is a strong cartesian morphism over φ. ∎

The same proof shows that a canonical trivial fibration is a geometric op-fibration, that is by definition a fibration between the opposite categories.

In the case where C is the poset of a DNN and $\mathcal{M}$ is the category Cat, we say that a model fibration $f : A \to B$, in $\mathcal{M}_C$ is a *geometric fibration* if it is a Grothendieck-Giraud fibration, and if all the iso-fibrations that constitute the fibrant object A are Grothendieck-Giraud fibrations (see theorem 2.2).

Theorem 2.3. *Let C be a poset of DNN, there exists a canonical $M - L$ structure where contexts and types correspond to the geometric fibrations in the 2-category of contravariant functors* Cat_C*, and such that base change substitutions correspond to its 1-morphisms.*

Proof. We follow the lines of Arndt and Kapulkin [AK11, theorem 26]. The main point is to prove that if $f : A \to B$ is a geometric fibration in $\mathcal{M}_C$, the pullback functor $f^\star : \mathsf{Cat}|A \to \mathsf{Cat}_B$, has a left adjoint $f_! = \Sigma_f$ and a right adjoint $f_\star = \Pi_f$ that both preserve the geometric fibrations. For the first case it is the stability of Grothendieck-Giraud fibrations by composition. For the second one, this is Giraud theorem of bi-adjunction [Gir71]. ■

There exist several equivalent interpretations of such a type theory, as for the intuitionistic theory of Bell, Lambek Scott et al. (see Martin-Löf, *Intuitionistic Type Theory*, [ML80]). For instance the types are sets, the terms are elements, or a type is a proposition and a term is a proof, or a type is a problem (a task) and a term is a method for solving it. (For each interpretation, things are local over a context.)
In particular, *Identity types* are admitted, representing equivalence of elements, proofs or methods that are not strict equalities, like homotopies, or invertible natural equivalences.
The types of identities, as in Hofmann and Streicher [HS98], are fibrations $\mathsf{Id}_A : I_{\mathcal{A}} \to \mathcal{A} \times \mathcal{A}$ equipped with a cofibration $r : \mathcal{A} \to I_{\mathcal{A}}$ (with a section) such that $\mathsf{Id}_A \circ r = \Delta$, the diagonal morphism. They are considered as paths spaces.
For instance, given a groupoid A, $\mathsf{Id}_A = (\{0 \leftrightarrow 1\} \Rightarrow A = A^{\{0 \leftrightarrow 1\}}$ is an identity type.
Axioms of inference for the types are expressed by rules of formation, introduction and determination, specific to each type [ML80].

Let us compare to the semantics in a topos $C^\wedge$: a context is an object Γ which is a presheaf with values in Set, so a fibration in sets over C and a type is another object A; to get something over Γ we can consider the projection $\Gamma \times A \to \Gamma$. A section corresponds to a morphism $a : \Gamma \to A$, which is rightly a term of type A, $\Gamma \vdash a : \mathcal{A}$.
A substitution corresponds to a morphism $F : \Delta \to \Gamma$, and defines a pullback of trivial fibrations $\Delta \times A \to \Delta$. If we have a morphism $g : B \to \Gamma \times A$ in the topos, we can define its existential image $\exists_\pi g(B)$ and its universal image $\forall_\pi g(B)$ as subobjects of Γ, which can be seen as a trivial fibrations over Γ.

Therefore, we have analogs of M-L type theory in Set theory, but with trivial fibrations only and without fibrant restriction.

2.5 Classifying the M-L theory ?

In what precedes the category Grpd has replaced the category Set; it is also cartesian closed. Also we have seen that all small limits and colimits exist in Grpd_C (Giraud, Hollander, Lurie). However every natural transformation between two functors with values in Grpd is invertible. Thus in the 2-category, the morphisms in $\mathsf{Hom}_{\mathsf{Grpd}}(G, G')$ are like homotopies. In fact they become homotopies when passing to the nerves.

Let us introduce the categories of presheaves on every fibration in groupoids $\mathcal{A} \to C$, i.e. the classifying topos $\mathcal{E}_{\mathcal{A}}$ of the stack $\mathcal{A}$. Their objects are fibered in groupoids over C, because the fibers $\mathcal{E}_U$ for $U \in C$ are such (they take their values in IsoSet), but their morphisms, the natural transformations between functors, are taken in the sense of sets, not invertible.

In what follows we combine the type theory of topos with the groupoidal $M - L$ type theory.
We propose new types, associated to every object $X_{\mathcal{A}}$ in every $\mathcal{E}_{\mathcal{A}}$.

The fibration $\mathcal{A} \to \Gamma$ itself can be identified with the final object $\mathbf{1}_{\mathcal{A}} \in \mathcal{E}_{\mathcal{A}}$ in the context Γ.

Sections of $\mathcal{A} \to \Gamma$ are particular cases of objects. For the terms in an object X_A, we take any natural transformation from the object S corresponding to a section $\Gamma \to \mathcal{A}$ to the object X_A in $\mathcal{E}_{\mathcal{A}}$.
A simple section is a term to $\mathbf{1}_{\mathcal{A}}$, the final object, which is a usual M-L type.

Due to the adjunction for the topos of presheaves, the construction Σ and Π extend to the new types.

Now a classifier of subobjects $\Omega_{\mathcal{A}}$ is available for any M-L type $\mathcal{A}$.
We define relative subobjects using the correspondence $\lambda_\pi : \Omega_{\mathcal{A}} \to \pi^\star \Omega_\Gamma$.

This extension of M-L theory allows to define languages and semantics over DNNs with internal structure in the model category $\mathcal{M}$.

3

Dynamics and homology

3.1 Ordinary cat's manifolds

Some limits, in the sense of category theory, of the dynamical object X^w of $C^{\sim}$ describe the sets of activities in the DNN which correspond to some decisions taken by its output (the so called *cat's manifolds* in the folklore of Deep Learning).

Here we consider the case of supervised learning or the case of reinforcement learning, because the success or the failure of an action integrating the output of the network is also a kind of metric.

For instance, consider a proposition P_{out} about the input ξ_{in} which depends on the final states ξ_{out}. It can be seen as a function P on the product $X_B = \prod_b X_b$ of the spaces of states over the output layers to the boolean field $\Omega_{\mathsf{Set}} = \{0, 1\}$, taking the value 1 if the proposition is true, 0 if not. Our aim is to better understand the involvement of the full network in this decision; it is caused by the input data in a deterministic manner, but it results from the chosen weights and from the full functioning of the DNN. One of the many ways to express the situation in terms of category is to enlarge C (or $\mathbf{\Gamma}$) by several terminal layers (see figure 3.1):

1) a layer $B^{\star}$ which makes the product of the output layers, as we have done with forks, followed by the layer B (remark that this can be replaced by B only, with an arrow from $b \in x_{\text{out}}$);

2) a layer ω_b with one cell and two states in a set Ω_b, as in Ω_{Set}, with one arrow from ω_b to B, for translating the proposition P, followed by a last layer ω_1, with one arrow $\omega_b \to \omega_1$, the state's space X_{ω_1} being a singleton $\star_1$, and the map $\star_1 \to \Omega_b$ sending the singleton to $\mathbf{1}$. This gives a category C_+ enlarging C by a fork with handle $B \leftarrow \omega_b \to \omega_1$, and a unique extension X^w_+, depending on P, of the functor X^w from C^{op} to Set in a presheaf over C_+.

The space of sections singletons of X^w_+ is identified naturally with the space of sections of X^w such that the output satisfies P_{out}, i.e. the subset of the product of all the $X^w(a)$ when a describes C made by the coherent activities giving the assertion "P is true" at the output. In this picture, we also can consider that P is the weight over the arrow $B \leftarrow \omega_b$, and note $X^{w,P}_+$ the extension of X^w.
In other terms, the subset of activities of X which affirm the proposition P_{out} is given by a value of the

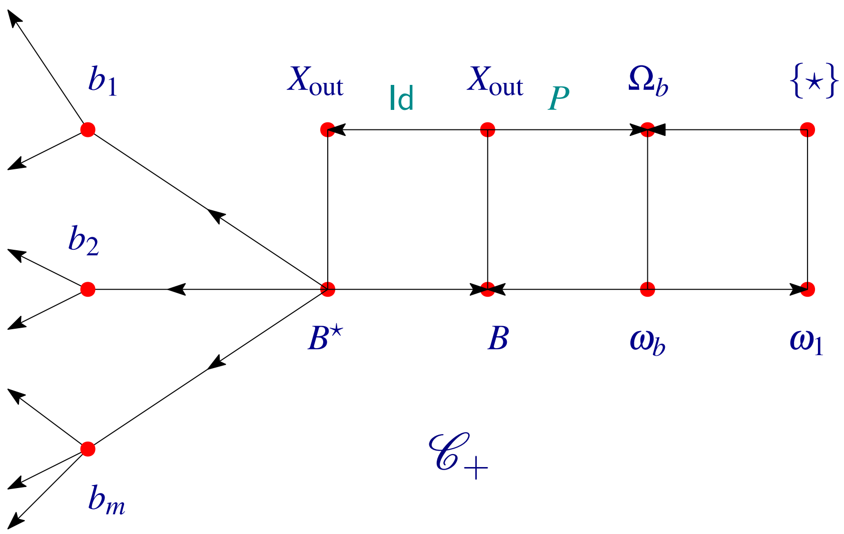


Figure 3.1: Interpretation of a proposition : categorical representation

right Kan extension of X_+ along the unique functor $p_+ : C_+^{\text{op}} \to \star$:

$$M(P_{\text{out}})(X) = \mathbf{R}\text{Kan}_{C_+}(X_+) = \lim_{a \in C_+^{\text{op}}} X_+^w(a) : \tag{3.1}$$

In the AI folklore, the set $M(P_{\text{out}})(X)$ is named a *cat's manifold*, alluding to the case where the network has to decide if yes or no a cat is present in the image. $M(P_{\text{out}})(X)$ can be identified with a subset of the product X_{in} of the input layers. It has to be compared with the assertion "P is true" made by an observer, then studied in function of the weights $\mathbb{W}$ of the dynamics.
However, in general, $M(P_{\text{out}})(X)$ cannot be identified with a product of subsets in the X_a's, for $a \in C$; it is a global invariant.
In fact, it is a particular case of a set of cohomology:

$$M(P_{\text{out}})(X) = H^0(C_+; X_+). \tag{3.2}$$

If the proposition P_{out} is always true, M coincides with the set of section of $X = X^w$, which can be identified with the product of the entry layers activities:

$$\Gamma(X) = H^0(C; X) \cong \prod_{a \in x_{\text{in}}} X_a \tag{3.3}$$

The construction of C_+ and the extension of X by X_+ can be seen as a conditioning. The map $X_+(\omega_b \to B)$ is equivalent to a proposition, the characteristic map of a subset of X_B. In this case we have

$$H^0(C_+; X_+) \subset H^0(C; X). \tag{3.4}$$

In the same manner, we define the manifold of a theory $\mathbb{T}_{\text{out}}$ expressed in a typed language $\mathbb{L}_{\text{out}}$ in the output layers, by replacing the above objects ω_b, ω_1, and the presheaf $X_+(P)$ over them, by larger sets and $X_+(\mathbb{T})$, as the set of sections of $X_+(\mathbb{T})$ over the whole C_+.

We will revisit the notion of cat's manifold when considering the homotopy version of semantic information.

3.2 Dynamics with spontaneous activity

In our approach of networks functioning, the feed-forward dynamic coincides with the limit set $H^0(X)$. The coincidence with the traditional notion of propagation from the input to the output relies on the particular choice of morphisms at the centers of forks (named tanks), product on one side and isomorphism on the other. But this can be generalized to other morphisms: the only condition being that the inner sources A and the input from the outer world I determine a unique section of the object X^w over C. In concrete terms, this happens if and only if the maps from A and I give coherent values at any tip of each fork.
This tuning involves the values in entry $\xi_0 \in \Xi$, the values of the inner sources $\sigma_0 \in \Sigma$ and the weights, in particular from an A to the $a', a", \ldots$'s. Therefore it depends on the learning process.

Then a possibility for defining coherent dynamical objects with spontaneous activities is to start with standard objects X^w, satisfying the restriction of products and isomorphisms, then to introduce small deformations of the projections maps, and obtain the global dynamics by using algorithms which realize the Implicit Function Theorem in the Learning process.

Another possibility, closer to the natural networks in animals, and more readable, is to keep unchanged the projections to the tips a',..., and to introduce new dynamical entries y_A for each tang A, then to send a message to the handle a according to the following formula

$$x_a = f_a^A(x_{a'}, x_{a"}, \ldots; y_A). \tag{3.5}$$

The state in A being described by $(x_{a'}, x_{a"}, \ldots, y_A)$.
In such a manner the coherence is automatically verified. Each collection of inputs and tangs modulations generates a unique section.
These spontaneous entries can be learned by backpropagation, as the weights, by minimizing a functional, or realizing a task with success.

It is important to remark that in natural brains, even for very small animals, having no more than several hundred neurons, the part of spontaneous activity is much larger than the part due to the sensory inputs. This activity comes from internal rhythms, regulatory activities of the autonomous system, internal motivations more or less planed. The neural network transforms them in actions or more general decisions. To make them efficient, corrections are necessary, due to reentrant architectures.

However these natural networks in general do not learn using fully supervised methods; they depend on reinforcement, by success of actions, or by unsupervised methods, involving maximization of mutual information quantities. This will require much further works to achieve this degree of integration in artificial networks. Also evolution plays a fundamental role, in particular by specifying the processes of weights transformations. But certainly experiments can be easily conducted in this direction, with simple networks as in *Logical Information Cells* [BBG21a], the experimental companion article.

3.3 Fibrations and cofibrations of languages and theories

In this section, we define several sheaves and cosheaves over the stack $\mathcal{F}$, that are naturally associated to languages and theories, defining moduli over monoids in the classifying topos $\mathcal{E}$, by using the semantic conditioning (see theorem 3.1), in such a manner that their homology or homotopy invariants, in the sense of topos, give tentative semantical Information quantities and spaces. In the most elementary cases, we recover in the following sections 3.4, 3.5, the definitions of Carnap and Bar-Hillel [CBH52], and their known generalizations [BBD$^+$11], [BBDH14], already used in Deep Learning (see for instance [XQLJ20]), but at the end of this chapter, we will also show new promising elements of information.
In this section and the following ones, we use the semantic functioning in usual $DNNs$ to define their semantic information content. Taking into account the internal dimensions given by the stacks $\mathcal{F}$ over C, several levels of information emerge. Without closing the subject, they reflect different meaning of the word *information*.

A first level concerns the pertinent types, or objects, to introduce in order to understand how the network performs a semantic task, in addition to the types coming from $\mathbb{L}_{\text{out}}$, that are put at the hand by the observer, and guide the learning process, by backpropagation or reinforcement. A first conjecture, that we will not study in the present text, is that new objects appear in cohomological forms, as obstructions for integrating correctly the input data in the output theory. It is not excluded that this can appear spontaneously in the network, but more probably it requires the intervention of the observer, for changing the functional (the metric) or the data, which induces a variation of the weights. We will describe below in section 3 examples of semantic groupoids which could generate or constrain these obstructions. More precisely, we expect that the new objects are vanishing cycles, in the sense of Grothendieck, Deligne, Laumon [Ill14], for convenient maps of sites, localized in the fibers $\mathcal{F}_U$, at points (U, ξ).
In some regions of the weights, the network should become able to develop a semantic functioning about the new objects, formalized by the languages $\mathbb{L}_U; U \in C$ similarly to what happens with singularities of functions or varieties, with imposed reality conditions. The analogy is made more precise in chapter 4.

A second level, perhaps not independent of the first one, concerns the information contained in some theories about other theories, or about decisions to take or actions to do, for instance $\mathbb{T}_{U'}$ in some layer, considered in relation to $\mathbb{T}_U$, when $\alpha : U \to U'$, or $\mathbb{T}_{\text{out}}$. As we saw, the expression of these theories in functioning networks depends on the given section ϵ of X^w. However, we expect that the notion of information also allows to compare the theories made by different networks about a some class of problems.

The semantic information that we want to make more precise must be attached to the communication between layers and the communication between networks, and attached to some problems to solve, for a view of the necessity to introduce interaction in a satisfying view of information. See Thom in [Tho83].

Some theories will be more informative than others, or more redundant, then we will be happy to attach quantitative notions of amount of information to the notion of semantic information. However, efficient numerical measures should also take care of the expression of theories by some axioms. Some systems of axioms are more economical than others, or more redundant than others. Redundancy is more the matter of axioms, ambiguity is more the matter of theories. In the present approach, the notion of ambiguity comes first.

In Shannon information theory, [SW49], the fundamental quantity is the entropy, which is in fact a measure of the ambiguity of the expressed knowledge with respect to an individual fact, for instance a message. Only some algebraic combinations of entropies can be understood as an information in the common sense, for instance the mutual information

$$I(X;Y) = H(X) + H(Y) - H(X,Y).$$

Here the theories $\mathbb{T}_U, U \in \mathcal{C}$ are seen as possible models, analogous to the probabilistic models $\mathbb{P}_X, X \in \mathcal{B}$ in Bayesian networks. The variables of the Bayesian network are analogous to the layers of the neural networks; the values of the variables are analogs of the states of the neurons of the layers. In some version of Bayes analysis, for instance presented by Pearl [Pea88], the Bayes network is associated to a directed graph, but in some other versions it is an hypergraph [YFW01], or a more general poset [BPSPV20].

In the case of the probabilistic models, Shannon theorems have revealed the importance of entropy and of mutual information. It has been shown in [BB15] and [Vig20]), that the entropy is a universal class of cohomology of degree one of the topos of presheaves over the Bayesian network, seen as a poset $\mathcal{B}$, equipped with a cosheaf $\mathcal{P}$ of probabilities (covariant functor of sets). The operation of joining variables gives a presheaf $\mathcal{A}$ in monoids over $\mathcal{B}$. On the other hand, the numerical functions on $\mathcal{P}$ form a sheaf $\mathcal{F}_{\mathcal{P}}$, which becomes an A-module by considering the mean conditioning of Shannon. The entropy belongs to the $\mathsf{Ext}^1_{\mathcal{A}}(K; \mathcal{F}_{\mathcal{P}})$ with coefficients in this module. Moreover, in this framework, higher mutual information quantities [McG54], [Tin62] belong to the homotopical algebra of cocycles of higher degrees

[BB15].

We conjecture that something analog appears in the case of $DNNs$ and theories $\mathbb{T}$, and of axioms for them.

The first ingredient in the case of probabilities was the operation of marginalization of a probability law, interpreted as the definition of a covariant functor (a copresheaf); it can be replaced here by the transfers of theories associated to the functors $F_\alpha : \mathcal{F}_{U'} \to \mathcal{F}_U$, and to the morphisms h in the fibers $\mathcal{F}_U$ from objects ξ to objects $F_\alpha(\xi')$, as we saw in the section 3. For logics, this transfer can go in two directions, depending on the geometry of F_α, from U' to U, and from U to U', as seen in section 2.2.

We start with the transfer from U' to U, having in mind the flow of information in the downstream direction to the output of the DNN; when it exists, a non-supervised learning should also correspond to this direction. However, the learning by backpropagation or by reinforcement goes from the output layers to the inner layers, then the inner layers have to understand something of the imposed language $\mathbb{L}_{\text{out}}$ and the useful theories $\mathbb{T}_{\text{out}}$ for concluding. Therefore we will also laterconsider this backward or upstream direction.

For an arrow $(\alpha, h) : (U, \xi) \to (U', \xi')$, the map

$$\Omega_{\alpha,h} : \Omega_{U'}(\xi') \to \Omega_U(\xi), \tag{3.6}$$

is obtained by composing the map $\lambda_\alpha = \Omega_\alpha$ at ξ', from $\Omega_{U'}(\xi')$ to $\Omega_U(F_\alpha \xi')$ with the map $\Omega_U(h)$ from $\Omega_U(F_\alpha \xi')$ to $\Omega_U(\xi)$.
More generally, for every object X' in $\mathcal{E}_{U'}$, the map $F_!^\alpha$ sends the subobjects of X' to the subobjects of $F_!^\alpha(X')$, respecting the lattices structures. Then for any natural transformation over $\mathcal{F}_U$, $h : X \to F_!^\alpha(X')$, we get a transfer

$$\Omega_{\alpha,h} : \Omega_{U'}^{X'} \to \Omega_U^X. \tag{3.7}$$

The object X or X' is seen as a local context in the topos semantics.
We assume in what follows that this mapping extends to the sentences in the typed languages $\mathbb{L}_U$, where the dependency on ξ reflects the variation of meaning in the included notions. In particular, the morphisms in the topos $\mathcal{E}_U$ express such variations. At the level of theories, this induces in general a weakening, something which is implied at (U, ξ) by the propositions at (U', ξ'), or more generally at the context X by telling what is true, or expected, in the context X'.
In what follows we note by $\mathcal{A} = \Omega^{\mathbb{L}}$ this presheaf of sentences in $\mathbb{L}$ over $\mathcal{F}$, and by $\mathbb{L}_{\alpha,h}$, or $\pi^\star_{\alpha,h}$, its transition maps, extending $\Omega_{\alpha,h}$.

Under the strong standard hypotheses on the fibration $\mathcal{F}$, for instance if it defines a fibrant object in the injective groupoids models, i.e. any F_α is a fibration, (see definition 2.1 above, following lemma 2.4

of section 2.2) there exists a right adjoint of $\Omega_{\alpha,h}$:

$$\Omega'_{\alpha,h} : \Omega^X_U \to \Omega^{X'}_{U'}. \tag{3.8}$$

It is given by extension of the operators λ'_α, associated to $F^\star_\alpha$, in the place of $F^\alpha_!$, plus a transposition.
In what follows we note by $\mathcal{A}' = {}^t\Omega^{\mathbb{L}}$ this copresheaf of sentences over $\mathcal{F}$, and by ${}^t\mathbb{L}'_{\alpha,h}$, or simply $\pi^{\alpha,h}_\star$, its transition maps. The extended strong hypothesis requires that $\pi^\star_{\alpha,h} \circ \pi^{\alpha,h}_\star = \mathsf{Id}$.

For fixed U and $\xi \in \mathcal{F}_U$, the operation $\wedge$ gives a monoid structure on the set $\mathcal{A}_{U,\xi} = \mathcal{A}'_{U,\xi}$, which is respected by the maps $\mathbb{L}_{\alpha,h}$ and ${}^t\mathbb{L}'_{\alpha,h}$.
Moreover, $\mathcal{A}_{U,\xi}$ has a natural structure of poset category, given by the external implication $P \leq Q$, for which $\mathbb{L}_{\alpha,h}$ and ${}^t\mathbb{L}'_{\alpha,h}$ are functors.
There exists a right adjoint of the functor $R \mapsto R \wedge Q$; this is the internal implication, $P \mapsto (Q \Rightarrow P)$. Then, by definition, $\mathcal{A}_{U,\xi} = \mathcal{A}'_{U,\xi}$ is a *closed monoidal category*. In fact this is the only structure that is essentially needed for the information theory below; this allows the linear generalization of appendix E.

The maps $\pi^\star$ and $\pi_\star$ give a fibration $\widetilde{\mathcal{A}}$ over $\mathcal{F}$, and a cofibration $\widetilde{\mathcal{A}'}$ over $\mathcal{F}$, in the sense of Grothendieck [Mal05]:
a morphism γ in $\widetilde{\mathcal{A}}$ from (U, ξ, P) to (U', ξ', P'), lifting a morphism (α, h) in $\mathcal{F}$ from (U, ξ) to (U', ξ'), is given by an arrow ι in $\Omega^{\mathbb{L}_U}$ from P to $\mathbb{L}_{\alpha,h}(P') = \pi^\star_{\alpha,h} P'$, that is an external implication

$$P \leq \mathbb{L}_{\alpha,h}(P'). \tag{3.9}$$

Similarly, an arrow in the category $\widetilde{\mathcal{A}'}$ lifting the same morphism (α, h) in $\mathcal{F}$, is an implication

$${}^t\mathbb{L}'_{\alpha,h}(P) \leq P'. \tag{3.10}$$

Remark that *a priori* the left adjunction $\pi^\star_{\alpha,h} \dashv \pi^{\alpha,h}_\star$ does not imply something between P and $\mathbb{L}_{\alpha,h}(P')$ when (3.10) is satisfied. However, under the strong hypothesis $\pi^\star \circ \pi_\star = \mathsf{Id}$, the relation (3.10) implies the relation (3.9). Then in this case, $\widetilde{\mathcal{A}'}$ is a subcategory of $\widetilde{\mathcal{A}}$.

Remark. An important particular case, where our standard hypotheses are satisfied, is when the $\Omega^{\mathbb{L}_{U,\xi}} = \mathcal{A}_{U,\xi}$ are the sets of open sets of a topological spaces $Z_{U,\xi}$, and when there exist continuous open maps $f_\alpha : Z_{U',\xi'} \to Z_{U,\xi}$ lifting the functors F_α, such that the maps $\pi^\star$ and $\pi_\star$ are respectively the direct images and the inverse images. The strong hypothesis holds when the f_α are topological fibrations.

$\widetilde{\mathcal{A}}$ and $\widetilde{\mathcal{A}'}$ belong to augmented model categories using monoidal posets [Rap10]. See section 2.5.

For $P \in \Omega^{\mathbb{L}_{U,\xi}} = \mathcal{A}_{U,\xi}$, we note $\mathcal{A}_{U,\xi,P}$ the set of proposition Q such that $P \leq Q$. They are sub-monoidal categories of $\mathcal{A}_{U,\xi}$. Moreover they are closed, because $P \leq Q, P \leq R$ implies $P \wedge Q = P$, then $P \wedge Q \leq R$, then $P \leq (Q \Rightarrow R)$.
When varying P, these sets form a presheaf over $\mathcal{A}_{U,\xi} = \mathcal{A}'_{U,\xi}$.

Lemma 3.1. *The monoids $\mathcal{A}_{U,\xi,P}$, with the functors $\pi^\star$ between them, form a presheaf over the category $\widetilde{\mathcal{A}}$.*

Proof. Given a morphism $(\alpha, h, \iota) : \mathcal{A}_{U,\xi,P} \to \mathcal{A}_{U',\xi',P'}$ in $\widetilde{\mathcal{A}}$, the symbol ι means $P \leq \pi^\star P'$, then, from $P' \leq Q'$, we deduce $P \leq \pi^\star P' \leq \pi^\star Q'$. ■

Lemma 2.4 in section 2.2 established the existence of a counit $\eta : \pi^\star \pi_\star \to \mathsf{Id}_U$, for every morphism $(\alpha, h) : (U, \xi) \to (U', \xi')$, then for every $P \in A_{U,\xi}$, we have $\pi^\star \pi_\star P \leq P$.

Under the stronger hypothesis on the fibration $\mathcal{F}$, that $\eta = \mathsf{Id}_{\Omega^{\mathbb{L}}}$, i.e. $\pi^\star \pi_\star P = P$, lemma 3.1 holds also true for the category $\widetilde{\mathcal{A}}'$.

Definition. *$\Theta_{U,\xi}$ is the set of theories expressed in the algebra $\Omega^{\mathbb{L}_U}$ in the context ξ. Under our standard hypothesis on $\mathcal{F}$, both $\mathbb{L}_\alpha$ and ${}^t\mathbb{L}_\alpha$ send theories to theories.*

Definition. *$\Theta_{U,\xi,P}$ is the subset of theories which imply the truth of proposition $\neg P$, i.e. the subset of theories excluding P.*

Remind that $\neg P \equiv (P \Rightarrow \bot)$ is the largest proposition R such that $R \wedge P \leq \bot$.
It is always true that $P \leq P'$ implies $\neg P' \leq \neg P$, but the reciprocal implication in general requires a boolean logic.
Then, for fixed U, ξ, the sets $\Theta_{U,\xi,P}$ when P varies in $\mathcal{A}_{U,\xi}$, form a presheaf over $\mathcal{A}_{U,\xi}$; if $P \leq Q$, any theory excluding Q is a theory excluding P.

Lemma 3.2. *Under the standard hypotheses on the fibration $\mathcal{F}$, without necessarily axiom (2.30), the sets $\Theta_{U,\xi,P}$ with the morphisms $\pi^\star$, form a presheaf over $\widetilde{\mathcal{A}}$.*

Proof. Let us consider a morphism $(\alpha, h, \iota) : \mathcal{A}_{U,\xi,P} \to \mathcal{A}_{U',\xi',P'}$, where ι denotes $P \leq \pi^\star P'$; we deduce $\pi^\star \neg P' = \neg \pi^\star P' \leq \neg P$; then $T' \leq \neg P'$ implies $\pi^\star T' \leq \pi^\star \neg P' \leq \neg P$. ■

Corollary. *Under the standard hypotheses on the fibration $\mathcal{F}$ plus the stronger one, the sets $\Theta_{U,\xi,P}$ with morphisms $\pi^\star$, also form a presheaf over $\widetilde{\mathcal{A}}'$.*

What happens to $\pi_\star$?

It is in general false that the collection $\mathcal{A}_{U,\xi,P}$ with the functors $\pi_\star^{\alpha,h}$ forms a copresheaf over $\widetilde{\mathcal{A}}'$. However, if we restrict ourselves to the smaller category $\widetilde{\mathcal{A}}'_{\text{strict}}$, with the same objects but with morphisms from $\mathcal{A}_{U,\xi,P}$ to $\mathcal{A}_{U',\xi',P'}$ only when $P' = \pi_\star^{\alpha,h} P$, this is true.

Proof. If $P \leq Q$, $\pi_\star P \leq \pi_\star Q$, then $P' \leq \pi_\star Q$.

■

The same thing happens to the collection of the $\Theta_{U,\xi,P}$ with the morphism $\pi_\star$: over the restricted category $\widetilde{\mathcal{A}}'_{\text{strict}}$, they form a copresheaf. *Proof*: if $T \leq \neg P$, we have $\pi_\star T \leq \pi_\star\neg P = \neg\pi_\star P = \neg P'$. However for the full category $\widetilde{\mathcal{A}}'$ (*resp.* the category $\widetilde{\mathcal{A}}$), the argument does not work: from $\pi_\star P \leq P'$ (*resp.* $P \leq \pi^\star P'$), it follows that $\neg P' \leq \neg\neg\pi_\star P = \pi_\star\neg P$ (*resp.* $\pi_\star P \leq \pi_\star\pi^\star P'$ then $\neg\pi_\star\pi^\star P' \leq \pi_\star P$, then by adjunction $\neg P' \leq \neg\pi_\star P = \pi_\star\neg P$); then $T \leq \neg P$ implies $\pi_\star T \leq \pi_\star\neg P$, not $\pi_\star T \leq \neg P'$.

To summarize what is positive with $\pi_\star$,

Lemma 3.3. *Under the strong standard hypothesis of defition 2.1, the collections $\mathcal{A}_{U,\xi,P}$ and $\Theta_{U,\xi,P}$ with the morphisms $\pi_\star$, constitute copresheaves over $\widetilde{\mathcal{A}}'_{\text{strict}}$.*

Note that the fibers $\mathcal{A}_{U,\xi,P}$ are not sub-categories of $\widetilde{\mathcal{A}}'_{\text{strict}}$, they are subcategoris of $\widetilde{\mathcal{A}}'$ and $\widetilde{\mathcal{A}}$.

Definition. *A theory $\mathbb{T}'$ is said* weaker *than a theory $\mathbb{T}$ if its axioms are true in $\mathbb{T}$. We note $\mathbb{T} \leq \mathbb{T}'$, as we made for weaker probabilistic models. This applies to theories excluding a proposition P, in $\Theta_{U,\xi,P}$.*

With respect to propositions in $\mathcal{A}_{U,\xi}$, if we take the joint R by the operation "and" of all the axioms $\vdash R_i; i \in I$ of $\mathbb{T}$, and the analog R' for $\mathbb{T}'$, the above relation corresponds to $R \leq R'$.
Remark: a weaker theory can also be seen as a simpler or more understandable theory; for instance in Θ_λ, the maximal theory $\vdash (\neg P)$ is dedicated to exclude P, and the propositions implying P.

Be careful that in the sense of sets of truth assertions, the pre-ordering by inclusion of the theories goes in the reverse direction. For instance $\{\vdash \perp\}$ is the strongest theory, in it everything is true, thus every other theory is weaker.

Now we introduce a notion of *semantic conditioning*.

Definition 3.1. *For fixed U,ξ, $P \leq Q$ in $\Omega^{\mathbb{L}_{U,\xi}}$, and $\mathbb{T}$ a theory in the language $\mathbb{L}_{U,\xi}$, we define a new theory by the internal implication:*

$$Q.\mathbb{T} = (Q \Rightarrow \mathbb{T}). \tag{3.11}$$

More precisely: the axioms of $Q.\mathbb{T}$ are the assertions $\vdash (Q \Rightarrow R)$ where $\vdash R$ describes the axioms of $\mathbb{T}$. We consider $Q.\mathbb{T}$ as the conditioning *of $\mathbb{T}$ by Q, in the logical or semantic sense, and frequently we write the resulting theory $\mathbb{T}|Q$.*

At the level of propositions, the operation $\Rightarrow$ is the right adjoint in the sense of the Heyting algebra of the relation $\wedge$, i.e.

$$(R \wedge Q \leq P) \quad iff \quad (R \leq (Q \Rightarrow P)). \tag{3.12}$$

Proposition 3.1. *The conditioning gives an action of the monoid $\mathcal{A}_{U,\xi,P}$ on the set of theories in the language $\mathbb{L}_{U,\xi}$.*

Proof.

$$\begin{aligned}(R \wedge Q' \wedge Q \leq P) \quad \text{iff} \quad & (R \wedge Q') \leq (Q \Rightarrow P) \\ \text{iff} \quad & (R \leq (Q' \Rightarrow (Q \Rightarrow P))).\end{aligned} \tag{3.13}$$

Note that $Q \Rightarrow P$ is also the maximal proposition Q' (for $\leq$) such that $Q \wedge Q' \leq P$.

Therefore the theory $Q \Rightarrow \mathbb{T}$ is the largest one among all theories $\mathbb{T}'$ satisfying

$$Q \wedge \mathbb{T}' \leq \mathbb{T}. \tag{3.14}$$

This implies that $\mathbb{T}|Q$ is weaker than $\mathbb{T}$ and than $\neg Q$.

1) In $Q \wedge \mathbb{T}$, the axioms are of the form $\vdash (Q \wedge R)$ where $\vdash R$ is an axiom of $\mathbb{T}$, and from $\vdash (Q \wedge R)$, we deduce $\vdash R$.

2) Here Q (*resp.* $\neg Q$) is understood as the theory with unique axiom $\vdash Q$ (*resp.* $\vdash \neg Q$), then if $\vdash (Q \wedge \neg Q)$ we have $\vdash \perp$ and all theories are true.

■

Remark. The theory $\mathbb{T}|Q = (Q \Rightarrow \mathbb{T})$ can also be written $\mathbb{T}^Q$, by definition of the internal exponential, as the action by conditioning is also the internal exponential.

Notation: for being lighter, in what follows, we will mostly denote the propositions by the letters $P, Q, R, P', \ldots$ and the theories by the next capital letters $S, T, U, S', \ldots$.

The operation of conditioning was considered by Carnap and Bar-Hilled [CBH52], in the case of Boolean theories, studying the content of propositions and looking for a general notion of sets of semantic Information. In this case $Q \Rightarrow T$ is equivalent to $T \vee \neg Q = (T \wedge Q) \vee \neg Q$ (see the companion text on logicoprobabistic information for more details [BBG20]).
Their main formula for the concept of information was

$$\mathrm{Inf}(\mathbb{T}|P) = \mathrm{Inf}(\mathbb{T} \wedge P) \backslash \mathrm{Inf}(P); \tag{3.15}$$

assuming that $\mathrm{Inf}(A \wedge B) \supseteq \mathrm{Inf}(A) \cup \mathrm{Inf}(B)$.

Proposition 3.2. *The conditioning by elements of $\mathcal{A}_{U,\xi,P}$, i.e. propositions Q implied by P, preserves the set $\Theta_{U,\xi,P}$ of theories excluding P.*

Proof. Let T be a theory excluding P and $Q \geq P$; consider a theory T' such that $Q \wedge T' \leq T$, we deduce $T' \wedge P \leq T$, thus $T' \wedge P \leq T \wedge P$. But $T \wedge P \leq \perp$, then $T' \wedge P \leq \perp$. But $Q \Rightarrow T$ is the largest theory such that $Q \wedge T' \leq T$, therefore $Q \Rightarrow T$ excludes P, i.e. asserts $\neg P$. ■

Remark. Consider the sets $\Theta'_{U,\xi,P}$ of theories which imply the validity of the proposition P. These sets constitute a cosheaf over the category $\widetilde{\mathcal{A}}'_{\mathrm{strict}}$ for $\pi_\star$ and a sheaf for $\pi^\star$. However, the formula (3.11) does'nt give an action of the monoid $\mathcal{A}_{U,\xi,P}$ on the set $\Theta'_{U,\xi,P}$, even in the boolean case, where $(Q \Rightarrow T) = T \vee \neg Q$.

We can also consider the set of all theories over the largest category $\widetilde{\mathcal{A}}$, without further localization; they also form a sheaf for $\pi^\star$ and a cosheaf Θ for $\pi_\star$, which are stable by the conditioning.

When necessary, we note Θ_{loc} the presheaf for $\pi^\star$ made by the $\Theta_{U,\xi,P}$ over $\widetilde{\mathcal{A}}$.

The naturality over $\widetilde{\mathcal{A}}'_{strict}$ of the action of the monoids relies on the following formulas, for every arrow $(\alpha, h) : (U, \xi) \to (U', \xi')$ in $\mathcal{F}$, we have the arrows $(U, \xi, P) \to (U', \xi', \pi_\star P)$ in $\widetilde{\mathcal{A}}'_{\mathrm{strict}}$; in the presheaf of monoids $\mathcal{A}_{U,\xi,P}$, for the morphism $\pi^\star$, and the presheaf $\Theta_{U,\xi,P}$ with morphisms $\pi_\star$:

$$(\pi^\star Q').T = \pi^\star\left[Q'.\pi_\star(T)\right]. \tag{3.16}$$

This holds true under the strong hypothesis $\pi^\star\pi_\star = \mathsf{Id}$.

If we want to consider functions ϕ of the theories, two possibilities appear: $\pi_\star$ for Θ with $\pi^\star$ for the monoids $\mathcal{A}$, or the opposite $\pi^\star$ for Θ with $\pi_\star$ for the monoids $\mathcal{A}$. But Both cases give a cosheaf over $\widetilde{\mathcal{A}}$, however only the second one gives a functional module Φ over $\mathcal{A}$, even with the strong standard hypothesis,

Theorem 3.1. *Under the strong hypothesis, in particular $\pi^\star\pi_\star = \mathsf{Id}$, and over the restricted category $\widetilde{\mathcal{A}}'_{\mathrm{strict}}$, the cosheaf Φ' made by the measurable functions (with any kind of fixed values) on the theories $\Theta_{U,\xi,P}$, with the morphisms $\pi^\star$, is a cosheaf of modules over the monoidal cosheaf $\mathcal{A}'_{\mathrm{loc}}$, made by the monoidal categories $\mathcal{A}_{U,\xi,P}$, with the morphisms $\pi_\star$.*

Proof. Consider a morphism $(\alpha, h, \iota) : A_{U,\xi,P} \to A_{U',\xi',\pi_\star P}$, a theory T' in $\Theta_{U',\xi',\pi_\star P}$, a proposition Q in $A_{U,\xi,P}$, and an element ϕ_P in $\Phi'_{U,\xi,P}$, we have

$$\begin{aligned}
\pi_\star Q.(\Phi'_\star\phi_P)(T') &= (\Phi'_\star\phi_P)(T'|\pi_\star Q)\\
&= \phi_P[\pi^\star(T'|\pi_\star Q)]\\
&= \phi_P[\pi^\star(\pi_\star Q \Rightarrow T')]\\
&= \phi_P[\pi^\star\pi_\star Q \Rightarrow \pi^\star T']\\
&= \phi_P[Q \Rightarrow \pi^\star T']\\
&= \phi_P[\pi^\star(T')|Q].
\end{aligned}$$

■

Remark. The same kind of computation shows that, in the case of the sheaf Φ of functions on the cosheaf Θ with $\pi_\star$ and the sheaf $\mathcal{A}_{\mathrm{loc}}$ with $\pi^\star$, we have, for the corresponding elements Q', T, ϕ',

$$\pi^\star Q'.\Phi^\star(\phi')(T) = \phi'(\pi_\star\pi^\star Q'.\pi_\star T); \tag{3.17}$$

which is not the correct equation of compatibility, under our assumption. It should be true for the other direction, if $\epsilon = \pi_\star \pi^\star = \mathsf{Id}_{\mathcal{A}_{U',\xi'}}$.

However, there exists an important case where both hypotheses $\pi^\star \pi_\star = \mathsf{Id}_U$ and $\pi_\star \pi^\star = \mathsf{Id}_{U'}$ hold true; it the case where the languages over the objects (U, ξ) are all isomorphic. In terms of the intuitive maps f_α, this means that they are homeomorphisms. This case happens in particular when we consider the restriction of the story to a given layer in a network.

Remark. Considering lemmas 3.1 and 3.2, we could forget the functional point of view with a space Φ. In this case we do not have an Abelian situation, but we have a sheaf of sets of theories Θ_{loc}, on which the sheaf of monoids $\mathcal{A}_{\mathrm{loc}}$ acts by conditioning,

Proposition 3.3. *The presheaf Θ_{loc} for $\pi^\star$ is compatible with the monoidal action of the presheaf $\mathcal{A}_{\mathrm{loc}}$, both considered on the category $\widetilde{\mathcal{A}}$ (then over $\widetilde{\mathcal{A}}'$ by restriction, under the strong standard hypothesis on $\mathcal{F}$).*

Proof. If $T' \leq \neg P'$ and $P \leq \pi^\star P'$, we have $\neg \pi^\star P' \leq \neg P$, therefore $\pi^\star T' \leq \neg P$. ■

In the Bayesian case, the conditioning is expressed algebraically by the Shannon mean formula on the functions of probabilities:

$$Y.\phi(\mathbb{P}_X) = \mathbb{E}_{Y_\star \mathbb{P}_X}(\phi(\mathbb{P}|Y = y)) \tag{3.18}$$

This gives an action of the monoid of the variables Y coarser than X, as we find here for the fibers $A_{U,\xi,P}$ and the functions of theories $\Phi_{U,\xi,P}$.

Equation (3.15) was also inspired by Shannon's equation

$$(Y.H)(X;\mathbb{P}) = H((Y,X);\mathbb{P}) - H(Y;Y_\star \mathbb{P}). \tag{3.19}$$

However this set of equations for a system $\mathcal{B}$ can be deduced from the set of equations of invariance

$$(H_X - H_Y)|Z = H_{X \wedge Z} - H_{Y \wedge Z}. \tag{3.20}$$

In the semantic framework, two analogies appear with the bayesian framework: in one of them, in each layer, the role of random variables is played by the propositions P; in the other one, their role is played by the layers U, augmented by the objects of a groupoid (or another kind of category for contexts). The first analogy was chosen by Carnap and Bar-Hillel, and certainly will play a role in our toposic approach too, at each U, ξ, to measure the logical value of functioning. However, the second analogy is more promising for the study of DNNs, in order to understand the semantic adventures in the feedforward and feedback dynamics.

To unify the two analogies, we have to consider the triples (U, ξ, P) as the semantic analog of random variables, with the covariant morphisms of the category $\mathcal{D} = \widetilde{\mathcal{A}'}^{\mathrm{op}}_{\mathrm{strict}}$,

$$(\alpha, h, \pi_\star^{\alpha,h}) : (U, \xi, P) \to (U', \xi', P' = \pi_\star^{\alpha,h} P), \tag{3.21}$$

as analogs of the marginals.
In fact, a natural extension exists and will be also studied, replacing the monoids $\mathcal{A}_{U,\xi,P}$ by the monoids $\mathcal{D}_{U,\xi,P}$ of arrows in $\mathcal{D}$ going to (U,ξ,P), i.e. replacing $\widetilde{\mathcal{A}}_{\text{loc}}$ by the left slice $\mathcal{D}\backslash\mathcal{D}$. This will allow the use of combinatorial constructions over the nerves of $\mathcal{F}$ and C.

If we consider the theories in $\Theta_{U,\xi,P}$ as the analogs of the probability laws, the analogs of the values of a variable Q are the conditioned theories $T|Q$.

When a functioning network is considered, the neural activities in $X^w_{U,\xi}$, can also be seen as values of the variables, through a map $S_{U,\xi,P} : X^w_{U,\xi} \to \Theta_{U,\xi,P}$.

As defined in section 2.3, a *semantic functioning* of the neural network $\mathbb{X}$ is given by a function

$$S_{U,\xi} : \mathbb{X}_{U,\xi} \to \Theta_{U,\xi}. \tag{3.22}$$

The introduction of P, seen as logical localization, corresponds to a refined notion of semantic functioning, a quotient of the activities made by the neurons that express a rejection of this proposition. This generates a foliations in the individual layer's activities.

Remark. We could also consider the cosheaf Θ' or Θ'_{loc} over $\widetilde{\mathcal{A}}'_{\text{strict}}$, and obtain the cosheaf Σ', of all possible maps $S_{U,\xi} : X_{U,\xi} \to \Theta'_{U,\xi}; U \in C, \xi \in \mathcal{F}_U$, where the transition from U,ξ to U',ξ' over α, h is given by the contravariance of X and by the covariance of Θ':

$$\Sigma'_{\alpha,h}(s_U)_{U',\xi'} = {}^t\,\mathbb{L}'_{\alpha,h} \circ s_U \circ X_{\alpha,h}. \tag{3.23}$$

However the above discussion shows that the compatibility with the conditioning would require $\pi_\star \pi^\star = \mathsf{Id}$, which appears to be too restrictive.

In addition, the network's feed-forward dynamics X^w makes appeal to a particular class of inputs Ξ, and is more or less adapted by learning to the expected theories Θ_{out} at the output. Therefore a convenient notion of information, if it exists, must involve these ingredients.

By using functions of the mappings $S_{U,\xi}$, we could not apply them to particular vectors in $X_{U,\xi}$. But using functions on the $\Theta_{U,\xi}$ we can. Then this will be our choice. And this can give numbers (or sets or spaces) associated to a family of activities $x_\lambda \in X_\lambda$, and to their semantic expression $S_\lambda(x_\lambda) \in \Theta'_\lambda$. Moreover, we can take the sum over the set of x belonging to Ξ, then a sum of semantic information corresponding to the whole set of data and goals. Which seems preferable.

The relations

$$S_{U,\xi} \circ X^\star = \pi^\star \circ S_{U',\xi'}, \tag{3.24}$$

mean that the logical transmission of the theories expressed by U' (in the context ξ') coincide with the theories in U induced by the neuronal transmission from U' to U.

If this coherence is verified, the object Σ in the topos, replacing Σ', could be taken as the exponential object $\Theta^{\mathbb{X}}$ in the topos of presheaves over $\widetilde{\mathcal{A}}$. By definition, this is equivalent to consider the parameterized families of functioning

$$S_\lambda : \mathbb{X}_{U,\xi} \times Y_\lambda \to \Theta_{U,\xi,P}; \tag{3.25}$$

where Y is any object in the topos of presheaves over $\widetilde{\mathcal{A}}$.

Remark. In the experiments with small networks, we verified this coherence, but only approximatively, i.e. with high probability on the activities in X.

On another hand, a semantic information over the network must correspond to the impact of the inner functioning on the output decision, given the inputs. For instance, it has to measure how far from the output theory is the expressed theory at U', ξ'. We hope that this should be done by an analog of the *mutual information* quantities. If we believe in the analogy with probabilities, this should be given by the topological coboundary of the family of sections of the module $\Phi_\lambda; \lambda \in \widetilde{\mathcal{A}'}$ [BB15]

Then we enter the theory of topological invariants of the sheaves of modules in a ringed topos. Here Φ over $\mathcal{D}$, or $\mathcal{D}\backslash\mathcal{D}$.

The category $\mathcal{D} = \widetilde{\mathcal{A}'}^{\text{op}}_{\text{strict}}$ gives birth to a refinement of the cat's manifolds we have defined before in section 3.1:
Suppose, to simplify, that we have a unique initial point in C; it corresponds to the output layer U_{out}. Then look at a given $\xi_0 \in \mathcal{F}_{\text{out}}$, and a given proposition P_{out} in $\Omega_{\text{out}}(\xi_0) = \mathcal{A}_{U_{\text{out}},\xi_0}$; it propagates in the inner layers through $\pi_\star$ in $P \in \mathcal{A}_{U,\xi}$ for any U and any ξ linked to ξ_0, and can be reconstructed by $\pi^\star$ at the output, due to the hypothesis $\pi^\star \pi_\star = \mathsf{Id}$. Then we get a section over C of the cofibration $\mathcal{D}^{\text{op}} \to C$. This can be extended as a section of $\mathcal{D}^{\text{op}} \to \mathcal{F}$, by varying ξ_0, when the F_α are fibrations, which is the main case we have in mind.
Note that this does not give all the sections, because some propositions P in a $\mathcal{A}_\lambda$ are not in the image of $\pi_\star$, even if all of them are sent by $\pi^\star$ to an element of a set $\Omega_{\text{out}}(\xi_0)$.
However, these interesting sections are in bijection with the connected components of $\mathcal{D}^{\text{op}}$.

Let $\mathbb{K}$ be a commutative ring, and c_P a non zero element of $\mathbb{K}$; we define the (measurable) function δ_P on the theories in the $\Theta_\lambda(P)$, taking the value c_P over a point in the above connected component of $\mathcal{D}$, and 0 outside.

Looking at the semantic functioning $S : X_{U,\xi} \to \Theta_\lambda$, we get a function δ_P on the sets of local activities. This function takes the value c_P on the set of activities that form theories excluding P.
Several subtle points appear:

1) the function really depends on P, but when P varies, it does not change when two propositions have the same negation $\neg P$;

2) to conform with the before introduced notion of cat's manifold, we must assume that the activities in different layers which exclude P in their axioms, are coherent, i.e. form a section of the object X^w.

Without the coherence hypothesis between dynamics and logics, we have two different notions of cat's manifolds, one dynamic and one linguistic or logical. In a sense, only the agreement deserves to be really named semantics.

3.4 Semantic information. Homology constructions

Bar complex of functions of theories and conditioning by propositions.

We start with the computation of the Abelian invariants, therefore with the module of functions Φ on Θ in the cases where conditioning act.

We consider first the most interesting case described by theorem 3.1, given by the presheaf Θ over the category $\mathcal{D}$, fibred over $\mathcal{F}$ which is itself fibred over C. Note that over $\widetilde{\mathcal{A}}'_{\text{strict}}$ we get cosheaves, thus we prefer to work over the opposite $\mathcal{D}$. Then $\mathcal{A}'_{\text{loc}}$ with morphisms $\pi_\star$, becomes a sheaf of monoids over $\mathcal{D}$, and Θ'_{loc}, with morphisms $\pi^\star$, becomes a cosheaf of sets over $\mathcal{D}$, in such a manner that the functions Φ on Θ'_{loc} constitute a sheaf of $\mathcal{A}'_{\text{loc}}$ modules.

We suppose that the elements ϕ_λ in Φ_λ take their values in a commutative ring K (with cardinality at most continuous).

The method of relative homological algebra, used for probabilities in Baudot, Bennequin [BB15], and Vigneaux [Vig20], cited above, can be applied here, for computing $\mathsf{Ext}^\star_{\mathcal{A}'_{\text{loc}}}(K,\Phi)$ in the toposic sense. The action of $\mathcal{A}'_{\text{loc}}$ on K is supposed trivial.

We note $\mathcal{R} = K\left[\mathcal{A}'_{\text{loc}}\right]$ the cosheaf in K-algebras associated to the monoids $\mathcal{A}'_\lambda; \lambda \in \mathcal{A}'$. The non-homogeneous bar construction gives a free resolution of the trivial constant module K:

$$0 \leftarrow K \leftarrow B'_0 \leftarrow B'_1 \leftarrow B'_2 \leftarrow \dots \tag{3.26}$$

where $B'_n; n \in \mathbb{N}$, is the free $\mathcal{R}$ module $\mathcal{R}^{\otimes(n+1)}$, with the action on the first factor. In each object $\lambda = (U, \xi, P)$, the module $B'_n(\lambda)$ is freely generated over $K\left[\mathcal{A}'_\lambda\right]$ by the symbols $[P_1|P_2|\dots|P_n]$, where the P_i are elements of $\mathcal{A}'_\lambda$, i.e. propositions implied by P. Then the elements of $B'_n(\lambda)$ are finite sums of elements $P_0[P_1|P_2|\dots|P_n]$.

The first arrow from B'_0 to K is the coordinate along $[\emptyset]$.

The higher boundary operators are of the Hochschild type, defined on the basis by the formula

$$\partial[P_1|P_2|...|P_n] \;=\; P_1[P_2|...|P_n] \;+\; \sum_{i=1}^{n-1}(-1)^i[P_1|...|P_iP_{i+1}|...|P_n] \;+\; (-1)^n[P_1|P_2|...|P_{n-1}] \quad (3.27)$$

For each $n \in \mathbb{N}$, the vector space $\mathsf{Ext}^n_{\mathcal{A}'}(K,\Phi)$ is the n-th group of cohomology of the associated complex $\mathsf{Hom}_{\mathcal{A}'}(B^\star,\Phi)$, made by natural transformations which commutes with the action of $K[\mathcal{A}']$.

The coboundary operator is defined by

$$\delta f_\lambda(T;Q_0|...|Q_n) =$$

$$f_\lambda(T|Q_0;Q_1|...|Q_n)+\sum_{i=0}^{n-1}(-1)^{i+1}f_\lambda(T;Q_0|...|Q_iQ_{i+1}|...|Q_n)+(-1)^{n+1}f_\lambda(T;Q_0|...|Q_{n-1}). \quad (3.28)$$

A cochain of degree zero is a section $\phi_\lambda; \lambda \in \mathcal{D}$ of Φ, that is, a collection of maps $\phi_\lambda : \Theta'_\lambda \to K$, such that, for any morphism $\gamma : \lambda \to \lambda'$ in $\mathcal{D}^{\mathrm{op}}$, and any $S' \in \Theta'_{\lambda'}$, we have

$$\phi_{\lambda'}(S') = \phi_\lambda(\pi^\star S'). \quad (3.29)$$

If there exists a unique last layer U_{out}, as in the chain, this implies that the functions ϕ_μ are all determined by the functions ϕ_{out} on the sets of theories S_{out} in the final logic, excluding given propositions, by definition of the sets $\Theta'_{U,\xi,P}$. And a priori these final functions are arbitrary.

Acyclicity and fundamental cochains

To be a cocycle, ϕ must satisfy, for any $\lambda = (U,\xi,P)$, and $P \le Q$,

$$0 = \delta\phi([Q])(S) = Q.\phi_\lambda(S) - \phi_\lambda(S) = \phi_\lambda(Q \Rightarrow S) - \phi_\lambda(S). \quad (3.30)$$

However, for any P we have $P \le \top$, and $S|\top = \top$; then the invariance (3.30) implies that ϕ_λ is independent of S; it is equal to $\phi_\lambda(\top)$.

Then, a cocycle is a collection elements $\phi(\lambda)$ in K, satisfying $\phi_{\lambda'} = \phi_\lambda$ each time there exists an arrow from λ to λ' in $\widetilde{\mathcal{A}}'_{\mathrm{strict}}$, thus forming a section of the constant sheaf over $\widetilde{\mathcal{A}}'_{\mathrm{strict}}$.

This gives:

Proposition 3.4. *As*

$$\mathsf{Ext}^0_{\mathcal{A}'}(K,\Phi) = H^0(\widetilde{\mathcal{A}}'_{\mathrm{strict}};K) = K^{\pi_0\left(\widetilde{\mathcal{A}}'_{\mathrm{strict}}\right)}, \quad (3.31)$$

then degree zero cohomology counts the propositions that are transported by $\pi_\star$ from the output.

The discussion at the end of section 3.3 describes the relation between the zero cohomology of information and the cats manifolds, that was identified before with the degree zero cohomology in the sense of Čech.

A degree one cochain is a collection ϕ_λ^R of measurable functions on Θ'_λ, and $R \in \mathcal{A}'_\lambda$, which satisfies the naturality hypothesis: for any morphism $\gamma : \lambda \to \lambda'$ in $\mathcal{D}^{\mathrm{op}}$, and any $S' \in \Theta'_{\lambda'}$, we have

$$\phi_{\lambda'}^{\pi_\star R}(S') = \phi_\lambda^R(\pi^\star S'). \tag{3.32}$$

The cocycle equation is

$$\forall U, \xi, \forall P, \forall Q \geq P, \forall R \geq P, \forall S \in \Theta'_{U,\xi,P}, \phi_\lambda^{Q \wedge R}(S) = \phi_\lambda^Q(S) + \phi_\lambda^R(Q \Rightarrow S). \tag{3.33}$$

Let us define a family of elements of K by the equation

$$\psi_\lambda(S) = -\phi_\lambda^P(S). \tag{3.34}$$

Formula (3.32) implies formula (3.29), then ψ_λ is a zero cochain.
Take its coboundary

$$\delta\psi_\lambda([Q])(S) = -\phi_\lambda^P(S) + Q.\phi_\lambda^P(S). \tag{3.35}$$

using the cocycle equation and the fact that for any $Q \geq P$ we have $Q \wedge P = P$, this gives

$$\phi_\lambda^Q(S) = \phi_\lambda^{Q \wedge P}(S) - Q.\phi_\lambda^P(S) = -\delta\psi_\lambda([Q])(S). \tag{3.36}$$

Remark that the cochain ψ is not unique, the formula $\psi = -\phi_\lambda^P$ is only a choice. Two cochains ψ satisfying $\delta\psi = \phi$ differ by a zero cocycle, that is a family of numbers c_λ, dependent on P but not on S. Remind us that P is part of the object λ.

Therefore every one cocycle is a coboundary, or in other terms:

Proposition 3.5. $\mathrm{Ext}^1_{\mathcal{A}'}(K, \Phi) = 0$.

The same argument applies to every degree $n \geq 1$, giving,

Proposition 3.6. $\mathrm{Ext}^n_{\mathcal{A}'}(K, \Phi) = 0$.

Proof. If $\phi_\lambda^{Q_1;\ldots;Q_n}$ is a cocycle of degree $n \geq 1$, where $\lambda = (U, \xi, P)$, the formula

$$\psi_\lambda^{Q_1;\ldots;Q_{n-1}} = (-1)^n \phi_\lambda^{Q_1;\ldots;Q_{n-1};P} \tag{3.37}$$

defines a cochain of degree $n-1$ such that $\delta\psi = \phi$.
Extracting $\phi_\lambda^{Q_1;\ldots;Q_n}$ from the last term of the cocycle equation for ϕ, applied to $Q_1, \ldots, Q_{n+1}$ with $Q_{n+1} = P$, gives

$$(-1)^n \phi_\lambda^{Q_1;\ldots;Q_n} = Q_1.\phi_\lambda^{Q_2;\ldots;Q_n;P} + \sum_{i=1}^{n-1} \phi_\lambda^{Q_2;\ldots;Q_i Q_{i+1};\ldots;Q_n;P} + (-1)^n \phi_\lambda^{Q_2;\ldots;Q_n \wedge P}. \tag{3.38}$$

As $Q_n \wedge P = P$ in $\mathcal{A}_\lambda$, this is exactly the coboundary of ψ applied to $Q_1; \ldots; Q_n$. ∎

Remark. At first sight this is a deception; however, there is a morality here, because it tells that the measure of semantic information reflects a value of a theory at the output, depending on many elements that the network does not know, without knowing the consequences of this theory. Some of these consequences can be included in the metric for learning, some other cannot be.

When a cochain ψ as above is chosen, it defines the degree one cocycle ϕ by the formula

$$\phi_{\lambda}^{Q}(S) = \psi_{\lambda}(Q \Rightarrow S) - \psi_{\lambda}(S). \tag{3.39}$$

The cochain ψ satisfied (3.29), and the coboundary ϕ the equation (3.32).

All the arbitrariness is contained in the values of ψ_{out}, which are function of P and of the theory excluding P. Now examine the role of a proposition Q implied by P. It changes the value of ϕ according to the equation

$$\phi_{\text{out}}(Q;T) = \phi_{\text{out}}^{Q}(T) = \phi_{\text{out}}^{P}(T) - \phi_{\text{out}}^{P}(T|Q) = \psi_{\text{out}}(T|Q) - \psi_{\text{out}}(T), \tag{3.40}$$

then it subtracts from $\psi_{\text{out}}(T)$ the conditioned value $\psi_{\text{out}}(T|Q)$. And this is transmitted inside the network by the equation

$$\phi_{\lambda'}^{\pi_\star Q}(S') = \phi_{\lambda}^{Q}(\pi^\star S'); \tag{3.41}$$

which is equivalent to the simplest equation

$$\psi_{\lambda'}(S') = \psi_{\lambda}(\pi^\star S'). \tag{3.42}$$

Note that we are working under the hypothesis $\pi^\star \pi_\star = \mathsf{Id}$, then it can happen that a theory S', in the inner layers cannot be reconstructed (by $\pi_\star$) from its deduction $\pi^\star S'$ in the outer layer. Thus the logic inside is richer than the transmitted propositions, but the quantity $\psi_{\lambda'}(S')$ depends only on $\pi^\star S'$.
This corresponds fairly well with what we observed in the experiments about simple classification problems, with architectures more elaborated than a chain, (see Logical cells II, [BBG21b]). In some cases, the inner layers invent propositions that are not stated in the objectives. They correspond to proofs of these objectives.

Mutual information, classical and quantum analogies

We propose now an interpretation of the functions ϕ and ψ, when $\mathbb{K} = \mathbb{R}$, or an ordered ring, as $\mathbb{Z}$: the value $\phi_{\text{out}}^{P}(S)$ measures the ambiguity of S with respect to $\neg P$, then it is natural to assume that the value of $\psi_{\text{out}}(S)$ is growing with S, i.e. $S \leq T$ implies $\psi_{\text{out}}(S) \leq \psi_{\text{out}}(T)$.

Among the theories which exclude P, there is a minimal one, which is $\perp$, without much interest, even it has the maximal information in the sense of Carnap and Bar-Hillel, and a maximal theory, which is $\neg P$ itself; it is the more precise, but with the minimal information, if we measure information by the quantity of exclusions of propositions it can give. Thus ψ does not count the quantity of possible information, but the closeness to $\neg P$.

Consequently, $\phi_P^Q(S)$ is always a positive number, which is decreasing in Q when S is given. Therefore, we can take ψ negative, by choosing $\psi_\lambda = -\phi_\lambda^P$. In what follows we consider this choice for ψ.
The maximal value of $\phi_P^Q(S)$, for a given S is attained for $Q = P$, in this case $S|P = \neg P$, then the maximal value is $\phi_\lambda^P(S) - \phi_\lambda^P(\neg P)$.

The truth of the proposition $\neg Q$ can be seen as a theory excluding P when $P \leq Q$. Like a counterexample of P.

Note the following formula for $P \leq Q$:

$$\phi_\lambda^Q(S) = \phi_\lambda^P(S) - \phi_\lambda^P(S|Q). \tag{3.43}$$

Remind that the entropy function H of a joint probability is also always positive, and we have

$$I(X;Y) = H(X) - H(X|Y), \tag{3.44}$$

as it follows from the Shannon equation and the definition of I.
This also gives $I(X;X) = H(X)$.

Then we interpret $\phi_\lambda^Q(S)$ as a mutual information between S and $\neg Q$, and $\phi_\lambda^P(S)$ itself as a kind of entropy, thus measuring an ambiguity: the ambiguity of what is expressed in the layer λ about the exclusion of P at the output.
This is in agreement with next formula,

$$\phi_\lambda^{\pi_\star Q}(S) = \phi_{\text{out}}^Q(\pi^\star S). \tag{3.45}$$

Remark. In Quantum Information Theory, where variables are replaced by orthogonal decomposition of an Hilbert space, and probabilities are replaced by adapted positive hermitian operators of trace one [BB15], the Shannon entropy H (entropy of the associated classical law) appears as (minus) the coboundary of a cochain which is the Von Neumann entropy $S = -\log_2 \text{Trace}(\rho)$,

$$H_Y(Y;\rho) = S_X(\rho) - Y.S_X(\rho). \tag{3.46}$$

Thus in the present case, it is better to consider that theories are analogs of density matrices, propositions are analogs of the observables, the function ψ is an analog of the opposite of the Von-Neumann entropy, and the ambiguity ϕ an analog of the Shannon entropy.

Let us see what we get for a functioning network X^w, possessing a semantic functioning $S_{U,\xi} : X_{U,\xi} \to \Theta_{U,\xi}$, not necessarily assuming the naturality (3.25). We can even specialize by taking a family of neurons having an interest in the exclusion of some property P, and look at a family

$$S_\lambda : X_{U,\xi} \to \Theta'_\lambda, \tag{3.47}$$

where $\lambda = (U, \xi, P)$.

To a true activity x of the network, we get $x_{U,\xi}$, then, we define

$$H^Q_\lambda(x) = \phi^Q_\lambda(S_\lambda(x_{U,\xi})). \tag{3.48}$$

And we propose it as the ambiguity in the layer U, ξ, about the proposition P at the output, when Q is given as an example.

To understand better the role of Q, we apply the equation (3.32), which gives

$$H^{\pi_\star Q}_{\lambda'}(x') = \phi^Q_\lambda(\pi^\star S'(x')). \tag{3.49}$$

Therefore, evaluated on a proposition $\pi_\star Q$ which comes from the output, the above quantity $I(x')$ in the hidden layer U', is the mutual information of $\neg Q$ and the deduction in U_{out} by $\pi^\star$ of the theory $S'(x')$, expressed in U' in presence of the given section (feedforward information flow), coming from the input, by the activity $x' \in X_{U'}$.

Remark. Consider a chain $(U, \xi) \to (U', \xi') \to (U", \xi")$. We denote by $\rho_\star$ and $\rho^\star$ the applications which correspond to the arrow $(U', \xi') \to (U", \xi")$. Therefore $(\pi')^\star = \pi^\star \rho^\star$ and $\pi'_\star = \rho_\star \pi_\star$.
For any section x, and proposition P in the output (U, ξ), consider the particular case $P = Q$, where $(Q \Rightarrow S) = \neg P$ for every theory excluding P:

$$\begin{aligned}
H(x') - H(x") &= \phi^P_\lambda(\pi^\star S'(x')) - \phi^P_\lambda(\pi^\star S'(x')|P) - (\phi^P_\lambda((\pi')^\star S"(x")) \\
&\quad - \phi^P_\lambda((\pi')^\star S"(x")|P)) \\
&= \phi^P_\lambda(\pi^\star S'(x')) - \phi^P_\lambda((\pi')^\star S"(x")) \\
&= \psi_\lambda(\pi^\star \rho^\star S"(x")) - \psi_\lambda(\pi^\star S'(x'))
\end{aligned}$$

This is surely negative in practice, because the theory $S'(x')$ is larger than the theory $\rho^\star S"(x")$. For instance, at the end, we surely have $S_{\text{out}} = \neg P$, as soon as the network has learned.
Consequently this quantity has a tendency to be negative. Then it is not like the mutual information between the layers. It looks more as a difference of ambiguities. Because the ambiguity is decreasing in a functioning network, in reality.
This confirms that H is a measure of ambiguity.
Therefore, the mutual information should come out in a manner that involves a pair of layers.

To obtain a notion of mutual information, we make an extension of the monoids $\mathcal{A}_{U,\xi,P}$, which continues to act by conditioning on the sets $\Theta_{U,\xi,P}$.

For that, we consider a fibration over $\mathcal{A}'_{\text{strict}}$ made by monoids $\mathcal{D}_\lambda$ which contain $\mathcal{A}_\lambda$ as submonoids. By definition, if $\lambda = (U,\xi,P)$, an object of $\mathcal{D}_\lambda$ is an arrow $\gamma_0 = (\alpha_0, h_0, \iota_0)$ of $\widetilde{\mathcal{A}}'_{\text{strict}}$, going from a triple (U_0,ξ_0,P_0) to a triple $(U,\xi,\pi_\star P_0)$, where $P \leq \pi_\star P_0$, and a morphism from $\gamma_0 = (\alpha_0, h_0, \iota_0)$ to $\gamma_1 = (\alpha_1, h_1, \iota_1)$ is a morphism γ_{10} from (U_0,ξ_0,P_0) to $(U_1,\xi_1,Q_1 = \pi_\star^{\alpha_{10},h_{10}} P_0)$ such that $Q_1 \geq P_1$.

For the intuition it is better to see the objects as arrows in the opposite category $\mathcal{D}$ of $\widetilde{\mathcal{A}}'_{\text{strict}}$, in such a manner they can compose with the arrows $Q \leq R$ in the monoidal category $\mathcal{A}_\lambda$, then we get a variant of the right slice $\lambda|\mathcal{D}$, just extended by $\mathcal{A}_\lambda$. The category $\mathcal{D}_\lambda$ is monoidal and strict if we define the product by

$$\gamma_1 \otimes \gamma_2 = (U,\xi,\pi_\star^{\gamma_1} P_1 \wedge \pi_\star^{\gamma_0} P_2). \tag{3.50}$$

The identity being the truth $\top_\lambda$.

We also define the action of $\mathcal{D}_\lambda$ on Θ_λ as follows:
for every arrow $\gamma_0 : \lambda_0 \to \lambda_{\pi_\star P_0}$, where $\lambda_0 = (U_0,\xi_0,P_0)$, and where $\lambda_{\pi_\star P_0}$ denotes $(U,\xi,\pi_\star P_0)$, assuming $\pi_\star P_0 \geq P$, we define

$$\gamma_0.\mathbb{T} = (\pi_\star^{\gamma_0} P_0 \Rightarrow \mathbb{T}). \tag{3.51}$$

This gives an action of the monoid of propositions in $\mathcal{A}_{\lambda_0}$ which are implied by P_0, whose images by $\pi_\star$ are implied by P.
If $P_0 \leq Q_0$ and $P_0 \leq R_0$, we have $\pi_\star^{\gamma_0}(Q_0 \wedge R_0) = \pi_\star^{\gamma_0}(Q_0) \wedge \pi_\star^{\gamma_0}(R_0)$.

The monoidal categories $\mathcal{D}_\lambda; \lambda \in \mathcal{D}$ form a natural presheaf $\mathcal{D}\backslash\mathcal{D}$ over $\mathcal{D}$. For any morphism $\gamma = (\alpha, h, \iota)$ of $\widetilde{\mathcal{A}}'_{strict}$, going from (U,ξ,P) to $(U',\xi',\pi_\star P)$, and any object $\gamma_0 : \lambda_0 \to \lambda_{\pi_\star P_0}$ in $\mathcal{D}_\lambda$, we define $\gamma_\star(\gamma_0)$ by the composition $(\alpha,h) \circ (\alpha_0,\xi_0)$ and the proposition $\pi_\star^\gamma \circ \pi_\star P_0$ in $\mathcal{A}_{\lambda'}$.

The naturalness of the monoidal action on the theories follows from $\pi_\gamma^\star \pi_\star^\gamma = \mathsf{Id}_U$:

$$\begin{aligned}
\pi_\gamma^\star[\gamma_\star(\pi_\star P_0).T'] &= \pi_\gamma^\star[\pi_\star^\gamma \pi_\star P_0 \Rightarrow T'] \\
&= \pi_\gamma^\star \pi_\star^\gamma \pi_\star P_0 \Rightarrow \pi_\gamma^\star T' \\
&= \pi_\star P_0 \Rightarrow \pi_\gamma^\star T'
\end{aligned}$$

Then, defining $[\Phi_\star(\gamma)(\phi_\lambda)](T') = \phi_\lambda(\pi_\gamma^\star T')$, we get the following result

Lemma 3.4.

$$[\Phi_\star(\gamma)\phi_\lambda](\gamma_\star(\gamma_0).T') = \phi_\lambda(\gamma_0.\pi^\star T'). \tag{3.52}$$

Consequently the methods of Abelian homological algebra can be applied [Mac12].

The (non-homogeneous) bar construction makes now appeal to symbols $[\gamma_1|\gamma_2|...|\gamma_n]$, where the γ_i are elements of $\mathcal{D}_\lambda$. The action of algebra pass through the direct image of propositions $\pi_\star P_i; i = 1,...,n$.

Things are very similar to what happened with the precedent monoids $\mathcal{A}'_\lambda$:
the zero cochains are families ϕ_λ of maps on theories satisfying

$$\psi_\lambda(\pi^\star T') = \psi_{\lambda'}(T'), \tag{3.53}$$

where $\gamma : \lambda \to \lambda'$ is a morphism in $\widetilde{A}'_{\text{strict}}$.
The coboundary operator is

$$\delta\psi_\lambda([\gamma_1]) = \psi_\lambda(T|\pi_\star^{\gamma_1} P_1) - \psi_\lambda(T). \tag{3.54}$$

Then the cohomology is defined as before. We get analog propositions. For instance, the degree one cochains are collections of maps of theories $\phi_\lambda^{\gamma_1}$ satisfying

$$\phi_\lambda^{\gamma_1}(\pi^\star T') = \phi_{\lambda'}^{\gamma_\star \gamma_1}(T'); \tag{3.55}$$

the cocycle equation is

$$\phi_\lambda^{\gamma_1 \wedge \gamma_2} = \phi_\lambda^{\gamma_1} + \gamma_1.\phi_\lambda^{\gamma_2}. \tag{3.56}$$

One more time, the cocycles are coboundaries; the following formula is easily verified

$$\phi_\lambda^{\gamma_1} = (\delta\psi_\lambda)[\gamma_1] = \pi_\star P_1.\psi_\lambda - \psi_\lambda; \tag{3.57}$$

where

$$\psi_\lambda = -\phi_\lambda^{\mathsf{Id}_\lambda}. \tag{3.58}$$

The new interesting point is the definition of a *mutual information*. For that we mimic the formulas of Shannon theory: we apply a combinatorial operator to the ambiguity. Then we consider the canonical bar resolution for $\mathsf{Ext}^\star_{\mathcal{D}}(\mathbb{K}, \Phi)$, with the trivial action of $\mathcal{A}'|\lambda; \lambda \in \widetilde{\mathcal{A}}$. The operator is the combinatorial coboundary δ^t at degree two, and it gives:

$$I_\lambda(\gamma_1;\gamma_2) = \delta^t\phi_\lambda[\gamma_1,\gamma_2] = \phi_\lambda^{\gamma_1} - \phi_\lambda^{\gamma_1\wedge\gamma_1} + \phi_\lambda^{\gamma_2}. \tag{3.59}$$

This gives the following formulas

$$I_\lambda(\gamma_1;\gamma_2) = \phi_\lambda^{\gamma_1} - \gamma_2.\phi_\lambda^{\gamma_1} = \phi_\lambda^{\gamma_2} - \gamma_1.\phi_\lambda^{\gamma_2}. \tag{3.60}$$

More concretely, for two morphisms $\gamma_1 : \lambda_1 \to \lambda$ and $\gamma_2 : \lambda_2 \to \lambda$, denoting by P_1, P_2 their respective coordinates on propositions, and by $\psi_\lambda = -\phi_\lambda^\lambda$ the canonical 0-cochain, we have:

$$I_\lambda(\gamma_1;\gamma_2)(T) = \psi_\lambda(T|\pi_\star P_2) + \psi_\lambda(T|\pi_\star P_1) - \psi_\lambda(T|\pi_\star P_1 \wedge \pi_\star P_2) - \psi_\lambda(T)$$

Remark. We decided that the interpretation of ϕ_λ is better when ψ_λ is growing. Now, assuming the positivity of I_λ, we get a kind of concavity of ψ_λ.

More generally, we say that a real function ψ of the theories, containing $\vdash \neg P$, in a given language, is *concave* (*resp.* strictly concave), if for any pair of such theories $T \leq T'$ and any proposition $Q \geq P$, the following expression is positive (*resp.* strictly positive),

$$I_P(Q;T,T') = \psi(T|Q) - \psi(T) - \psi(T'|Q) + \psi(T'). \tag{3.61}$$

Remark that this definition extends *verbatim* to any closed monoidal category, because it uses only the pre-order and the exponential.
The positivity of the mutual information is the particular case where $T' = T|Q_1$.

This makes ψ look like the function log $(\ln P)$ for a domain $\perp < P \leq \neg P$, analog of the interval $]0,1[$ in the propositional context.

The functions ψ_λ can always be chosen such that $\phi_\lambda^P = -\psi_\lambda$. Then the above interpretation of ϕ as an informational ambiguity is compatible with an interpretation of $\psi(T)$ as a measure of the *precision* of the theory.

The Boolean case, comparing to Carnap and Bar-Hillel [CBH52]

In the finite Boolean case, *the opposite of the content* defined by Carnap and Bar-Hillel gives such a function ψ, strictly increasing and concave. Remind that the content set $C(T)$ is the set of elementary propositions that are excluded by the theory T. Here we assimilate a theory with the language and its axioms, and with a subset of a finite set E. If $T < T'$, there is less excluded points by T' than by T, then $-c(T') - (-c(T)) > 0$. If $P \leq Q$, the content set of $T \vee \neg Q$ is the intersection of $C(T)$ and $C(\vdash \neg Q) = C(Q)^c$, and the content of $T' \vee \neg Q$ the intersection of $C(T)$ and $C(\vdash \neg Q) = C(Q)^c$, then the complement of $C(T' \vee \neg Q)$ in $C(T')$ is contained in the complement of $C(T \vee \neg Q)$ in $C(T)$. Consequently

$$\psi(T|Q) - \psi(T) - (\psi(T'|Q) - \psi(T')) = c(T) - c(T|Q) - (c(T') - c(T'|Q)) \geq 0. \tag{3.62}$$

It is zero when $T' \wedge (\neg Q) \leq T$.
A natural manner to obtain a strictly concave function is to apply the logarithm function to the function $(c_{\max} - c(T))/c_{\max}$.
Therefore a natural formula in the boolean case is

$$\psi_P(\mathbb{T}) = \ln \frac{c(\perp) - c(\mathbb{T})}{c(\perp) - c(\neg P)} \tag{3.63}$$

But we also could take a uniform normalization:

$$\psi_\perp(\mathbb{T}) = \ln \frac{c(\perp) - c(\mathbb{T})}{c(\perp)} \tag{3.64}$$

Amazingly, this was the definition of the amount of information (with a minus sign) of Carnap and Bar-Hillel [CBH52].
A generalization along their line consists to choose any strictly positive function m of the elementary propositions and to define the numerical content $c(T)$ as the sum of the values of m over the elements excluded by T. This corresponds to the attribution of more or less value to the individual elements.
We essentially recover the basis of the theory presented by Bao, Basu et al. [BBD+11], [BBDH14].

Question. *Does a natural formula exist, that is valid in every Heyting algebra, or at least in a class of Heyting algebras larger than Boole algebras?*

Example. The open sets of a topology on a finite set X. The analog of the content of T is the cardinality of the closed set $X \setminus T$. Then a preliminary function ψ is the cardinality of T itself, which is naturally increasing with T. However simple examples show that this function can be non-concave. The set $T|Q\backslash T$ is made by the points x of $X\backslash T$ having a neighborhood V such that $V \cap V \subset T$, there exists no relation between this set and the analog set for T' larger than T, but smaller than $\neg P$.

However, appendix D constructs a good function ψ for the sites of DNNs and the injective finite sheaves. This applies in particular to the chains $0 \to 1 \to \ldots \to n$.

A remark on semantic independency

In their 1952 report [CBH52], Carnap and bar-Hillel gave a different justification than us for taking the logarithm of a normalized version of the content. This was in the Boolean situation, $n = 0$, but our appendix D extends what they said to some non-Boolean situations.

They had in mind that independent assertions must give an addition of the amounts of information of the separate assertions. However, as they recognized themselves, the concept of semantic independency is not very clear [CBH52, page 12]. In fact they studied a particular case of typed language that they named $\mathcal{L}_n^\pi$, where there exists one type of subjects with n elements, $a, b, c, \ldots$, that can have a given number π of attributes (or predicate). The example is three humans, their gender (male or female), and their age (old or young). For every elementary proposition Z_i, i.e. a point inn E, they choose a number $m_P(Z_i)$ in $]0, 1|$, and define, as in the preceding section with μ, the function m of any proposition L, by taking the sum of the m_i over the elements of L, viewed as a subset of E.

Carnap and Bar-Hillel imposed several axioms on m_P, for instance the invariance under the natural action of the symmetry group $\mathfrak{S}_n \times \mathfrak{G}_\pi$, where $\mathfrak{G}_\pi$ describes the symmetries between the predicates, and the normalization by $m(E) = 1$. The *content* is an additive normalization of the opposite of m. The number $c(L)$ evaluates the quantity of elementary propositions excluded by L.

At some moment, they introduce axiom h, [CBH52, page 14], $m(Q \wedge R) = m(Q)m(R)$, if Q and R do not consider any common predicate. This axiom was rarely considered in the rest of the paper. However it is followed by a definition: two assertions S and T were said inductively independent (with respect to m_P) if an only if

$$m(S \wedge T) = m(S)m(T). \tag{3.65}$$

This was obviously inspired from the theory of probabilities [Car50], where primitive predicates are considered in relation to probabilities.

If we think of the example with the age and the gender, the axiom is not very convincing from the point of view of probability, because in most sufficiently large population of humans it is not true that age and gender are independent. However, from a *semantic point of view*, this is completely justified!

Now, if we come to the amount of information, taking the logarithm of the inverse of $m(T)$ to measure $\mathsf{inf}(T)$ makes that independency (inductive) is equivalent to the additivity:

$$\psi(S \wedge T) = \psi(S) + \psi(T). \tag{3.66}$$

Under this form, the definition still has a meaning, for any function ψ. Even with values in a category of models, with a good notion of colimit, as the disjoint union of sets.

In Shannon's theory, with the set theoretic interpretation of Hu Kuo Ting, [Tin62], we recover the same thing.

Comparison of information between layers

Another way to obtain a comparison between layers, i.e. objects (U, ξ), comes from the ordinary cohomology of the object Φ in the topos of presheaves over the opposite category of $\widetilde{\mathcal{A}}'_{\text{strict}}$, that we named $\mathcal{D}$.

This cohomology can be computed following the method exposed by Grothendieck and Verdier in SGA 4 [AGV63], using a canonical resolution of Φ. This resolution is constructed from the nerve $\mathcal{N}(\mathcal{D})$, made by the sequences of arrows $\lambda \to \lambda_1 \to \lambda_2 \ldots$ in $\widetilde{\mathcal{A}}'_{\text{strict}}$, then associated to the fibration by the slices category $\lambda|\mathcal{D}$ over $\mathcal{D}$. Be carefull that in $\mathcal{D}$, the arrows are in reverse order.

The nerve $\mathcal{N}(\mathcal{D})$ has a natural structure of simplicial set whose n simplices are sequences of composable arrows $(\gamma_1, \ldots, \gamma_n)$ between objects $\lambda_0 \to \cdots \to \lambda_n$ in $\widetilde{\mathcal{A}}'_{\text{strict}}$, and whose face operators $d_i; i = 0, \ldots, n$ are given by the following formulas:

$$\begin{aligned}
d_0(\gamma_1, \ldots, \gamma_n) &= (\gamma_2, \ldots, \gamma_n) \\
d_i(\gamma_1, \ldots, \gamma_n) &= (\gamma_1, \ldots, \gamma_{i+1} \circ \gamma_i, \ldots, \gamma_n) \text{ if } 0 < i < n \\
d_n(\gamma_1, \ldots, \gamma_n) &= (\gamma_1, \ldots, \gamma_{n-1}).
\end{aligned}$$

This allows to define a *canonical cochain complex* $(C^n(\mathcal{D},\Phi),d)$ which cohomology is $H^\star(\mathcal{D},\Phi)$.

The n-cochains are

$$C^n(\mathcal{D},\Phi) = \prod_{\lambda_0\to\cdots\to\lambda_n} \Phi_{\lambda_n} \tag{3.67}$$

and the coboundary operator $\delta : C^{n-1}(\mathcal{D},\Phi) \to C^n(\mathcal{D},\Phi)$ is given by

$$(\delta\phi)_{\lambda_0\to\cdots\to\lambda_n} = \sum_{i=0}^{n-1}(-1)^i\phi_{d_i(\lambda_0\to\cdots\lambda_n)} + (-1)^n\Phi_\star(\gamma_n)\phi_{d_n(\lambda_0\to\cdots\lambda_n)}. \tag{3.68}$$

For instance at degree zero, this gives, for $\gamma : \lambda \to \lambda'$

$$\delta\phi^0_\gamma(S') = \phi^0_{\lambda'}(S') - \phi^0_\lambda(\pi^\star S'). \tag{3.69}$$

For our cocycle ϕ^Q_λ, with $P \leq Q$, a more convenient sheaf over $\mathcal{D}$ is given by the sets Ψ_λ of functions of the pairs (S,Q), with S excluding P and P implying Q, with morphisms

$$\Psi_\star(\gamma)(S',Q') = \psi(\pi^\star S',\pi^\star Q'). \tag{3.70}$$

This gives

$$\delta\phi^0_\gamma(S',Q') = \phi^0_{\lambda'}(S',Q') - \phi^0_\lambda(\pi^\star S',\pi^\star Q'). \tag{3.71}$$

In our case, with $\phi^0_\lambda(S,Q) = \phi^Q_\lambda(S)$, we get the measure of the evolution of the ambiguity along the network.

From now on, we change topic and consider the reverse direction of propagation of theories and propositions.

The particular case of natural isomorphisms

Until the end of this subsection, we consider the particular case of isomorphisms between the logics in the layers, i.e. $\pi^\star\pi_\star = \mathsf{Id}_U$ and $\pi_\star\pi^\star = \mathsf{Id}_{U'}$.
As we will see, this is rather deceptive, giving a particular case of the preceding notion of ambiguity and information, obtained without the hypothesis of isomorphism, then it can be skipped easily, but it seemed necessary to explore what possibilities were offered by the contravariant side of $\widetilde{A}$.

In this case we are allowed to consider the sheaf of propositions $\mathcal{A}$ for $\pi^\star$ together and the cosheaf of theories Θ for $\pi_\star$ over the category $\widetilde{\mathcal{A}}$. The action of $\mathcal{A}$ by conditioning on the sheaf Φ of measurable functions on Θ is natural, (see proposition 3.3).
Thus we can apply the same strategy as before, using the bar complex.

The zero cochains satisfy

$$\psi_{\lambda'}(\pi_\star T) = \psi_\lambda(T). \tag{3.72}$$

This equation implies the naturality (3.29). However, there is a difference with the preceding framework, because we have more morphisms to take in account, i.e. the implications $P \leq P'$. This implies that, for U, ξ fixed, ϕ does not depend on P; there exists a function $\psi_{U,\xi}$ on all the theories such that ψ_λ on $\Theta(U, \xi, P)$ is its restriction.

Proof: for any pair $P \leq Q$ in $\mathcal{A}_\lambda$, and any theory which excludes Q then P, we have $\psi_P(S) = \psi_Q(S)$. Therefore $\psi_P = \psi_\perp$.

The equation of cocycle is the same as before, i.e. (3.30). It implies that $\psi_{U,\xi}$ is invariant by the action of $\mathcal{A}_\lambda$. In every case, boolean or not, this implies that $\phi_{U,\xi}$ is also independent of the theory T. Therefore the H^0 now simply counts the sections of $\mathcal{F}$.

The degree one cochains satisfy

$$\phi^{R'}_{\lambda'}(\pi_\star S) = \phi^{\pi^\star R'}_\lambda(S). \tag{3.73}$$

In particular, for any triple $P \leq Q \leq R$, and any $S \in \Theta_P$, we have

$$\phi^R_{U,\xi,Q}(S) = \phi^R_{U,\xi,P}(S), \tag{3.74}$$

which allows us to consider only the elements of the form ϕ^P_λ, that we denote simply ϕ_λ.

The cocycle equation is as before, (3.33): And taking $\psi_\lambda = -\phi_\lambda$ gives canonically a zero whose coboundary is ϕ:

$$\phi^Q_\lambda(S) = \psi_\lambda(S) - \psi_\lambda(S|Q). \tag{3.75}$$

Which defines the dependency of ϕ in Q.

The naturality, in the case of isomorphisms, for a connected network, with a unique output layer, tells that everything can be computed in the output layer. The intervention of the layers is illusory. Then it is sufficient to consider the case of one layer and logical calculus.

What follows is only a verification that things transport naturally to the whole category $\widetilde{A}$.

The extension of monoids is made via the left slices categories $\lambda|\mathcal{A}$; the action of $\lambda|\mathcal{A}$ on Θ_λ is given by

$$\gamma.\mathbb{T} = (\pi^\star_\gamma P' \Rightarrow \mathbb{T}) = \mathbb{T}|\pi^\star_\gamma P' \tag{3.76}$$

where $\gamma : \lambda \to \lambda'$, $\lambda = (U, \xi, P)$, $\lambda' = (U', \xi', P')$, $P \leq \pi^\star P'$, and $\pi_\gamma = (\alpha, h)$ is the projected morphism of $\mathcal{F}$.

This defines an action of the monoid of propositions in $\mathcal{A}_{\lambda'}$ which are implied by P'. If $P' \leq Q'$ and $P' \leq R'$, we have $\pi^\star_\gamma(Q' \wedge R') = \pi^\star_\gamma(Q') \wedge \pi^\star_\gamma(R')$.

A natural structure of monoid is given by

$$\gamma_1.\gamma_2 = (U, \xi, \pi^\star\gamma_1 \wedge \pi^\star\gamma_2). \tag{3.77}$$

This works because, for a morphism $\gamma : \lambda \to \lambda'$, we have $P \leq \pi_\gamma^\star P'$.
The identity is the truth $\top_\lambda$.

Lemma 3.5. *The naturality of the operations over $\mathcal{A}'$ follows from the further hypothesis: for every morphism (α, h), we assume that the counit $\pi^\star\pi_\star$ is equal to $\mathsf{Id}_{\mathbb{L}_{U,\xi}}$.*

Proof. Consider an arrow $\rho : \lambda \to \lambda_1$; it gives a morphism $\rho^\star : \lambda_1|\mathcal{A} \to \lambda|\mathcal{A}$.
For a morphism $\gamma_1 : \lambda_1 \to \lambda_1'$, $\rho^\star(\lambda_1) = \gamma_1 \circ \rho$.
If $\gamma_1 : \lambda_1 \to \lambda_1'$ is an arrow in $\mathcal{A}'$, where $\lambda_1' = (U_1', \xi_1', P_1')$, and T a theory in Θ_λ, we have

$$\begin{aligned}
\rho^\star(\gamma_1).T &= \pi^\star_{\gamma_1\circ\rho}P_1' \Rightarrow T \\
&= \pi^\star_\rho\pi^\star_{\gamma_1}P_1' \Rightarrow \pi^\star_\rho(\pi_\rho)_\star T \\
&= \pi^\star_\rho[\pi^\star_{\gamma_1}P_1' \Rightarrow (\pi_\rho)_\star T] \\
&= \pi^\star_\rho[\gamma_1.(\pi_\rho)_\star T] \\
&= \rho^\star(\gamma_1.\rho_\star T)
\end{aligned}$$

■

The monoids $\lambda|\widetilde{\mathcal{A}}$ is a presheaf over $\widetilde{\mathcal{A}}$, only in the case of isomorphisms, i.e. $\pi_\star\pi^\star = \mathsf{Id}_{\lambda'}$.

The bar construction now makes appeal to symbols $[\gamma_1|\gamma_2|...|\gamma_n|$, where the γ_i are arrows issued from λ. The action of algebra pass through the inverse image of propositions $\pi^\star P_i$.
The zero cochains are families ϕ_λ of maps on theories satisfying

$$\psi_\lambda(T) = \psi_{\lambda'}(\pi_\star T), \tag{3.78}$$

where $\gamma : \lambda \to \lambda'$ is a morphism in $\widetilde{A}$.
The coboundary operator is

$$\delta\psi_\lambda([\gamma_1]) = \psi_\lambda(T|\pi^\star_\gamma P_1) - \psi_\lambda(T). \tag{3.79}$$

Then the cohomology is as before.
The one cochains are collections of maps of theories $\phi_\lambda^{\gamma_1}$ satisfying

$$\phi_{\lambda'}^{\gamma_1'}(\pi_\star T) = \phi_\lambda^{\gamma_1'\circ\gamma}(T). \tag{3.80}$$

The cocycle equation is

$$\phi_\lambda^{\gamma_1\wedge\gamma_2} = \phi_\lambda^{\gamma_1} + \gamma_1.\phi_\lambda^{\gamma_2}. \tag{3.81}$$

One more time, the cocycles are coboundaries; the following formula is easily verified

$$\phi_\lambda^{\lambda_1} = (\delta\psi_\lambda)[\lambda_1] = \pi^\star P_1.\psi_\lambda - \psi_\lambda; \tag{3.82}$$

where

$$\psi_\lambda = -\phi_\lambda^{Id_\lambda}. \tag{3.83}$$

The combinatorial coboundary δ^t at degree two gives:

$$I_\lambda(\gamma_1;\gamma_2) = \delta^t \phi_\lambda[\gamma_1,\gamma_2] = \phi_\lambda^{\gamma_1} - \phi_\lambda^{\gamma_1\wedge\gamma_1} + \phi_\lambda^{\gamma_2}. \tag{3.84}$$

This gives the following formulas

$$I_\lambda(\gamma_1;\gamma_2) = \phi_\lambda^{\gamma_1} - \gamma_2.\phi_\lambda^{\gamma_1} = \phi_\lambda^{\gamma_2} - \gamma_1.\phi_\lambda^{\gamma_2}. \tag{3.85}$$

More concretely, for two morphisms $\gamma_1 : \lambda_1 \to \lambda$ and $\gamma_2 : \lambda_2 \to \lambda$, denoting by P_1, P_2 their respective coordinates on propositions, and by $\psi_\lambda = -\phi_\lambda^\lambda$ the canonical 0-cochain, we have:

$$I_\lambda(\gamma_1;\gamma_2)(T) = \psi_\lambda(T|\pi^\star P_1 \wedge \pi^\star P_2) - \psi_\lambda(T|\pi^\star P_1) - \psi_\lambda(T|\pi^\star P_2) + \psi_\lambda(T) \tag{3.86}$$

In a unique layer U, for a given context ξ, we get

$$I(P_1;P_2)(T) = \psi(T|P_1 \wedge P_2) - \psi(T|P_1) - \psi(T|P_2) + \psi(T). \tag{3.87}$$

This is the particular case of the mutual information we got before, see equation (3.59), because now, the generating function ψ is the restriction to $\Theta(P)$ of a function that is defined on $\Theta = \Theta(\perp)$.

3.5 Homotopy constructions

Abelian homogeneous bar complex of information

We start by describing an homogeneous version of the information cocycles, giving first the differences of ambiguities, from which the above ambiguity can be derived by reducing redundancy. For that purpose we consider equivariant cochains as in [BB15].

The sets Θ_λ, where $\lambda = (U,\xi,P)$, are now extended by the symbols $[\gamma_0|\gamma_1|...|\gamma_n]$, where $n \in \mathbb{N}$, and the $\gamma_i; i = 0,...,n$, are objects of the category $\mathcal{D}_\lambda$ or arrows in $\widetilde{\mathcal{A}}'_{\text{strict}}$ abutting to $\lambda_R = (U,\xi,R)$ for $P \leq R$. This extension with $n+1$ symbols is denoted by Θ_λ^n. It represents the possible theories in the local language and its context U,ξ, excluding the validity of P, augmented by the possibility to use counter-examples $\neg Q_i, i = 0,...,n$. There is a natural simplicial structure on the union $\Theta_\lambda^\bullet$ of these sets. The face operators $d_i; i = 0,...,n$ being given by the following formulas:

$$\begin{aligned}
d_0(\gamma_0,...,\gamma_n) &= (\gamma_1,...,\gamma_n) \\
d_i(\gamma_0,...,\gamma_n) &= (\gamma_0,...,\gamma_{i-1},\gamma_{i+1}...,\gamma_n) \text{ if } 0 < i < n \\
d_n(\gamma_0,...,\gamma_n) &= (\gamma_0,...,\gamma_{n-1}).
\end{aligned}$$

By definition, the geometric realization of $\Theta_\lambda^\bullet$ is named the space of theories at λ or localized at λ. Its homotopy type is named the *algebraic homotopy type* of theories, also at λ.

Remind that a simplicial set K is a presheaf over the category Δ, with objects $\mathbb{N}$ and morphisms from m to n, the non decreasing maps from $[m] = \{1,...,m\}$ to $[n] = \{1,...,n\}$. The *geometric realization* $|K|$ of a simplicial set K is the topological space obtained by quotienting the disjoint union of the products $K_n \times \Delta(n)$, where $K_n = K([n])$ and $\Delta(n) \subset \mathbb{R}^{n+1}$ is the geometric standard simplex, by the equivalence relation that identifies $(x, \varphi_\star(y))$ and $(\varphi^\star(x), y)$ for every nondecreasing map $\varphi : [m] \to [n]$, every $x \in K_n$ and every $y \in \Delta(m)$; here $f^\star$ is $K(f)$ and $f_\star$ is the restriction to $\Delta(n)$ of the unique linear map from $\mathbb{R}^{n+1}$ to $\mathbb{R}^{m+1}$ that sends the canonical vector e_i to $e_{f(i)}$. In this construction, for $n \in \mathbb{N}$, K_n is equipped with the discrete topology and $\Delta(n)$ with its usual topology, then compact, the topology on the union over $n \in \mathbb{N}$ is the weak topology, i.e. a subset is closed if and only if its intersection with each closed simplex is closed, and the realization is equipped with the quotient topology, the finest making the quotient map continuous. In particular, even it is not obvious at first glance, the realization of the simplicial set Δ^k is the standard simplex $\Delta(k)$.

Let $\mathbb{K}$ be commutative ring of cardinality at most continuous (conditions of measurability will be considered later). We consider the rings $\Phi^n_\lambda; n \in \mathbb{N}$ of (measurable) functions on the respective Θ^n_λ with values in $\mathbb{K}$.
The above simplicial structure gives a differential complex on the graded sum $\Phi^\bullet_\lambda$ of the $\Phi^n_\lambda; n \in \mathbb{N}$, with the simplicial (or combinatorial) coboundary operator

$$(\delta_\lambda \phi)^{\gamma_0|\cdots|\gamma_n}_\lambda = \sum_{i=0}^{n} (-1)^i \phi^{\gamma_0|\cdots|\widehat{\gamma_i}|\cdots|\gamma_n}. \tag{3.88}$$

We call *algebraic cocycles* the elements in the kernel.

As we have seen, the arrows $\gamma_Q \in \mathcal{D}_\lambda$ can be multiplied, using the operation $\wedge$ on propositions in $\mathcal{A}_\lambda$, and this defines an action of monoid on Θ_λ by the conditioning operation. Therefore we can define the *homogeneous functions* or *homogeneous algebraic cochains* of degree $n \in \mathbb{N}$ as the (measurable) functions $\phi^{\gamma_0;\gamma_1;...;\gamma_n}_\lambda$ on Θ_λ, such that for any γ_Q in $\mathcal{D}_\lambda$, abutting in (U, ξ, Q), for $P \leq Q$, and any $T \in \Theta_\lambda$, thus excluding P,

$$\phi^{\gamma_Q \wedge \gamma_0;\gamma_Q \wedge \gamma_1;...;\gamma_Q \wedge \gamma_n}_\lambda(T) = \phi^{\gamma_0;\gamma_1;...;\gamma_n}_\lambda(T|Q). \tag{3.89}$$

The above operator δ_λ preserves the homogeneous algebraic cochains. The kernel restriction of δ_λ defines the *homogeneous algebraic cocycles*.

A morphism $\gamma : \lambda \to \lambda'$ naturally associates $\phi^{\gamma_0|\gamma_1|...|\gamma_n}_\lambda$ with $\phi^{\gamma'_0|\gamma'_1|...|\gamma'_n}_{\lambda'}$ through the formula

$$\phi^{\gamma_0|\gamma_1|...|\gamma_n}_\lambda(\pi^\star T') = \phi^{\gamma_\star\gamma_0|\gamma_\star\gamma_1|...|\gamma_\star\gamma_n}_{\lambda'}(T'). \tag{3.90}$$

Then the hypothesis $\pi^\star \pi_\star = \mathsf{Id}_{U',\xi'}$ allows to define a cosheaf $\Phi^n_\lambda; \lambda \in \mathcal{D}$ over $\mathcal{D}$, not a sheaf, by

$$(\Phi_\star \phi_{\lambda'})^{\gamma_0|\gamma_1|...|\gamma_n}(T) = \phi^{\gamma_\star\gamma_0|\gamma_\star\gamma_1|...|\gamma_\star\gamma_n}_{\lambda'}(\pi_\star T). \tag{3.91}$$

However the first equation (3.90) is more precise, and we take it as a definition of *natural algebraic homogeneous cochains*.

Remark. We cannot consider it as a sheaf because of a lack of definition of $\gamma^\star \gamma'_i$.

The operation of conditioning preserves the naturality, in reason of the following identity, involving $\gamma : \lambda \to \lambda'$, $\gamma_Q \in \mathcal{D}_\lambda$, $S' \in \Theta^n_{\lambda}$:

$$\pi^\star_\gamma [S'|\gamma_\star(\gamma_Q)] = \pi^\star_\gamma S'|\gamma_Q. \tag{3.92}$$

Therefore we can speak of *natural homogeneous algebraic cocycles*.

For $n = 0$, the cochains are collections of functions $\psi^{\gamma_0}_\lambda$ of the theories in $\mathcal{A}_\lambda$ such that

$$\psi^{\gamma_Q \wedge \gamma_0}_\lambda (S) = \psi^{\gamma_0}_\lambda (S|Q), \tag{3.93}$$

and such that, for any morphism $\gamma : \lambda \to \lambda'$,

$$\psi^{\gamma_0}_\lambda (\pi^\star_\gamma T') = \psi^{\gamma_\star \gamma_0}_{\lambda'} (T'). \tag{3.94}$$

From the first equation, we can eliminate γ_0. We define $\psi_\lambda = \psi^\top_\lambda$, and get

$$\psi^{\gamma_Q}_\lambda (S) = \psi_\lambda (S|Q). \tag{3.95}$$

The second equation, with the transport of truth, is equivalent to

$$\psi_\lambda (\pi^\star_\gamma T') = \psi_{\lambda'} (T'). \tag{3.96}$$

A cocycle corresponds to a collection of constant c_λ, which are natural, then to the functions of the connected components of the category $\mathcal{D}$.
Thus we recover the same notion as in the preceding section.

In degree one, the homogeneous cochain $\phi^{\gamma_0;\gamma_1}_\lambda$ cannot be *a priori* expressed through the collection of functions $\varphi^{\gamma_0}_\lambda = \phi^{\gamma_0;\top}_\lambda$, but, if it is a cocycle, it can:

$$\phi^{\gamma_0;\gamma_1}_\lambda = \varphi^{\gamma_0}_\lambda - \varphi^{\gamma_1}_\lambda; \tag{3.97}$$

as this follows directly from the algebraic cocycle relation applied to $[\gamma_0|\gamma_1|\top_\lambda]$.

But we also have, by homogeneity

$$Q.\varphi^{\gamma_Q} = Q.\phi^{\gamma_Q|\top} = \phi^{\gamma_Q \wedge \gamma_Q|\gamma_Q \wedge \top} = \phi^{\gamma_Q|\gamma_Q} = \varphi^{\gamma_Q} - \varphi^{\gamma_Q} = 0. \tag{3.98}$$

Then, the homogeneity equation gives the particular case

$$\varphi^{Q \wedge Q_O} - \varphi^{Q \wedge Q} = Q.\varphi^{\gamma_{Q_O}} - Q.\varphi^{\gamma_Q} = Q.\varphi^{\gamma_{Q_O}}, \tag{3.99}$$

therefore

$$\varphi^{Q\wedge Q_O} = \varphi^{\gamma_Q} + Q.\varphi^{\gamma_{Q_O}}; \tag{3.100}$$

which is the cocycle equation we discussed in the preceding section, under the form of Shannon.

Remark. All that generalizes to any degree, in virtue of the comparison theorem between projective resolutions, proved in the relative case in MacLane "Homology" [Mac12], or in SGA 4 [AGV63], more generally, because the above homogeneous bar complex and in-homogeneous bar complex are such resolutions of the constant functor $\mathbb{K}$.

Semantic Kullback-Leibler distance

In [BB15], it was also shown that the Kullback-Leibler distance (or divergence) $D_{KL}(X;\mathbb{P};\mathbb{P}')$ between two probability laws on a random variable X defines a cohomology class in the above sense. The cochains depend on a sequence $\mathbb{P}_0,...,\mathbb{P}_n$ of probabilities and a sequence of variables $X_0,...,X_m$ less fine than a given variable X; the conditioning the $n+1$ laws by the value y of a variable $Y \geq X$ is integrated over $Y_\star\mathbb{P}_0$, for giving an action on the set of measurable functions of the $n+1$ laws, then the homogeneity is defined as before, and the coboundary is the standard combinatorial one, as before. For $n=1$, the universal degree one class is shown to be the difference of divergences.
Remind that the $K-L$ divergence is given by the formula

$$D_{KL}(X;\mathbb{P};\mathbb{P}') = -\sum_{x_i} p_i \log \frac{p'_i}{p_i}. \tag{3.101}$$

In our present case, we consider functions of $n+1$ theories and $m+1$ propositions, all works as for $n=0$. In degree zero, the cochains are defined by functions $\psi_\lambda(S_0,S_1)$ satisfying

$$\psi_\lambda(\pi^\star_\gamma S'_0;...;\pi^\star_\gamma S'_n) = \psi_{\lambda'}(S'_0;...;S'_n), \tag{3.102}$$

for any morphism $\gamma:\lambda\rightarrow\lambda'$.
The formula for the homogeneous cochain is

$$\psi^{\gamma_Q}_\lambda(S_0;...;S_n) = \psi_\lambda(S_0|Q;...;S_n|Q). \tag{3.103}$$

The non-homogeneous zero cocycles are the functions of P only, invariant by the transport $\pi_\star$.
In degree one, the cocycles are defined by any function $\varphi^Q_\lambda(S_0;...;S_n)$ which satisfies

$$\varphi^Q_\lambda(\pi^\star_\gamma S'_0;...;\pi^\star_\gamma S'_n) = \varphi^{\pi_\star(Q)}_{\lambda'}(S'_0;...;S'_n), \tag{3.104}$$

for any morphism $\gamma:\lambda\rightarrow\lambda'$, and verifies the cocycle equation

$$\varphi^{Q\wedge R}_\lambda(S_0;...;S_n) = \varphi^Q_\lambda(S_0;...;S_n) + \varphi^R_\lambda(S_0|Q;...;S_n|Q). \tag{3.105}$$

The homogeneous cocycle associated to φ is defined by

$$\phi_{\lambda}^{\gamma_{Q_0};\gamma_{Q_1}}(S_0;...;S_n) = \varphi_{\lambda}^{Q_0}(S_0;...;S_n) - \varphi_{\lambda}^{Q_1}(S_0;...;S_n). \tag{3.106}$$

As for $n = 0$, there exists a function $\psi_\lambda(S_0;...;S_n)$ such that for any $Q \in \mathcal{A}_\lambda$, i.e. $Q \geq P$, we have

$$\varphi_{\lambda}^{Q}(S_0;...;S_n) = \psi_\lambda(S_0|Q;...;S_n|Q) - \psi_\lambda(S_0;...;S_n). \tag{3.107}$$

In the particular case $n = 1$, we can consider a basic real function $\psi_\lambda(S)$, seen as a logarithm of theories as before, and define

$$\psi_\lambda(S_0;S_1) = \psi_\lambda(S_0 \wedge S_1) - \psi_\lambda(S_0). \tag{3.108}$$

If the function $\psi_\lambda(S)$ is supposed increasing in S (for the relation of weakness $\leq$, as before), this gives a negative function.
We obtain

$$\phi_{\lambda}^{Q}(S_0;S_1) = \psi_\lambda(S_0 \wedge S_1|Q) - \psi_\lambda(S_0 \wedge S_1) - \psi_\lambda(S_0|Q) + \psi_\lambda(S_0). \tag{3.109}$$

The positivity of this quantity is equivalent to the concavity of $\psi_\lambda(S)$ on the pre-ordered set of theories. Assuming this property we obtain an analog of the Kullback-Leibler divergence.
If $\psi_\lambda(S)$ is strictly concave, that is the most convenient hypothesis, this function takes the value zero if and only if $S_0 = S_1$. Therefore it can be taken as a natural *semantic distance*, depending on the data of Q, as candidate from a counter-example of P.
As in the case of D_{KL} this function is not symmetric, then it could be more convenient to take the sum

$$\sigma_{\lambda}^{Q}(S_0;S_1) = \phi_{\lambda}^{Q}(S_0;S_1) + \phi_{\lambda}^{Q}(S_1;S_0) \tag{3.110}$$

to have a good notion of distance between two theories.

Simplicial homogeneous space of histories of theories

Another argument to justify the consideration of the homogeneity is the interest of taking a *pushout* of the theories.

The sheaf of monoidal categories $\mathcal{D}_\lambda$ over $\mathcal{D}$ acts in two manners on the algebraic space of theories $\Theta_{\lambda}^{\bullet}$:

$$\gamma_Q.(S \otimes [\gamma_0;...;\gamma_n]) = (S|Q) \otimes [\gamma_0;...;\gamma_n], \tag{3.111}$$

$$\gamma_Q \wedge (S \otimes [\gamma_0;...;\gamma_n]) = S \otimes [\gamma_Q\gamma_0;...;\gamma_Q\gamma_n]. \tag{3.112}$$

Then we can consider the colimit $\Theta_{\lambda}^{\bullet}/\mathcal{D}$ of these pairs of maps over all the arrows γ_Q, i.e. over $\mathcal{D}_\lambda$: this colimit is the disjoint union of the coequalizers for each arrow. This is a quotient simplicial set. The homogeneous cochains are just the (measurable) functions on this simplicial set.

This can be realized directly as a pushout, or coequalizer, of a unique pair of maps, by taking the union Z of the products $\Theta_\lambda^\bullet \times \mathcal{D}_\lambda$, and the two natural maps μ, ν to $T = \Theta_\lambda^\bullet$ given by multiplication and conditioning respectively.

Remark that the two operations in (3.111) and (3.112) are adjoint of each other, then we can speak of *adjoint gluing*.

Also interesting is the *homotopy quotient*, taking into account that, geometrically, Z has a higher degree in propositions belonging to $\mathcal{D}_\lambda$, due to the presence of γ_Q. This homotopy colimit is the simplicial set $\Sigma^\bullet$ obtained from the disjoint union $(Z \times [0,1]) \sqcup (T \times \{0\}) \sqcup (T \times \{1\})$ by taking the identification of $(z,0)$ with $\mu(z)$ and of $(z,1)$ with $\nu(z)$. It can be named a *homotopy gluing*, because the arrows are used geometrically as continuous links between points in $T \times \{0\}$ and $T \times \{1\}$. The simplicial set $\Sigma^\bullet$ is equipped with a natural projection onto the ordinary coequalizer $\Theta_\lambda^\bullet / \mathcal{D}_\lambda$. See for instance Dugger [Dug08] for a nice exposition of this notion, and its interest for homotopical stability with respect to the ordinary colimit. Then we propose that a more convenient notion of homogeneous cochains could be the functions on $\Sigma^\bullet$.

Similarly, we have two natural actions of the category $\mathcal{D}$ of arrows leading to λ and issued from λ': the first one being of the type

$$\Theta_{\lambda'} \otimes \mathcal{D}_\lambda^{\otimes(n+1)} \to \Theta_\lambda^n; \tag{3.113}$$

the second one of the type

$$\Theta_{\lambda'} \otimes \mathcal{D}_\lambda^{\otimes(n+1)} \to \Theta_{\lambda'}^n. \tag{3.114}$$

They are respectively defined by the following formulas:

$$\gamma^\star(S' \otimes [\gamma_0; ...; \gamma_n]) = (\pi_\gamma^\star S')_\lambda \otimes [\gamma_0; ...; \gamma_n] \tag{3.115}$$

The second one is

$$\gamma_\star(S' \otimes [\gamma_0; ...; \gamma_n]) = S' \otimes [\pi_\star^\gamma \gamma_0; ...; \pi_\star^\gamma \gamma_n] \tag{3.116}$$

They are both compatibles with the quotient by the actions of the monoids, then they define maps at the level of $\Sigma^\bullet$.

The *natural* cochains are the functions that satisfy, for each $\gamma : \lambda \to \lambda'$, the equation

$$\phi_\lambda \circ \gamma^\star = \phi_{\lambda'} \circ \gamma_\star. \tag{3.117}$$

Note that no one of the above equations, for homogeneity and naturality, necessitates numerical values, but the second necessitates values in a constant set or a constant category, at least along the orbits of $\mathcal{D}$.

And it is important for us that the cochains can take their values in a category $\mathcal{M}$ admitting limits, like Set or Top, non necessarily Abelian, because our aim is to obtain a theory of *information spaces* in the sense searched by Carnap and Bar-Hillel in 1952 [CBH52].

Define a set Θ_1^n (*resp.* Θ_0^n) by the coproduct, or disjoint union, over $\gamma : \lambda \to \lambda'$ (*resp.* λ) of the sets $\Theta_{\lambda'} \otimes \mathcal{D}_\lambda^{\otimes(n+1)}$ (*resp.* Θ_λ^n). When the integer n varies, we note the sum by $\Theta_1^\bullet$ (*resp.* $\Theta_0^\bullet$). They are canonically simplicial sets.

The collections of maps $\gamma^\star$ and $\gamma_\star$ define two (simplicial) maps from $\Theta_1^\bullet$ to $\Theta_0^\bullet$, that we will denote respectively ϖ and ϑ, for *past* and *future*. The colimit or *coequalizer* of these two maps, is the quotient $\mathrm{H}_0^\bullet$ of $\Theta_0^\bullet$ by the equivalence relation

$$(\pi_\gamma^\star S')_\lambda \otimes [\gamma_0;...;\gamma_n]_\lambda \quad \sim \quad S'_{\lambda'} \otimes [\pi_\star^\gamma \gamma_0;...;\pi_\star^\gamma \gamma_n]_{\lambda'}. \tag{3.118}$$

Once iterated over the arrows, this relation represents the complete story of a theory, from the source of its formulation in the network to the final layer.
It is remarkably conform to the notion of cat's manifold, and compatible with the possible presence of inner sources in the network.

Remark that the two operations in (3.115) and (3.116) are also adjoint relative to each other, then again the corresponding colimit can be named an adjoint gluing.

Remark. The above equivalence relation is more fine than the relation we would have found with the covariant functor, i.e.

$$(\pi_\star^\gamma S)_{\lambda'} \otimes [\pi_\star^\gamma \gamma_0;...;\pi_\star^\gamma \gamma_n]_{\lambda'} \quad \sim \quad S_\lambda \otimes [\gamma_0;...;\gamma_n]_\lambda; \tag{3.119}$$

because this relation is implied by the former, when we applied it to $S' = \pi_\star S$, in virtue of our hypothesis $\pi^\star \pi_\star = \mathsf{Id}$.

The two relations ar equivalent if and only if $\pi_\star \pi^\star = \mathsf{Id}$, that is the case of isomorphic logics among the network.

We define the *natural cochains* as the (measurable) functions on $\mathrm{H}_0^\bullet$, and the *natural homogeneous cochains* as the functions on the quotient $\mathrm{H}_0^\bullet/\mathcal{D}$ by the identification of junction with conditioning. And we are more interested in the homogeneous case.

However, in a non-Abelian context, the stability under homotopy will be an advantage, therefore we also consider the homotopy colimit of the maps ϖ and ϑ, or homotopy gluing between past and future, and propose that this colimit $\mathrm{I}_0^\bullet$ (or $ho\mathrm{I}$ if we reserve I for the usual colimit) is a better notion of the histories of theories in the network. It is also naturally a simplicial set. Then the *natural homotopy* *homogeneous cochains* will be functions on the homotopy gluing $ho\mathrm{I}$.

The homotopy type of the theories histories $\mathrm{I}_0^\bullet$ itself is an interesting candidate for representing the information, and information flow in the network.

For instance, its connected components gives the correct notion of zero-cycles, and the functions on them are zero-cocycles. The Abelian construction is sufficient to realize these cocycles.

We will later consider functions from the space $\mathrm{I}_0^\bullet$ to a closed model category $\mathcal{M}$, their homotopy type in the sense of Quillen can be seen as a non-Abelian set of cococycles.

What we just have made above for the cochains (homogeneous and/or natural) is a particular case of a *homotopy limit*.

The notion of homotopy limit was introduced in Bousfield-Kan 1972, [BK72, chapter XI] where it generalized the classical bar resolution in a non-linear context, see MacLane's book "Homology" [Mac12]. The authors attributed its origin to Milnor, in the article "On axiomatic homology theory" [Mil62]. For this notion and more recent developments (see [Hir03], [DHKS04], or [Dug08]).

In this spirit, we extend now the two maps ϖ, ϑ from $\Theta_1^\bullet$ to $\Theta_0^\bullet$, in higher degrees, by using the nerve of the category $\mathcal{D}$.

The nerve $\mathcal{N} = \mathcal{N}(\mathcal{D})$ of the category $\mathcal{D}$ is the simplicial set made by the sequences A of successive arrows in $\mathcal{D}$. For $k \in \mathbb{N}$, $\mathcal{N}_k$ is the set of sequences of length k. A sequence is written $(\delta_1, ..., \delta_k)$, where $\delta_i; i = 1, ..., k$ goes from λ_{i-1} to λ_i in $\mathcal{D}$. We use the symbols $\delta_i^\star$, or the letters γ_i when there is no ambiguity, for the arrow δ_i considered in the opposite category $\mathcal{D}^{op} = \widetilde{\mathcal{A}}'_{\text{strict}}$; this reverse the direction of the sequence, going now upstream. When necessary, we write $\delta_i(A), \lambda_{i-1}(A), ...$, for the arrows and vertices of a chain A.

For $k \in \mathbb{N}$, we define Θ_k^n as the disjoint union over $A = (\delta_1, ..., \delta_k)$ of the sets $\Theta_{\lambda_0} \otimes \mathcal{D}_{\lambda_k}^{\otimes(n+1)}$. Thus the theory is attached to the beginning in the sense of $\mathcal{D}$, and the involved propositions are at the end. The chain in $\mathcal{D}$ goes in the dynamical direction, downstream. When the integers n and k vary, we note $\Theta_\star^\bullet$ the sum (disjoint union). This is a bi-simplicial set.

We have $k+1$ canonical maps $\vartheta_i; i = 1, ..., k+1$ from Θ_{k+1}^n to Θ_k^n. Each map deletes a vertex, moreover at the extremities it also deletes the arrow, and inside the chain, it composes the arrows at $i-1$ and i. In λ_0, the map $\pi_{\gamma_1}^\star$ is applied to the theory, to be transmitted downstream, and in λ_{k+1}, the map $\pi_\star^{\gamma_{k+1}}$ is applied to the $n+1$ elements γ_{Q_j} in $\mathcal{D}_{\lambda_{k+1}}$, to be transmitted upstream.

By analogy with the definition of the homotopy colimit of a diagram in a model category cf. references upcit, we take for a more complete space of histories, the whole geometric realization of the simplicial functor $\Theta^\bullet_\star$, seen now as a *simplicial space* with the above skeleton in degree k, and the above gluing maps ϑ_i. The expression $g\mathrm{I}$ denotes this space, that we understand as the geometrical space of complete histories of theories.

The extension of information over the nerve incorporates the topology of the categories $\mathcal{C},\mathcal{F},\mathcal{D}$. The degree n was for the logic, the degree k is for its transfer through the layers.

$g\mathrm{I}$, or its homotopy type, represents for us the logical part of the available information; it takes into account

1) the architecture $\mathcal{C}$,

2) the pre-semantic structure, through the fibration $\mathcal{F}$ over $\mathcal{C}$, which constrains the possible weights, and also generates the logical transfers $\pi^\star$, $\pi_\star$,

3) the terms of a language through $\widetilde{\mathcal{A}}$, and the propositional judgements through $\mathcal{D}$ and Θ.

The dynamic is given by the semantic functioning $S^w : X^w \to \Theta$, depending on the data and the learning. Its analysis needs an intermediary, a notion of cocycles of information, that we describe now.

The information appears as a tensor $F^{\gamma_0,\dots,\gamma_n}_{\delta_1,\dots,\delta_k}(S)$. *A priori* its components take their values in the category $\mathcal{M}$, that can be Set or Top.

The points in $g\mathrm{I}$ are classes of elements

$$u = S \otimes [\gamma_0,\dots,\gamma_n] \otimes [\delta_1,\dots,\delta_k](t_0,\dots,t_n; s_1,\dots,s_k) \tag{3.120}$$

where the $t_i; i = 0,\dots,n$ and $s_j; j = 1,\dots,k$ are respectively barycentric coordinates in $\Delta(n)$ and $\Delta(k-1)$.

It is tempting to interpret the coordinates t_i as weights, or values, attributed to the propositions Q_i, and the numbers s_j as times, conduction times perhaps, along the chain of mappings.

Therefore we see the tensor F as a local system $F_u; u \in g\mathrm{I}$ over $g\mathrm{I}$.

Simplicial dynamical space of a DNN, information content

Considering a semantic functioning $S : X \to \Theta$, we can enrich it by the choice of propositions in each layer U and context ξ_U (or better collections of elements of $\mathcal{D}_\lambda$), and consider sequences over the networks, relating activities and enriched theories. Then, for each local activity, and each chain of arrows in the network, equipped with propositions at one end (downstream), the function F gives a space of information.

More precisely, we form the topological space of activities $g\mathbb{X}$, by taking the homotopy colimit of the object $\mathbb{X}$, fibred over the object $\mathbb{W}$, in the classifying topos of $\mathcal{F}$, lifted to $\mathcal{D}$, and seen as a diagram over $\mathcal{D}$. This space is defined in the same manner $gI_\star$ was defined from $\Theta_\star$ over $\mathcal{D}$; it is the geometric realization of the simplicial set $g\mathrm{X}_\star$, whose k-skeleton is the sum of the pairs (A_k, x_λ) where A is an element of length k in $\mathcal{N}(\mathcal{D})$ and x_λ an element in $\mathbb{X}_\lambda$, at the origin of A in $\mathcal{D}$. The degeneracies $d_i; i = 1, ..., k+1$ from X_{k+1} to X_k are given for $1 < i < k+1$, by composition of the morphisms at i, by forgetting $\delta_{k+1}(A)$ for $i = k+1$, and by forgetting δ_1 and transporting x_λ by $X^\star_w$ for $i = 1$.

Then we can ask for an extension of the semantic functioning to a continuous or simplicial map

$$gS : g\mathrm{X} \to g\mathrm{I}. \tag{3.121}$$

This implies a compatibility between dynamical functioning in $\mathbb{X}$ and logical functioning in Θ. However, this map factorizes by a quotient, that can be small, when the semantic functioning is poor. It is only for some regions in the weight object $\mathbb{W}$, giving itself a geometrical space $g\mathbb{W}$, that the semantic functioning is interesting.

Given $F : g\mathrm{I} \to \mathcal{M}$, this gives a map $F \circ gS$ from $g\mathrm{X}$ to $\mathcal{M}$, that can be seen as the information content of the network.

To have a better analog on the Abelian quantities, we suppose that $\mathcal{M}$ is a closed model category, and we pass to the homotopy type

$$ho.F \circ gS : g\mathrm{X} \to ho\mathcal{M}. \tag{3.122}$$

For real data inputs and spontaneous internal activities, this gives a homotopy type for each image.

For instance, the degree one homogeneous cocycle $\phi^Q_\lambda(S)$ deduced from a precision function $\psi_\lambda(S)$ with real values, is replaced by a map to topological spaces, associated to some "propositional" paths between two points of $g\mathrm{I}$; a degree two combinatorial cocycles, as the mutual information, is replaced by a varying space associated to a "propositional" triangle, up to homotopy.

Non-Abelian inhomogeneous fundamental cochains and cocycles. A tentative

Remember that the fundamental zero cochain $\psi^{Q_0}_\lambda$ with real coefficients, satisfied $\psi^Q_\lambda(S) = \psi_\lambda(S|Q) \geq \psi_\lambda(S)$. Then, in the nonlinear framework, it is tempting to assume the existence in $\mathcal{M}$ of a class of morphisms replacing the inclusions of the sets, namely cofibrations, and to generalize the increasing of the function ψ_λ of S, by the existence of a cofibration, $F(S) \rightarrowtail F(S|Q)$, or more generally a cofibration $F(S) \rightarrowtail F(S')$ each time $S \leq S'$.

This is sufficient for defining an object of ambiguity, then an information object (non-homogeneous), by generalizing the relation between precision and ambiguity of the Abelian case:

$$H^Q(S) = F(S|Q) \backslash F(S); \tag{3.123}$$

where the subtraction is taken in a geometrical or homotopical sense.
All that supposes that $\mathcal{M}$ is a closed model category of Quillen.

This invites us to assume that F is covariant under the action of the monoidal categories $\mathcal{D}_\lambda$, i.e. for every arrow γ_Q in $\mathcal{D}_\lambda$, and every theory S in Θ_λ, there exists a morphism $F(\gamma_Q;S): F(S) \to F(S|Q)$ in $\mathcal{M}$, and for two arrows γ_Q, $\gamma_{Q'}$,

$$F(\gamma_{Q'}\gamma_Q;S) = F(\gamma_{Q'};S|Q) \circ F(\gamma_Q;S) \tag{3.124}$$

and we assume that every $F(\gamma_Q;S)$ is a cofibration.

In the same manner, the generalization of the concavity of the real function ψ_λ^Q is the hypothesis that, for two arrows γ_Q, $\gamma_{Q'}$, there exists a cofibration of the quotient objects H:

$$H(Q,Q';S): H^Q(S|Q') \rightarrowtail H^Q(S). \tag{3.125}$$

The same thing happening for $H^{Q'}(S|Q) \rightarrowtail H^{Q'}(S)$.
The difference space is the model category version of the *mutual information* between Q and Q':
by definition

$$I_2(Q;Q') = H^Q \backslash [H^{Q\otimes Q'} \backslash H^{Q'}], \tag{3.126}$$

or in other terms,

$$I_2(Q;Q') = (Q.F\backslash F)\backslash[(Q\otimes Q')F\backslash Q'.F], \tag{3.127}$$

Reasoning on subsets of $H^{Q\otimes Q'}$, this gives the symmetric relation

$$I_2(Q;Q') \sim H^Q \cap H^{Q'}. \tag{3.128}$$

The general concavity condition is the existence of a natural cofibration $H^Q(S') \rightarrowtail H^Q(S)$ as soon as there is an inclusion $S \leq S'$.

This stronger property of concavity for the functor F implies in particular, for any pair of theories S_0, S_1, the existence of a cofibration

$$J_Q(S_0;S_1): H^Q(S_0) \to H^Q(S_0 \wedge S_1). \tag{3.129}$$

This allows to define a homotopical notion of Kullback-Leibler divergence space in $\mathcal{M}$, between two theories falsifying P, at a proposition $Q \geq P$:

$$D^Q(S_0;S_1) = H^Q(S_0 \wedge S_1)\backslash F_\star H^Q(S_0). \tag{3.130}$$

Comparison between homogeneous and inhomogeneous non-Abelian cochains and cocycles

To be complete, we have to relate these maps $F, H, I, D, \ldots$ from theories and constellations of propositions to $\mathcal{M}$ with the homogeneous tensors $F^{\gamma_0,\ldots,\gamma_n}_{\delta_1,\ldots,\delta_k}(S)$. For that, the natural idea is to follow the path we had described from the homogeneous Abelian bar-complex to the non-homogeneous one, at the beginning of this section. This will give a homotopical/geometrical version of the MacLane comparison in homological algebra.

We consider the bi-simplicial set $\mathbf{I}^{\bullet}_{\star}$ as a simplicial set $\mathbf{I}_{\star}$ in the algebraic exponent n for $\bullet$, then it is a contravariant functor from the category Δ to the category of simplicial sets Δ_{Set}. The morphisms of Δ from $[m]$ to $[n]$ are the non-decreasing maps, their set is noted $\Delta(m,n)$.

Our hypothesis is that the above tensors form a *cosimplicial local system* Φ with values in the category $\mathcal{M}$ over the simplicial presheaf $\mathbf{I}_{\star}$, in the sense of the preprint *Extra-fine sheaves and interaction decompositions* [BPSPV20]. In an equivalent manner, we consider the category $\mathcal{T} = \mathsf{Set}(\mathbf{I}_{\star})$ which objects are the simplicial cells u of $\mathbf{I}_{\star}$ and arrows from v of dimension n to u of dimension m are the non-decreasing maps $\varphi \in \Delta(m,n)$ (morphisms in the category Δ) such that $\varphi^{\star}(v) = u$. Here the map $\varphi^{\star}$ is simplicial in the index k for $\star$, concerning the nerve complex of $\mathcal{D}$; then the cosimplicial local system is a contravariant functor from $\mathcal{T}$ to $\mathcal{M}$.
All that is made to obtain a non-Abelian version of the propositional (semantic) bar-complex. Following a recent trend, we name *spaces* the elements of $\mathcal{M}$.

We add that an inclusion of theories $S \leq S'$ gives a cofibration $\Phi(S') \rightarrowtail \Phi(S)$, in a functorial manner over the poset of theories.

Let us repeat the arguments to go from homogeneous cochains or cocycles to non-homogeneous ones.

First, a zero-cochain is defined over the cells $S_{\lambda} \otimes [\gamma_0]$, where the arrow γ_0 abuts in a propositions $Q_0 \geq P$. The associated non-homogeneous space $F(S)$ corresponds to $Q_0 = \top$. The relation between conditioning and multiplication gives the way to recover $\Phi^{Q_0}(S)$.

Second, we name degree one homogeneous cocycle a sheaf of spaces $\Phi^{[\gamma_0,\gamma_1]}(S)$, over the one skeleton of $\varphi^{\star}$, which satisfies that for the triangle $[\gamma_0, \top, \gamma_1]$, the space $\Phi^{[\gamma_0,\gamma_1]}$ is homotopy equivalent to the *difference* of the spaces $\Phi^{[\gamma_0,\top]}$ and $\Phi^{[\gamma_1,\top]}$.
Remark: more generally a degree one cocycle should satisfies this axiom for every zigzag $\gamma_0 \leq \gamma_{\frac{1}{2}} \geq \gamma_1$.
This definition supposes that we have a notion of difference in $\mathcal{M}$, satisfying the same properties that the difference $A \backslash (A \cap B)$ satisfies in subsets of set. If all the theories considered contain a minimal one, then

spaces are subspaces of a given space, and this hypothesis has a meaning. However, this is the case in our situation, considering the sets Θ_P, because we consider only propositions $Q, Q_0, Q_1, ...$ that are implied by P.

To the degree one cocycle $\Phi^{[\gamma_0,\gamma_1]}(S)$ we associate the space $H^{\gamma_0}(S) = \Phi^{[\gamma_0,\top]}(S)$, obtained by replacing γ_1 by $\top$. The space $G^{\gamma_1}(S)$ is obtained by replacing γ_0 by $\top$ in Φ.

Note the important point that H and G are in general non-homogeneous cocycles.

Applying the definition of 1-cocycle to the triangle $[\gamma_0, \top, \gamma_1]$, we obtain that

$$\Phi^{[\gamma_0,\gamma_1]}(S) \sim H^{\gamma_0}(S) \setminus H^{\gamma_1}(S). \tag{3.131}$$

Lemma 3.6. *The cocyclicity of* Φ *implies*

$$Q.H^Q \sim H^{Q\otimes Q} \backslash H^Q. \tag{3.132}$$

Proof.

$$Q.H^Q = Q.\Phi^{Q|\top} = \Phi^{Q\otimes Q|Q\otimes\top} = \Phi^{Q\otimes Q|Q} = H^{Q\otimes Q} \backslash H^Q. \tag{3.133}$$

■

From that we deduce,

Proposition 3.7. *The homogeneity of* Φ *implies*

$$H^{Q\otimes Q'} \backslash H^{Q\otimes Q} \sim Q.H^{Q'} \backslash [H^{Q\otimes Q} \backslash H^Q]. \tag{3.134}$$

Proof.

$$H^{Q\otimes Q'} \backslash H^{Q\otimes Q} = Q.H^{Q'} \backslash Q.H^Q \sim Q.H^{Q'} \backslash [H^{Q\otimes Q} \backslash H^Q]. \tag{3.135}$$

■

In the Abelian case of ordinary difference this is equivalent to

$$H^{Q\otimes Q'} \sim Q.H^{Q'} \cup H^Q. \tag{3.136}$$

This is the usual Shannon equation; then (3.134) can be seen as a non-Abelian Shannon equation. Taking homotopy in $Ho(\mathcal{M})$ probably gives a more intrinsic meaning of semantic information.

It is natural to admit that, at the level of information spaces, $H^{Q\otimes Q} \sim H^Q$. Under this hypothesis, we get the usual Shannon's formula under

$$H^{Q\otimes Q'} \backslash H^Q \sim Q.H^{Q'}. \tag{3.137}$$

That is, for every theory S falsifying P:

$$H^{Q\otimes Q'}(S)\backslash H^{Q}(S)\sim H^{Q'}(S|Q). \tag{3.138}$$

Remind there is no reason *a priori* that $H^{Q}\rightarrowtail H^{Q\otimes Q'}$. Then the above difference is after intersection.

If F is any non-homogeneous zero-cochain, we have a cofibration $F\rightarrowtail Q.F$, where $Q.F(S)=F(S|Q)$. In this case we already defined a space H^{Q} by

$$H^{Q}(S)=F(S|Q)\backslash F(S). \tag{3.139}$$

Proposition 3.8. *H^{Q} automatically satisfies equation* (3.134).

Proof. we have $F\rightarrowtail (Q\otimes Q')F$ and $F\rightarrowtail (Q\otimes Q)F$, then

$$\begin{aligned}H^{Q\otimes Q'}\backslash H^{Q\otimes Q}&=((Q\otimes Q')F\backslash F)\backslash((Q\otimes Q)F\backslash F)\\&\sim(Q\otimes Q')F\backslash(Q\otimes Q)F.\end{aligned}$$

Using $F\rightarrowtail Q.F\rightarrowtail (Q\otimes Q)F$, and assuming $Q.F\rightarrowtail (Q\otimes Q')F$, we get

$$\begin{aligned}Q.H^{Q'}\backslash[H^{Q\otimes Q}\backslash H^{Q}]&=Q.(Q'F\backslash F)\backslash[((Q\otimes Q)F\backslash F)\backslash(Q.F\backslash F)]\\&=(Q\otimes Q')F\backslash Q.F)\backslash[(Q\otimes Q)F\backslash Q.F]\\&\sim(Q\otimes Q')F\backslash(Q\otimes Q)F.\end{aligned}$$

Therefore, as wanted,

$$H^{Q\otimes Q'}\backslash H^{Q\otimes Q}\sim Q.H^{Q'}\backslash[H^{Q\otimes Q}\backslash H^{Q}]. \tag{3.140}$$

∎

We also had suggested above to define the mutual information $I_2(Q;Q')$ associated to a cocycle H by the formula $I_2(Q:Q')=H^{Q}\backslash Q'.H^{Q}$.
The restricted concavity condition on H is the existence of a natural cofibration $Q'.H^{Q}\rightarrowtail H^{Q}$.

Remark. This goes in the opposite direction to F: the more precise the theory S is, the bigger $H^{Q}(S)$ is, i.e. $S\leq S'$ implies $H^{Q}(S')\rightarrowtail H^{Q}(S)$.

We assume also that for all pair Q,Q' we have $H^{Q\otimes Q'}\sim H^{Q'\otimes Q}$.

Proposition. *under the above hypothesis and the assumption that $H^{Q\otimes Q}\sim H^{Q}$ and $H^{Q'\otimes Q'}\sim H^{Q'}$, we can consider H^{Q} and $H^{Q'}$ as subsets of $H^{Q\otimes Q'}$, and we have*

$$I_2(Q;Q')=I_2(Q';Q)=H^{Q}\cap H^{Q'}. \tag{3.141}$$

Proof. The Shannon formula (3.137) tells that $Q.H^{Q'}$ is $H^{Q\otimes Q'}\backslash H^{Q}$ and $Q'.H^{Q}$ is $H^{Q'\otimes Q}\backslash H^{Q'}$, then

$$I_2(Q;Q') = H^{Q}\backslash[H^{Q\otimes Q'}\backslash H^{Q'}] \sim H^{Q}\cap H^{Q'}. \tag{3.142}$$

■

Remark. We cannot write the relation with the usual union, but, under the above hypotheses, there is a cofibration

$$j\vee j' : H^{Q}\vee H^{Q'} \rightarrowtail H^{Q\otimes Q'}, \tag{3.143}$$

giving rise to a quotient

$$I_2(Q;Q') \cong H^{Q}\times_{H^{Q\otimes Q'}} H^{Q'}. \tag{3.144}$$

Generalizing the suggestion of Carnap and Bar-Hillel, and a Shannon theorem in the case of probabilities, we propose, to tell that Q,Q' are *independent* (with respect to P) at the theory S, when $H^{Q}\cap H^{Q'}$ is empty (initial element of $\mathcal{M}$).

With I_2, we can continue and get a semantic version of the *synergy* quantity of three variables:

$$I_3(Q_1;Q_2;Q_3)(S) = I_2(Q_1;Q_2)(S)\backslash I_2(Q_1;Q_2)(S|Q_3). \tag{3.145}$$

However, there is no reason why it must be a true space, because in the Abelian case it can be a negative number; (see [BTBG19] for the relation with the Borromean rings).

Remark. This invites us to go to $Ho(\mathcal{M})$, where there exists a notion of relative objects: for a zigzag $A \twoheadleftarrow C \rightarrowtail B$, with a trivial fibration to the left, and a cofibration to the right, the deduced arrow $A \longrightarrow B$ in $Ho(\mathcal{M})$, can be considered as a kind of difference of spaces as in Jardine, Cocyle categories [Jar09], and Zhen Lin Low, Cocycles in categories of fibrant objects [Low15]. Before Quillen and Jardine this kind of homotopy construction was introduced by Gabriel and Zisman [GZ67], as a calculus of fraction, in the framework of simplicial objects, their book being the first systematic exposition of the simplicial theory.

With respect to the Shannon information, what is missing is an analog of the expectation of functions over the states of the random variables. In some sense, this is replaced by the properties of growing and concavity of the function ψ, or spaces F and H, which give a manner to compare the theories. The true semantic information is not the value attributed to each individual theory, it is the set of relations between these values, either numerical, either geometric, as expressed by functors over the simplicial space $gI^{\bullet}_{\star}$, or better, more practical, over the part of ot that is accessible to a functioning network $g\mathbb{X}$.

The example of the theory $\mathcal{L}_3^2$ of Carnap and Bar-Hillel

Let us try to describe the structure of Information, as we propose it, in the simple (static) example that was chosen for development by Carnap and Bar-Hillel in their report in 1952, [CBH52].

The authors considered a language $\mathcal{L}_n^\pi$ with n subjects $a, b, c, \ldots$ and π attributes of them $A, B, \ldots$, taking some possible values, respectively $\pi_A, \pi_B, \ldots$. In their developed example $n = 3$, $\pi = 2$ and every π_i equals 2. The subjects are human persons, the two attributes are the gender G, male M or female F, and the age A, old O or young Y.

The elementary, or ultimate, states, $e \in E$ of the associated Boolean algebra $\Omega = \Omega^E$ are given by choosing values of all the attributes for all the subjects. For instance, in the language $\mathcal{L}_3^2$, we have $4^3 = 64$ elementary states.

The propositions $P, Q, R, \ldots$ are the subsets of Ω, their number is 2^{64}. The theories $S, T, \ldots$, in this case, are also described by their initial assertion, that is the truth of a given proposition, obtained by conjunction, and also named $S, T, \ldots$.

With our conventions, for conditioning and information spaces or quantities, it appears practical to define the propositions by the disjunction of their elements $e_I = e_{i_1} \vee \ldots \vee e_{i_k}$ and the theories by the conjonction of the complementary sets $\neg e_i = S_i$, that is $S_I = (\neg e_{i_1}) \wedge \ldots \wedge (\neg e_{i_k})$. Experimentally [BBG21a] the theories exclude something, like P, i.e. contain $\neg P$, then with S_I we see that $P = e_I$ is excluded, as are all the e_{i_j} for $1 \leq j \leq k$. A proposition Q which is implied by P, corresponds to a subset which contains all the elementary propositions e_{i_j} for $1 \leq j \leq k$.

In what follows, the models of "spaces of information" that are envisaged are mainly groupoids, or sets, or topological spaces.

A zero cochain $F_P(S)$ gives a space for any theory excluding P, in a growing manner, in the sense that $S \leq S'$ (inclusion of sets) implies $F(S) \leq F(S')$. The coboundary $\delta F = H$, gives a space $H_P^Q(S)$ for any proposition Q such that $P \leq Q$, whose formula is

$$H_P^Q(S) = F_P(S \vee \neg Q) \backslash F_P(S). \tag{3.146}$$

By concavity, this function (space) is assumed to be decreasing with S, i.e. if $S \leq S'$,

$$H_P^Q(S) \hookleftarrow H_P^Q(S'). \tag{3.147}$$

And by monotonicity of F, it is also decreasing in Q, i.e. if $Q \leq Q'$,

$$H_P^Q(S) \hookleftarrow H_P^{Q'}(S'). \tag{3.148}$$

In particular, we can consider the smaller $F_P(S)$ that is $F_P(\perp)$, as it is contained in all the spaces $F_P(S)$, we choose to take it as the empty space (or initial object in $\mathcal{M}$), then

$$H_P^Q(\perp) = F_P(\neg Q). \tag{3.149}$$

As we saw in general for every one-cocycle, not necessarily a coboundary, we have for any pair Q, Q' larger than P,

$$H_P^{Q\wedge Q'}(S)\backslash H_P^{Q'}(S) \approx H_P^Q(S|Q') = H_P^Q(S\vee\neg Q'). \tag{3.150}$$

Therefore, in the boolean case, every value of H can be deduced from its value on the empty theory:

$$H_P^Q(\neg Q') \approx H_P^{Q\wedge Q'}(\perp)\backslash H_P^{Q'}(\perp). \tag{3.151}$$

We note simply $H_P^Q(\perp) = H_P^Q = F_P(\neg Q)$.
And they are the spaces to determine.

The localization at P (i.e. the fact to exclude P) consists in discarding the elements e_i belonging to P from the analysis. Therefore we begin by considering the complete situation, which corresponds to $P = \perp$.

In this case we note simply $H^Q = H_\perp^Q = F(\neg Q)$.

The concavity of F is expressed by the existence of embeddings (or more generally cofibrations) associated to each set of propositions R_0, R_1, R_2, R_3 such that $R_0 \leq R_1 \leq R_3$ and $R_0 \leq R_2 \leq R_3$:

$$F(R_3)\setminus F(R_1) \rightarrowtail F(R_2)\setminus F(R_0). \tag{3.152}$$

In particular, for any pair of proposition Q, Q', we have $\perp \leq \neg Q \leq \neg(Q\wedge Q')$ and $\perp \leq \neg Q' \leq \neg(Q\wedge Q')$, and $F(\perp) = H^\top = \emptyset$, then

$$j : H^{Q\wedge Q'}\setminus H^{Q'} \rightarrowtail H^Q, \tag{3.153}$$

and

$$j' : H^{Q\wedge Q'}\setminus H^{Q} \rightarrowtail H^{Q'}. \tag{3.154}$$

Then we introduced the hypothesis that the subtracted spaces of both situations give equivalent results, and defined the mutual information $I_2(Q;Q')$:

$$H^Q\setminus(j(H^{Q\wedge Q'}\setminus H^{Q'})) \approx I_2(Q;Q') \approx H^{Q'}\setminus(j'(H^{Q\wedge Q'}\setminus H^{Q})). \tag{3.155}$$

Importantly, to get a cofibration, the subtraction cannot be replaced by a collapse with marked point, but it can in general be a collapse without marked point.

Consequently, the main axioms for the brut semantic spaces H^Q are: (i) the existence of natural embeddings (or cofibrations) when $Q \leq Q'$:

$$H^{Q'} \rightarrowtail H^Q, \tag{3.156}$$

and (ii) the above formulas (3.153) and (3.154) defining the same space $I_2(Q;Q')$, as in (3.155), which can perhaps all be interpreted after intersection.

We left open the relation between $I_2(Q;Q')$ and $H^{Q\vee Q'}$, however the axioms (ii) imply that there exist natural embeddings

$$H^{Q\vee Q'} \rightarrowtail I_2(Q;Q'). \tag{3.157}$$

The idea, to obtain a coherent set of non-trivial information spaces, is to exploit the symmetries of the language, or other elements of structure, which give an action of a category on the language, and generate constraints of naturalness for the spaces.

There exists a Galois group G of the language, generated by the permutation of the n subjects, the permutations of the values of each attribute and the permutations of the attributes that have the same number of possible values.

To be more precise, we order and label the subjects, the attribute and the values, with triples xY_i. In our example, $x = a,b,c$, $Y = A,G$, $i = 1,2$, the group of subjects permutation is $\mathfrak{S}_3$, the transposition of values are $\sigma_A = (A_1A_2)$ and $\sigma_G = (G_1G_2)$, and the four exchanges of attributes are $\sigma = (A_1G_1)(A_2G_2)$, $\kappa = (A_1G_1A_2G_2)$, $\kappa^3 = \kappa^{-1} = (A_1G_2A_2G_1)$, and $\tau = (A_1G_2)(A_2G_1)$.
We have

$$\sigma_A \circ \sigma_G = \sigma_G \circ \sigma_A = (A_1A_2)(G_1G_2) = \kappa^2; \tag{3.158}$$

$$\sigma \circ \sigma_A = \sigma_G \circ \sigma = \kappa; \quad \sigma_A \circ \sigma = \sigma \circ \sigma_G = \kappa^{-1}; \tag{3.159}$$

$$\sigma_A \circ \sigma \circ \sigma_G = \tau; \quad \sigma_A \circ \tau \circ \sigma_G = \sigma \tag{3.160}$$

The group generated by $\sigma, \sigma_A, \sigma_G$ is of order 8; it is the dihedral group D_4 of all the isometries of the square with vertices $A_1G_1, A_1G_2, A_2G_2, A_2G_1$. The stabilizer of a vertex is a cyclic group C_2, of type σ or τ, the stabilizer of an edge is of type σ_A or σ_G, noted C_2^A or C_2^A.

Therefore, in the example $\mathcal{L}_3^2$, the group G is the product of $\mathfrak{S}_3$ with a dihedral group D_4.

In the presentation given by the present article, the language $\mathcal{L}$ is a sheaf over the category G, which plays the role of the fiber $\mathcal{F}$. We have only one layer U_0, but the duality of propositions and theories corresponds to the duality between questions and answers (i.e. theories) respectively.

The action of G on the set Ω is deduced from its action on the set E, which can be described as follows:

1) One orbit of four elements, where a,b,c have the same gender and age. The stabilizer of each element is $\mathfrak{S}_3 \times C_2$, or order 12.

2) One orbit of 24 elements made by a pair of equal subjects and one that differs from them by one attribute only. The stabilizer being the $\mathfrak{S}_2$ of the pair of subjects.

3) One orbit of 12 elements made by a pair of equal subjects and one that differs from them by the two attributes. The stabilizer being the product $\mathfrak{S}_2 \times C_2$, where C_2 stabilizes the characteristic of the pair, which is the same as stabilizing the character of the exotic subject.

4) One last orbit of 24 elements, where the three subjects are different, then two of them differ by one attribute and differ from the last one by the two attributes. The stabilizer is the stabilizer C_2' of the missing pair of values of the attributes.

The action of G on the set E corresponds to the conjugation of the inertia subgroups.

Remark. All that looks like a Galois theory; however there exist subgroups of G, even normal subgroups, that cannot happen as stabilizers in the language, without adding terms or concepts. For instance, the cyclic group $\mathfrak{A}_3 \subset \mathfrak{S}_3$; if it stabilizes a proposition P, this means that the subjects appear in complete orbits of $\mathfrak{A}_3$, but these orbits are orbits of $\mathfrak{S}_3$ as well, then the stabilizer contains $\mathfrak{S}_3$. The notion of cyclic ordering is missing.

The collection of all the ultimate states of a given type defines a proposition, noted T, describing I, II, III, IV. This proposition has for stabilizer the group G itself. Its space of information must have a form attached to G, but it also must take into account the structure of its elements.

Ansatz 1. *The information space of type T corresponds to the natural groupoid of type T*

Remark that each type corresponds to a well formed sentence in natural languages: type I is translated by "all the subjects have the same attributes"; type II by "all the subjects have the same attributes except one which differs by only one aspect"; type III "one subject is opposite to all the others"; type IV "all the subjects are distinguished by at least one attribute".
The union of the types II and III is described by the sentence "all the subjects have the same attributes except one".
The information space of $(II) \vee (III)$ is (naturally) a groupoid with 12 objects and fundamental group $\mathfrak{S}_2$. A good exercise is to determine the information spaces of all the unions of the four orbits. It should convince the reader that something interesting happens here, even if the whole tentative here evidently needs to be better formalized.

Remark that other propositions have non-trivial inertia, and evidently support interesting semantic information. The most important for describing the system are the *numerical statements*, for instance "there exist two female subjects in the population". Its inertia is $\mathfrak{S}_3 \times C_2^A$.

By definition, a *simple* proposition is given by the form aA, telling that one given subject has one given value for one given attribute. There exist twelve such propositions, they are permuted by the group

G. The simple propositions form an orbit of the group G, of type III above.

Amazingly, the set of the twelve simples is selfdual under the negation:

$$\neg(aA) = a\overline{A}, \tag{3.161}$$

where $\overline{A}$ denotes the opposite value.

Ansatz 2. *Each simple corresponds to a groupoid with one object, and four arrows, that form a Klein sub-group of G which fixes the subject a and fixes the attribute A corresponding to C_2, generated by the transposition σ_A, also preserving $\overline{A}$.*

Another ingredient, introduced by Carnap and Bar-Hillel, is the *mutual independency* of the 12 *simple* propositions.

According to the definition of the spaces $I_2(Q,Q')$, this implies:

Ansatz 3. *The spaces of the simples are disjoint; the maximal information spaces, associated to full populations e, are unions of them, after some gluing.*

It is natural to expect that for each individual population $e \in X$, the information space H^e is a kind of marked groupoid H^T_e, that is a groupoid with a singularized object. A good manner to mark the point e in H^e is to glue to the space H^T of its type a space H^P, where P is the proposition which characterizes e among the elements of the orbit T. The groupoid of this space H^P can contains several objects.

All kinds of gluing that we had to consider are realized by identifying two spaces H_1, H_2 with marked points along a subspace K (representing a mutual information or the space of the "or"), as asked by the axiom (ii) above.
Therefore in general, the subspace has strictly less marked points than any of the spaces that are glued.
When we mention *cylinders* in this context, this means that one of the spaces, say H_1 is a cylinder with basis K, and we say that H_1 is grafted on the other space H_2.

Ansatz 4. *The information space of the ultimate element e is obtained by gluing a cylinder to the space of its type, based on a subspace associated to it, and containing as many objects as we need simple pieces*

For type I, one object is added; for type II and III, two objects are added and for type IV, three objects.

Illustration. Associate to each e a trefoil knot, presented as a braid with three colored strands, corresponding to its three simple constituents.
Each subject corresponds to a strand, each pair of values A, G of the attributes to a color, red, blue, green and black for the vertices ofthe square, red and green and blue and black being in diagonal.

Any proposition is a union of elementary ones, then to go farther, we have to delete pieces of the maximal spaces H^e, for obtaining its information spaces.

The existence of a full coherent set of spaces is non-trivial and is described in detail in the forthcoming preprint, *A search of semantic spaces* [BB22].

Then to describe the information of the more general propositions, we have to combine the forms given by the groups and groupoids, as for H^T and H^e, with a combinatorial counting of information, deduced from the content, as in Carnap and Bar-Hillel.

A suggestion is to represent the combinatorial aspect by a dimension: all propositions are ranged by their numerical content, for instance e has $c(e) = 63$, $\neg e$ has $c = 1$, and aA has $c = 58$. We represent the groups and groupoids by $CW-$ complexes of dimension 2 or ∞, associated to a presentation by generators and relations of their fundamental group, possibly marked by several base points. The spaces of information H^Q are obtained by thickening the complexes, by taking the product with a simplex or a ball of the dimension corresponding to Q. However, note that any manner to code this dimension by a number, for instance, connected components, would work as well.

For some propositions, we cannot expect a form of information in addition of the dimension. This concerns propositions that are complex and not used in natural languages; example: "in this population, there is two old mans, or there is a young woman, or there exist a woman that has the same age of a man". This is pure logical calculus, not really semantic.

The general construction shows that the number of non-trivial semantic spaces is far from 2^{64}, it is of the order of 64^{α}, with α between 3 or 4.

Then, on this simple example we see that "spaces" of semantic information are more interesting and justified than numerical estimations, but also that this concerns only few propositions, the ones which seem too have more sense. Then the structure of spaces has to be completed by calculus and combinatorics for most of the 2^{64} sentences. This touches the sensitive departure point from the *admissible* sentences, more relevant to Shannon theory, and the *significant* sentences, more relevant for a future semantic theory, that we hope to find in the above direction of homotopy invariants of spaces of theories and questions.

4 Unfoldings and memories, LSTMs and GRUs

This chapter presents evidences that some architectures of $DNNs$, which are known to be efficient in syntactic and semantic tasks, rely on internal invariance supported by some groupoids of braids, which also appear in enunciative linguistic, in relation with cognition and representation of notions in natural languages.

4.1 RNN lattices, LSTM cells

Artificial networks for analyzing or translating successions of words, or any timely ordered set of data, have a structure in lattice, which generalizes the chain: the input layers are arranged in a corner: horizontally $x_{1,0}$, $x_{2,0}$, ..., named data, vertically $h_{0,1}$, $h_{0,2}$, ..., named hidden memories.

Generically, there is a layer $x_{i,j}$ for each $i = 1, 2, ..., N$, $j = 0, 1, 2, ..., M$, and a layer $h_{i,j}$ for each $i = 1, 2, ..., N$, $j = 0, 1, 2, ..., M$. The information of $x_{i,j-1}$ and $h_{i-1,j}$ are joined in a layer $A_{i,j}$, which sends information to $x_{i,j}$ and $h_{i,j}$.
Then in our representation, the category $C_{\mathbf{X}}$ has one arrow from $x_{i,j}$ to $A_{i,j}$, from $h_{i,j}$ to $A_{i,j}$, from $x_{i,j-1}$ to $A_{i,j}$ and from $h_{i-1,j}$ to $A_{i,j}$, and it is all (see figure 4.1). If we want, we could add the layers $A^{\star}_{i,j}$, but there is no necessity.

The output is generally a up-right corner horizontally $y_1 = x_{1,M}$, $y_2 = x_{2,M}$, ..., named the result (a classification or a translation), and vertically $h_{N,1}$, $h_{N,2}$, ..., (which could be named future memories).

However, the inputs and outputs can have the shape of a more complex curves, transverse to vertical and horizontal propagation. Things are organized as in a two dimensional Lorentz space, where a space coordinate is $x_{i,j-1} - h_{i-1,j}$ and a time coordinate $x_{i,j-1} + h_{i-1,j}$. Input and output correspond to spatial sections, related by causal propagation.

Remark. In many applications, several lattices are used together, for instance a sentence or a book can be read backward after translation, giving reverse propagation, without trouble. We will discuss these aspects with the modularity.

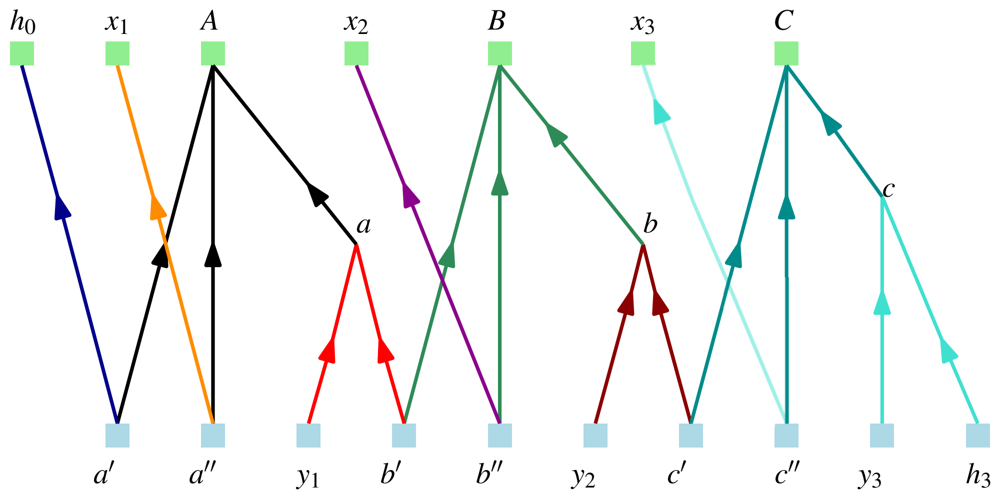


Figure 4.1: Categorical representation of a RNN

Most $RNNs$ have a dynamic of the type a non-linearity applied to a linear summation:
we denote the vectorial states of the layers by greek letters ξ for layers x and η for layers h, like $\xi^a_{i,j}$ and $\eta^b_{k,l}$; the lower indices denote the coordinates of the layer and the upper indices denote the neuron, that is the real value of the state. In most applications, the basis of neurons plays an important role.
In the layer $A_{i,j}$ the vector of state is made by the pairs $(\xi^a_{i,j-1}, \eta^b_{i-1,j}); a \in x_{i,j-1}, b \in h_{i-1,j}$.
The dynamic X^w has the following form:

$$\xi^a_{i,j} = f^a_x\left(\sum_{a'} w^a_{a';x,i,j}\xi^{a'}_{i,j-1} + \sum_{b'} u^a_{b';x,i,j}\eta^{b'}_{i-1,j} + \beta^a_{x,i,j}\right); \tag{4.1}$$

$$\eta^b_{i,j} = f^b_h\left(\sum_{a'} w^b_{a';h,i,j}\xi^{a'}_{i,j-1} + \sum_{b'} u^b_{b';h,i,j}\eta^{b'}_{i-1,j} + \beta^b_{x,i,j}\right). \tag{4.2}$$

The functions f are sigmoids or of the type $\tanh(Cx)$, the real numbers β are named *bias*, and the numbers w and u are the weights.
In practice, everything here is important, the system being very sensitive, however theoretically, only the overall form matters, thus for instance we can incorporate the bias in the weights, just by adding a formal neuron in x or h, with fixed value 1. The weights are summarized by the matrices $W_{x,i,j}$, $U_{x,i,j}$, $W_{h,i,j}$, $U_{h,i,j}$.
All these weights are supposed to be learned by backpropagation, or analog more general reinforcement.

Experiments during the eighties and nineties showed the strongness of the RNNs but also some weaknesses, in particular for learning or memorizing long sequences. Then Hochreiter and Schmidhuber, in a remarkable paper in Neural Computation [HS97], introduced a modification of the simple RNN, named the Long Short Term Memory, or $LSTM$, which overcame all the difficulties so efficiently that more than thirty years after it continues to be the standard.

The idea is to duplicate the layers h by introducing parallel layers c, playing the role of longer time memory states, and just called cell states, by opposition to hidden states for h.

In what follows we present the cell which replaces $A_{i,j}$ without insisting on the lattice aspect, which is unchanged for many applications.

The sub-network which replaces the simple crux $A = A_{i,j}$ is composed of five tanks A, F, I, H', V, plus the inputs $C_{t-1}, H_{t-1}, X_{t-1}$, and has nine tips $c'_{t-1}, h'_{t-1}, x'_t, f, i, o, \widetilde{h}, v_i, v_f$ plus the three outputs c_t, h_t, y_t. However, y_t being a function of h_t only, it is forgotten in the analysis below.
In A, the two layers h' and x' (where we forget the indices $t-1$ and t respectively) join to give by formulas like (4.2) the four states of $i, f, o, \widetilde{h}$ respectively called input gate, forget gate, output gate, combine gate, the first three are sigmoidal, the fourth one is of type tanh, indicating a function of states separations. The weights in these operations are the only parameters to adapt, they form matrices W_i, U_i, W_f, U_f, W_o, U_o and W_h, U_h; which makes four times more than for a RNN (because the output $\xi_{i,j}$ is not taken in account).
Then the states in v_f and v_i are respectively given by combining c' with f and $\widetilde{h}$ with i, in the simplest bilinear way:

$$\xi_v^a = \gamma^a \varphi^a; a \in v; \tag{4.3}$$

where γ denotes the states of c' or $\widetilde{h}$, and φ the states of f or i respectively.
Note that the above formulae have a sense if and only of the dimensions of c and f and v_f are equal and the dimension of $\widetilde{h}$ and i and v_i are equal. This is an important restriction.
At the level of vectors this diagonal product is name the *Hadamard product* and is written

$$\xi_v = \gamma \odot \varphi. \tag{4.4}$$

It is free of parameters. Only the dimension is free for a choice.
Then, v_i and v_f are joined by a Hadamard sum, adding term by term, to give the new cell state

$$\xi_c = \xi_{v_f} \oplus \xi_{v_i}; \tag{4.5}$$

which implies that v_i and v_f have the same dimension.
And finally, a new Hadamard product gives the new hidden state:

$$\eta_h = \xi_o \odot \tanh \xi_c. \tag{4.6}$$

We get an additional degree of freedom with the normalization factor C in $\tanh Cx$ but this is all. However this implies that c and o and h have the same dimension.

Therefore the $LSTM$ has a discrete invariant, which is the dimension of the layers and is named its *multiplicity* m.
Only the layers x can have other dimensions; in what follows, we denote n this dimension (see figure 4.2).

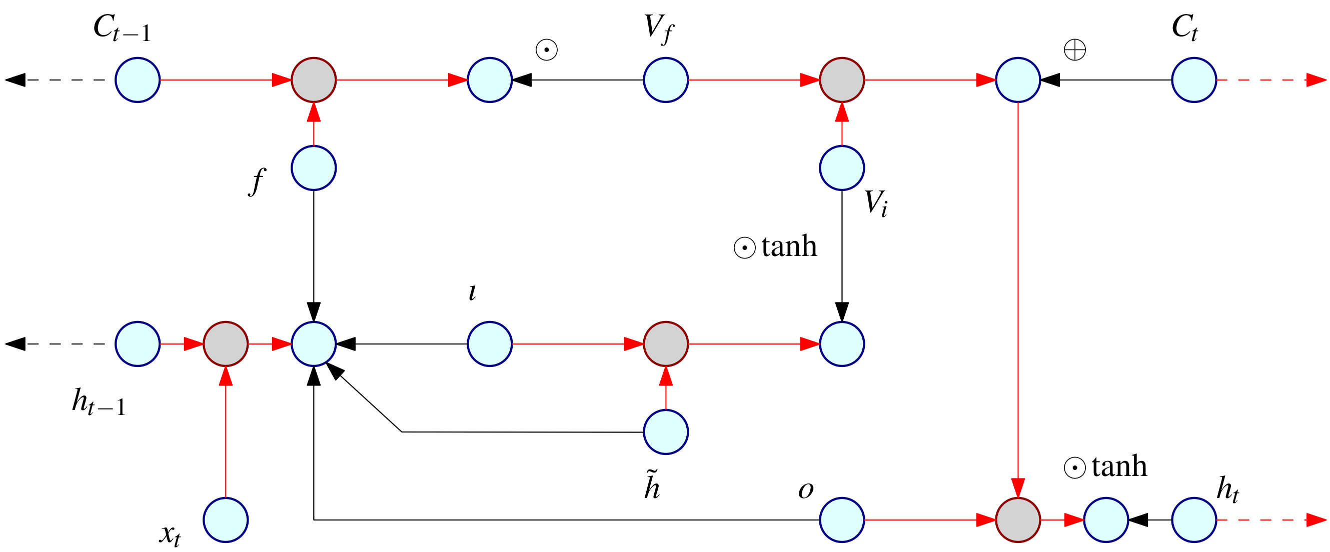


Figure 4.2: Grothendieck site representing a LSTM cell

Symbolically, the dynamics can be summarized by the two formulas:

$$c_t = c_{t-1} \odot \sigma_f(x_t, h_{t-1}) \oplus \sigma_i(x_t, h_{t-1}) \odot \tau_h(x_t, h_{t-1}) \tag{4.7}$$

$$h_t = \sigma_o(x_t, h_{t-1}) \odot \tanh c_t, \tag{4.8}$$

where σ_k (*resp.* τ_k) denotes the application of σ (*resp.* tanh) to a linear or affine form.
In what follows, x_t is replaced by x' and h_{t-1}, c_{t-1} by h', c', like their tips.

Due to the non-linearities σ and tanh, there are several regimes of functioning, according to the fact that some of the variables give or not a saturation; this can generate almost linear transformations or the opposite, a discrete-valued transformation. For instance, ± 1 when tanh is applied, or $\in \{0, 1\}$ if σ is applied. Here appears the fundamental aspect of discretization in the functioning of $DNNs$.

In the linear regime, the new state c appears as a polynomial of degree 2 in the vectors x, h' and degree 1 in c', and h appears as a polynomial of degree 3 in x', h'.
Introducing the linear (or affine with bias) forms $\alpha_f, \alpha_i, \alpha_o, \alpha_h$, before application of σ or tanh, we have

$$h_t = \alpha_o \odot (c' \odot \alpha_f \oplus \alpha_i \odot \alpha_h). \tag{4.9}$$

The dominant term in x', h' is decomposable: $\alpha_o \odot \alpha_i \odot \alpha_h$; the term of degree 2 in x', h' is $\alpha_o \odot c' \odot \alpha_f$, and there is no linear term, because we forgot the bias. When separating x' from h', we obtain all possible degrees ≤ 3.
However, experiments with alternative memory cells, named GRU and their simplifications, have shown that the degree in x' is apparently less important then the degree in h'. All trials with degree < 3 in h' gave a dramatic loss of performance, but this was not the case for x', where degree 1 appears to be sufficient.

The number of parameters to tune is $4m^2+4mn$ or $4m^2+dmn$, with $1 \le d \le 4$ is for the dependencies in x in the four operations $\alpha_f, \alpha_i, \alpha_o, \alpha_h$. At least $d = 1$ for α_h or for α_f seems to be necessary from the study of MGU.

4.2 GRU, MGU

Several attempts were made for diminishing the quantity of parameters to adapt in $LSTM$ without diminishing the performance. The most popular solution is known as Gated Recurrent Unit, or GRU (see [CvMBB14] and [CGCB14] from Bengio's group). Then this cell has been simplified into several kinds of Minimal Gated Units, MGU ([ZWZZ16] or [HS17]).

The idea is to replace several gated layers by one, at the cost of a more complex architecture's topology.

In the standard GRU, the pair h_t, c_t is replaced by h_t alone, as in the original RNN; there exists two input layers X_t, H_{t-1}, the number of joins, our tanks, is six: R, F, I, V, W, H', the number of tips is six, $z, r, v_{1-z}, v_r, v_x, v_h$ and one output h_t.
The dynamic begins with two non-linear linear transform, of type $\sigma\sum$, like (4.2) in R, giving z and r from x' and h'; then in I, there is a Hadamard product $v_z = h' \odot (1-z)$, where $1-z$ designates the Hadamard difference between the saturation and the values of the states of z. Moreover, in F, there is another Hadamard product $v_r = h' \odot r$. A $\tanh\sum$, like (4.2) with $f = \tanh$, joins x' with v_r in W to give v_x, which joins z in H' to give v_h by a third Hadamard product. Finally, v_h and v_{1-z} are joined together by a Hadamard sum in V, giving $h = v_z \oplus v_h$.

Symbolically, with the same conventions used for $LSTM$, the dynamic can be summarized by the following formula

$$h_t = (1-\sigma_z(x_t, h_{t-1})) \odot h_{t-1} \oplus \sigma_z(x_t, h_{t-1}) \odot \tanh(W_x(x_t) + U_x(\sigma_r(x_t, h_{t-1}) \odot h_{t-1})). \tag{4.10}$$

In a GRU as in a $LSTM$ we have three Hadamard products and one Hadamard sum, plus three non-linear-linear transforms NLL (one with tanh); $LSTM$ had four NLL transforms (two with tanh), but the complexity of GRU stays in the succession of two NLL with adaptable parameters.
Remark that $LSTM$ also contains a succession of non-linearities, tanh being applied to c_t, which is a sum of product on non-linear terms of type σ or tanh.

In the linear (or affine) regime, the GRU gives

$$h_t = [(1-\alpha_z) \odot h_{t-1}] \oplus [\alpha_z \odot [Wx_t + U(\alpha_r \odot h_{t-1})]]. \tag{4.11}$$

For the same reason than $LSTM$ a GRU has a multiplicity m, and a dimension n of data input. The parameters to be adapted are the matrices W_z, U_z, W_r, U_r and W_x, U_x in W. This gives $3m^2+3mn$ real numbers to adapt, in place of $4m^2+4mn$ for a complete $LSTM$.

The simplification which was proposed by Zhou et al. in [ZWZZ16] for MGU consists in taking $\sigma_z = \sigma_r$, thus reducing the parameters to $2m^2+2mn$. This unique vector is denoted σ_f, assimilated to the forget gate f of $LSTM$.
It seems that the performance of MGU was as good as the ones of GRU, which are almost as good as $LSTM$ for many tasks.

Heck and Salem [HS17] suggested further radical simplifications, some of them being as good as MGU. $MGU1$ consists in suppressing the dependency of the unique σ_f in x', and $MGU2$ in suppressing also the bias β_f. An $MGU3$ removed x' and h', just keeping a bias, but it showed poor learning and accuracy in the tests.

The experimental results proved that $MGU2$ is excellent in all tests, even better than GRU.
Note that both $MGU2$ and $MGU1$ continue to be of degree 3 in h'. This reinforces the impression that this degree is an important invariant of the memory cells. But these results indicate that the degree in x' is not so important.

Consequently we may assume

$$h_t = (1-\sigma_z(h_{t-1}))\odot h_{t-1} \oplus \sigma_z(h_{t-1})\odot \tanh(W_x(x_t)+U_x(\sigma_z(h_{t-1})\odot h_{t-1})). \tag{4.12}$$

And in the linear regime

$$h_t = [(1-\alpha_z)\odot h'] \oplus [\alpha_z \odot [Wx_t + U(\alpha_z \odot h')]]. \tag{4.13}$$

Only two vectors of linear (or affine) forms intervene, $\alpha_z^a(h'); a = 1,\dots,m$ and h' itself, i.e. $\eta^a(h'); a = 1,\dots,m$.
The parameters to adapt are U_z, giving α_z, and $U_x = U$, $W_x = W$, giving the polynomial of degree two in parenthesis, i.e. the state of the layer called v_h.

The number of free parameters in $MGU2$ is $2m^2+mn$, twice less than the most economical $LSTM$.

The graph Γ of a GRU or a MGU has five independent loops, a fundamental group free of rank five; it is non-planar. The categorical representation of a $LSTM$ has only three independent loops, and is planar (see figure 4.2).

4.3 Universal structure hypothesis

A possible form of dynamic covering the above examples is a vector of dimension m of non-linear functions of several vectors σ_{α^a}, σ_{β^b}, ..., that are σ of th functions of linear (or perhaps affine) forms of the variables ξ^a, η^b, for a, b, c varying from 1 to m. More precisely

$$\eta_t^a = \sum_{b,c,d} t_b^a \sigma_{\alpha^b} \tanh \left[\sum_{c,d} u_{c,d}^a \sigma_{\beta^c} \sigma_{\gamma^d} + \sum_c v_c^a \sigma_{\beta^c} + \sum_d w_d^a \sigma_{\gamma^d} + \sigma_{\delta^a} \right]. \tag{4.14}$$

Remark: we have written $\sigma_\alpha, \sigma_\beta, \ldots$ for the application to a linear form of a sigmoid or a tanh indifferently; but for a more precise discussion of the examples, we must distinguish and write $\tau_\alpha, \tau_\beta, \ldots$ when tanh is applied. However, sometimes in the following lines, we will use τ when we are sure that a tanh is preferable to a σ.

The tensor $u_{c,d}^a$ would introduce m^3 parameters, leading to great computational difficulties. A natural manner to limit the degrees of freedom at Km^2, inspired by $LSTM$ and GRU, is to use the Hadamard product, for instance $\sigma_{\beta^a} \sigma_{\gamma^a}$.
A second simplification, justified by the success of MGU consists to impose $\alpha^a = \gamma^a$.
A third one, justified by the success of $MGU2$ is to limit the degree in x' to 1. This can be done by reserving the dependency on x' to the forms β and δ.
All that gives

$$\eta_t^a = \sigma_{\alpha^a}(\eta) \tanh \left[\sigma_{\alpha^a}(\eta) \sigma_{\beta^a}(\eta, \xi) + \sigma_{\beta^a}(\eta, \xi) + \sigma_{\delta^a}(\xi) \right]. \tag{4.15}$$

This contains $2m^2 + 2mn$ free parameters to be adapted.

Remark. Here we have neglected the addition of the alternative term in the dynamic which is $(1 - \sigma_{\alpha^a})\eta^a$ in GRU and MGU, but this term is probably very important, therefore, we must keep in mind that it can be added in the applications. At the end it will reappear in the formulas we suggest below.

For $MGU1, 2$, the term of higher degree has no dependency in x', then we can simplify further in

$$\eta_t^a = \sigma_{\alpha^a}(\eta) \tanh \left[\sigma_{\alpha^a}(\eta) \sigma_{\beta^a}(\eta) + \sigma_{y^a}(\xi) \sigma_{\beta^a}(\eta) + \tau_{\delta^a}(\xi) \right]. \tag{4.16}$$

Moreover, as $MGU2$ is apparently better than $MGU1$ in the tested applications, the forms α^a can be taken linear, not affine.

It looks like a simplified $LSTM$, if we define for the state of c_t the following vector:

$$\gamma_t^a = \sigma_{\alpha^a}(\eta) \sigma_{\beta^a}(\eta) + \sigma_{y^a}(\xi) \sigma_{\beta^a}(\eta) + \tau_{\delta^a}(\xi), \tag{4.17}$$

and impose the recurrence $y^a(\xi) = \gamma_{t-1}^a$.
This gives a kind of minimal $LSTM$, so-called $MLSTM$,

$$\gamma_t^a = \sigma_{\alpha^a}(\eta) \sigma_{\beta^a}(\eta) + \gamma_{t-1}^a \sigma_{\beta^a}(\eta) + \tau_{\delta^a}(\xi), \tag{4.18}$$

$$\eta_t^a = \sigma_{\alpha^a}(\eta)\tanh[\gamma_t^a]. \tag{4.19}$$

Or with the forgotten alternative term,

$$\eta_t^a = \sigma_{\alpha^a}(\eta)\tanh[\gamma_t^a] + (1-\sigma_{\alpha^a}(\eta))\eta^a. \tag{4.20}$$

Now we suggest to look at these formulas from the point of view of the deformation of singularities having polynomial universal models, and trying to keep the main properties of the above dynamics:

1) on a generic straight line in the input space h', and in any direction of the output space h, we have *every possible shape of a 1D polynomial function of degree* 3, when modulating by the functions of x';

2) the presence of non-linearity σ applied to forms in h' and th applied to forms in x' allow discretized regimes for the full application, but also a regime where the dynamic is close to a simple polynomial model.

In the above formulas the last application of th renders possible the degeneration to degree 1 in h' and x', we suggest to forbid that, and to focus on the coefficients of the polynomial. In fact the truncation of the linear forms by σ or th is sufficient to warranty the saturation of the polynomial map.
From this point of view the terms of degree 2 are in general not essential, being absorbed by a Viete transformation. Also the term of degree zero, does not change the shape, only the values; but this can be non-negligible.
In the simplest form this gives

$$\eta_t^a = \sigma_{\alpha^a}(\eta)^3 + u^a(\xi)\sigma_{\alpha^a}(\eta) + v^a(\xi); \tag{4.21}$$

where u and v are th applied to a linear form of ξ, and σ_α is a σ applied to a linear form in η. This gives only $m^2 + 2mn$ free parameters, thus one order less than $MGU2$ in m.

However, we cannot neglect the forgotten alternative $(1-z)h'$ of GRU, or more generally the possible function in the transfer of a term of degree two, even if structurally, from the point of view of the deformation of shapes, it seems not necessary, thus the following form could be preferable:

$$\eta_t^a = \sigma_{\alpha^a}(\eta)^3 + (1-\sigma_{\alpha^a}(\eta))\eta^a + u^a(\xi)\sigma_{\alpha^a}(\eta) + v^a; \tag{4.22}$$

or more generally, with $2m^2 + 2mn$ free parameters:

$$\eta_t^a = \sigma_{\alpha^a}(\eta)^3 + \sigma_{\alpha^a}(\eta)[\sigma_{\beta^a}(\eta) + u^a(\xi)] + v^a(\xi); \tag{4.23}$$

where β is a second linear map in η.

Description of an architecture for this dynamic : it has two input layers H_{t-1}, X_t, three sources or tanks A, B, C, and seven internal layers that give six tips, α,β, v_β, u, v, $v_{\alpha\beta}$, $v_{\alpha\alpha\alpha}$, and one output layer

h_t. First h_{t-1} gives σ_α and σ_β, and x_t gives u and v; then σ_β joins u in A to give $v_\beta = \sigma_\beta \oplus u$, then σ_α joins v_β in B to give $v_{\alpha\beta} = \sigma_\alpha \odot v_\beta$. In parallel, σ_α is transformed along an ordinary arrow in $v_{\alpha\alpha\alpha} = \sigma_\alpha^{\odot 3}$. And finally, in C, the sum of v, $v_{\alpha\alpha\alpha}$ and v_β produces the only output h_t.

The simplified network is for $\beta = 0$. It has also three tanks, A, B and C, but only five tips, α, u, v, v_α, $v_{\alpha\alpha\alpha}$. The schema is the same, without the creation of β, and v_β (*resp.* $v_{\alpha\beta}$) replaced by v_α (*resp.* $v_{\alpha\alpha}$).

Remark. In the models with tanh like (4.20) the sign of the terms of effective degree three can be minus or plus; in the model (4.23) it is always plus, however this can be compensated by the change of sign of the efferent weights in the next transformation.

Equation (4.15) could induce the belief that 0 goes to 0, but in general this is not the case, because the function σ contrarily to tanh has only strictly positive values. For instance the standard $\sigma(z) = 1/1 + \exp(-z)$ gives $\sigma(0) = 1/2$.

However, the point 0 plays apparently an important role, even if it is not preserved: 1) in $MGU2$ the absence of bias in α^a confirms this point; 2) the functions σ and th are almost linear in the vicinity of 0 and only here. Therefore, let us define the space H of the activities of the memory vectors h_{t-1} and h_t, of real dimension m; it is pointed by 0, and the neighborhood of this point is a region of special interest.

We also introduce the line U of coordinate u and the plane $\Lambda = U \times \mathbb{R}$ of coordinates u, v, where 0 and its neighborhood is also crucial. The input from new data x_t is sent to Λ, by the two maps $u(\xi)$ and $v(\xi)$. By definition this constitutes an *unfolding* of the degree three map in $\sigma_\alpha(\eta)$.

A more complex model of the same spirit is

$$\eta_t^a = \sigma_{\alpha^a}(\eta)^3 \pm \sigma_{\alpha^a}(\eta)[\sigma_{\beta^a}(\eta)^2 + u^a(\xi)] + v^a(\xi)\sigma_{\beta^a}(\eta) + w^a(\xi)[\sigma_{\alpha^a}(\eta)^2 + \sigma_{\beta^a}(\eta)^2] + z^a(\xi); \quad (4.24)$$

it has $2m^2 + 4mn$ free parameters. The expression of x_t is much richer and we will see below that it shares many good properties with the model (4.21), in particular stability and universality. The corresponding space U has dimension 3 and the corresponding space Λ has dimension 4.

4.4 Memories and braids

In every DNN, the dynamic from one or several layers to a deeper one must have a sort of stability, to be independent of most of the details in the inputs, but it must also be plastic, and sensitive to the important details in the data, then not too stable, able to shift from a state to another one, for constructing a kind of discrete signification. These two aspects are complementary. They were extensively discussed a long time before the apparition of DNNs in the theory of dynamical systems. The framework was different because most concepts in this theory were asymptotic, pertinent when the time tends to infinity, and here in deep learning, to the contrary, most concepts are transient: one shot transformations for feed forward, and gradient descent or open exploration for learning; however, with respect to the shape of individual transformation, or with respect to the parameters of deformation, the two domains encounter similar problems, and probably answer in similar manners.

Structural stability is the property to preserve the shape after small variation of the parameters. In the case of individual map between layers, this means that little change in the input has little effect on the output. In the case of a family of maps, taking in account a large set of different inputs, this means that varying a little the weights, we get little change in the global functioning and the discrimination between data. The second level is deeper, because it allows to understand what are the regions of the manifolds of input data, where the individual dynamics are stable in the first sense, and what happens when individual dynamics changes abruptly, how are made the transitions and what are the properties of the inputs at the boarders. A third level of structural stability concerns the weights, selected by learning: in the space of weights it appears regions where the global functioning in the sense of family is stable, and regions of transitions where the global functioning changes; this happens when the tasks of the network change, for instance detect a cat versus a dog. This last notion of stability depends on the architecture and on the forms of dynamical maps that are imposed.

With $LSTM$, GRU and their simplified versions like MGU, $MGU2$, we have concrete examples of these notions of structural stability.

The transformation is X^w from (h_{t-1}, x_t) to h_t. The weights w are made by the coefficients of the linear forms, $\alpha^a(\eta), \beta^a(\eta), u^a(\xi)$, $v^a(\xi)$, but the structure depends on the fixed architecture and the non-linearities, of two types, the tensor products and sums, and the applied sigmoids and $tanh$.

For simplicity we assume a response of the cell of the form (4.21), but the discussion is not very different with the other cell families (4.23), (4.16) or (4.20).

We have a linear endomorphism α of coordinates $\alpha^a; a \in h$ of $\mathbb{R}^m = H$; when we apply to it the sigmoid function coordinate by coordinate, we obtain a map ϕ from H to a compact domain in H. The invariance of the multiplicity m of the memory cell suggests the hypothesis (to be verified experimentally) that ϕ is a diffeomorphism from H to its image. However, as we will see just below, other reasons like redundancy suggests the opposite, therefore we left open this hypothesis, with a preference for diffeomorphism, for mathematical or structural reasons. Probably, depending on the application, there exists a range of dimensions m which performs the task, such that ϕ is invertible.

We also have the two mappings $u^a(\xi); a \in h$ and $v^a(\xi); a \in h$ from the space $X = \mathbb{R}^n$ of states x_t, to $\mathbb{R}^m$. This gives a complete description of the set of weights $W_{h;h',x'}$.

The formula (4.21) defines the map X^w from $H \times X$ to H.

We also consider the restriction X^w_ξ at a fixed state ξ of x_t.

Theorem 4.1. *The map X^w is not structurally stable on H or $H \times X$, but each coordinate η^a_t, seen as* *function on a generic line of the input h_{t-1} and a generic line of the input x_t, or as a function on H or $H \times X$, is stable (at least in the bounded regions where the discretization does not apply).*

These coordinates represent the activities of individual neurons, then we get structural stability at the level of the neurons and not at the level of the layers.

As we justify in the following lines, this theorem follows from the results of the universal unfolding theory of smooth mappings, developed by Whitney, Thom, Malgrange and Mather (see [GWDPL76] and [Mar82]).
The main point here (our hypothesis) is the observation that, for each neuron in the h_t layer, the cubic degeneracy z^3 can appear, together with its deformation by the function u.
For the deformation of singularities of functions, and their unfolding, see [Arn73] and [AGZV12a].

The universal unfolding of the singularity z^3 is given by a polynomial

$$P_u(z) = z^3 + uz, \tag{4.25}$$

This means that for every smooth real function F, from a neighbor of a point 0 in $\mathbb{R}^{1+M}$, such that

$$F(z, 0, ..., 0) = z^3, \tag{4.26}$$

there exist a smooth map $u(Y)$ and a smooth family of maps $\zeta(z, Y)$ such that

$$F(z, Y) = \zeta(z, Y)^3 + u(Y)\zeta(z, Y) \tag{4.27}$$

Equivalently, the smooth map

$$(z, u) \mapsto (P_u(z), u), \tag{4.28}$$

in the neighbor of $(0, 0)$ is stable: every map sufficiently near to it can be transformed to it by a pair of diffeomorphisms of the source and the goal. This result on maps from the plane to the plane, is the starting point of the whole theory, found by Whitney: the stability of the gathered surface over the plane v, u.
The stability is not true for the product

$$(z, u, w, v) \mapsto (P_u(z), u, P_v(w), v) \tag{4.29}$$

The infinitesimal criterion of Mather is not satisfied (see [GWDPL76], [Mar82]).

There also exists a notion of universal unfolding for maps from a domain of $\mathbb{R}^n$ to $\mathbb{R}^p$ in the neighborhood of a point 0, however in most cases, there exists no universal unfolding, at the opposite of the case of functions, when $p = 1$.
Here $n = p = m$, the transformation from h_{t-1} to h_t is an unfolding, dependent of $\xi \in x_t$, but it does not

admit a universal deformation. It has an infinite codimension in the space of germs of maps.
Also for mappings, universality of and unfolding and its stability as a map are equivalent (another theorem from Mather).

Our non-linear model from equation (4.21) with u free being equivalent to the polynomial model by diffeomorphism, we can apply to it the above results. This establishes theorem 4.1.

Corollary. *Each individual cell plays a role.*

This does not contradict the fact that frequently several cells send similar message, i.e. there exists a redundancy, which is opposite to the stability or genericity of the whole layer. However, as said before, in some regime and/or for m sufficiently small, the redundancy is not a simple repetition, it is more like a creation of characteristic properties.

Let us look at a neuron $a \in h_t$, and consider the model (4.21). If $u = u^a(\xi)$ does not change of sign, the dynamic of the neuron a is stable under small perturbations. For $u > 0$, it looks like a linear function, it is monotonic. For $u < 0$ there exist a unique stable minimum and a unique saddle point which limits its basin of attraction. But for $u = 0$ the critical points collide, the individual map is unstable. This is named the *catastrophe* point. For the whole theory, see [Tho72], [AGZV12a].

If we are interested in the value of η_t^a, as this is the case in the analysis of the cat's manifolds seen before, for understanding the information flow layer by layer, we must also consider the levels of the function, involving v^a then Λ. This asks to follow a sort of inversion of the flow, going to the past, by finding the roots z of the equations

$$P^a(z) = c. \tag{4.30}$$

Depending on u and v, there exist one root or three roots. For instance, for $c = 0$, the second case happens if an only if the numbers $u^a(\xi), v^a(\xi)$ satisfy the inequality $4u^3 + 27v^2 < 0$. When the point $(u^a(\xi), v^a)$ in the plane Λ belongs to the *discriminant* curve Δ of equation $4u^3 + 27\eta^2 = 0$, things become ambiguous, two roots collide and disappear together for $4u^3 + 27v^2 > 0$.
These accidents create ramifications in the cat's manifolds.

This analysis must be applied independently to all the neurons $a = 1, \ldots, m$ in h, that is to all the axis in H. If α is an invertible endomorphism, the set of inversions has a finite number of solutions, less than 3^m.

Remind that the region around 0 in the space H is especially important, because it is only here that the polynomial model applies numerically, σ and tanh being almost linear around 0. Therefore the set of data η_{t-1} and ξ_t which gives some point η_t in this region have a special meaning: they represent ambiguities in the past for η_{t-1} and critical parameters for ξ_t. Thus the discriminant Δ of equation $4u^3 + 27v^2 = 0$ in

Λ plays an important role in the global dynamic.

The inversion of $X_\xi^w : H \to H$ is impossible continuously along a curve in ξ whose u^a, v^a meet Δ for some component a. It becomes possible if we pass to complex numbers, and lift the curve in Λ to the universal covering $\Lambda_\star^\sim(\mathbb{C})$ of the complement $\Lambda_\mathbb{C}^\star$ of $\Delta_\mathbb{C}$ in $\Lambda_\mathbb{C}$ [AGZV12b].
The complex numbers have the advantage that every degree k polynomials has k roots, when counted with multiplicities. The ambiguity in distinguishing individual roots along a path is contained in the Poincaré fundamental group $\pi_1(\Lambda_\mathbb{C}^\star)$. However the precise definition of this group requires the choice of a base point in $\Lambda_\mathbb{C}^\star$, then it is more convenient to consider the fundamental groupoid $\Pi(\Lambda_\mathbb{C}^\star) = \mathcal{B}_3$, which is a category, having for points the elements of $\Lambda_\mathbb{C}^\star$ and arrows the homotopy classes of paths between two points. The choice of an object λ_0 determine $\pi_1(\Lambda_\mathbb{C}^\star; \lambda_0)$, which is the group of homotopy classes of loops from λ_0 to itself, i.e. the isomorphisms of λ_0 in $\mathcal{B}_3$. This group is isomorphic to the Artin braid group B_3 of braids with three strands [AGZV12b].

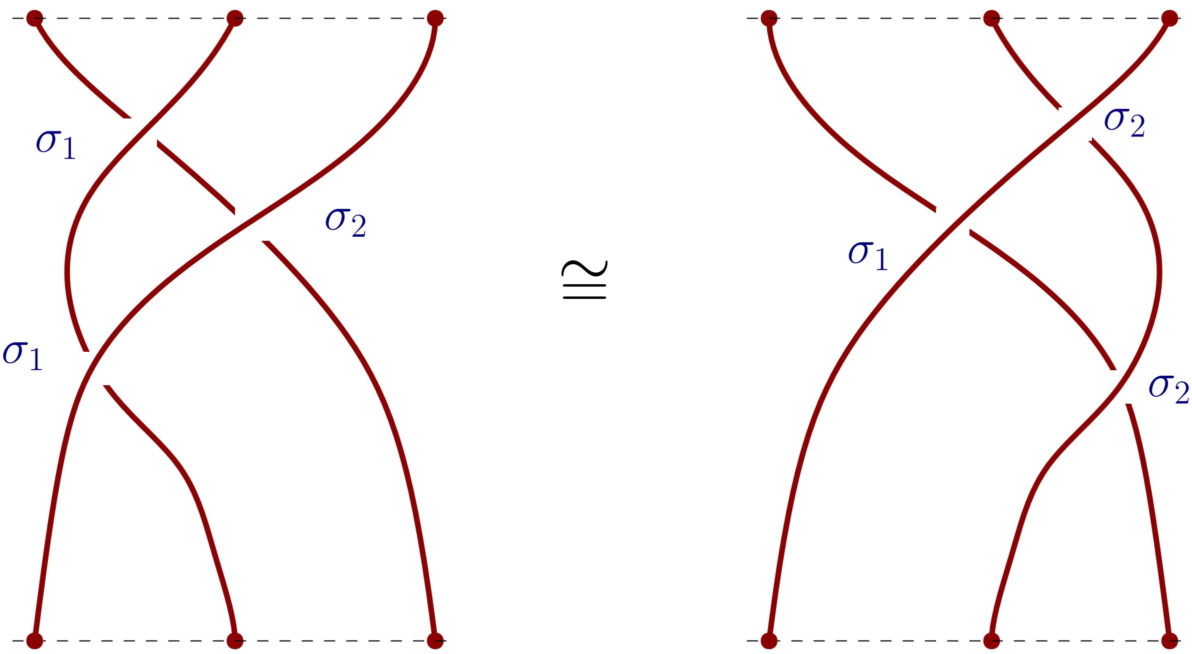


Figure 4.3: Two homotopic braids

This group B_3 is generated by two loops σ_1, σ_2 that could be defined as follows: take a line $u = u_0 \in \mathbb{R}_- \subset \mathbb{C}$,, with complex coordinate v, and let v_0^+, v_0^- be the positive and negative square roots of $-\frac{4}{27}u_0^3$; the loop $\sigma_1 = \sigma^+$ (*resp.* $\sigma_2 = \sigma^-$) is based in 0, contained in the line $u = u_0$ and makes one turn in the trigonometric sense around v_0^+ (*resp.* v_0^-). The relations between σ_1 and σ_2 are generated by $\sigma_1\sigma_2\sigma_1 = \sigma_2\sigma_1\sigma_2$.
The center C of B_3 is generated by $c = (\sigma_1\sigma_2)^3$. The quotient by this center is isomorphic to the group B_3/C generated by $a = \sigma_1\sigma_2\sigma_1$ and $b = \sigma_1\sigma_2$ satisfying $a^2 = b^3$; the quotient of B_3/C by a^2 is the Möbius group $PSL_2(\mathbb{Z})$ of integral homographies, and the quotient of B_3/C by a^4 is the modular group $SL_2(\mathbb{Z})$ of integral matrices of determinant one, then a two fold covering of $PSL_2(\mathbb{Z})$. The quotient $\mathfrak{S}_3$ of B_3 is

defined by the relations $\sigma_1^2 = \sigma_2^2 = 1$, and by the relation which defines B_3, i.e. $\sigma_1\sigma_2\sigma_1 = \sigma_2\sigma_1\sigma_2$ (see figure 4.3).

Of course the disadvantage of the complex numbers is the difficulty to compute with them in DNNs, for instance σ and tanh extended to $\mathbb{C}$ have poles. Moreover all the dynamical regions are confounded in $\Lambda^\star_{\mathbb{C}}$; in some sense the room is too wide. Therefore, we will limit ourselves to the sub-category $\Pi_{\mathbb{R}} = \mathcal{B}_3(\mathbb{R})$, made by the real points of $\Lambda^\star$, but retaining all the morphisms between them, that is a full sub-category of $\mathcal{B}_3$. This means that only the paths are imaginary in $\mathcal{B}_3(\mathbb{R})$.

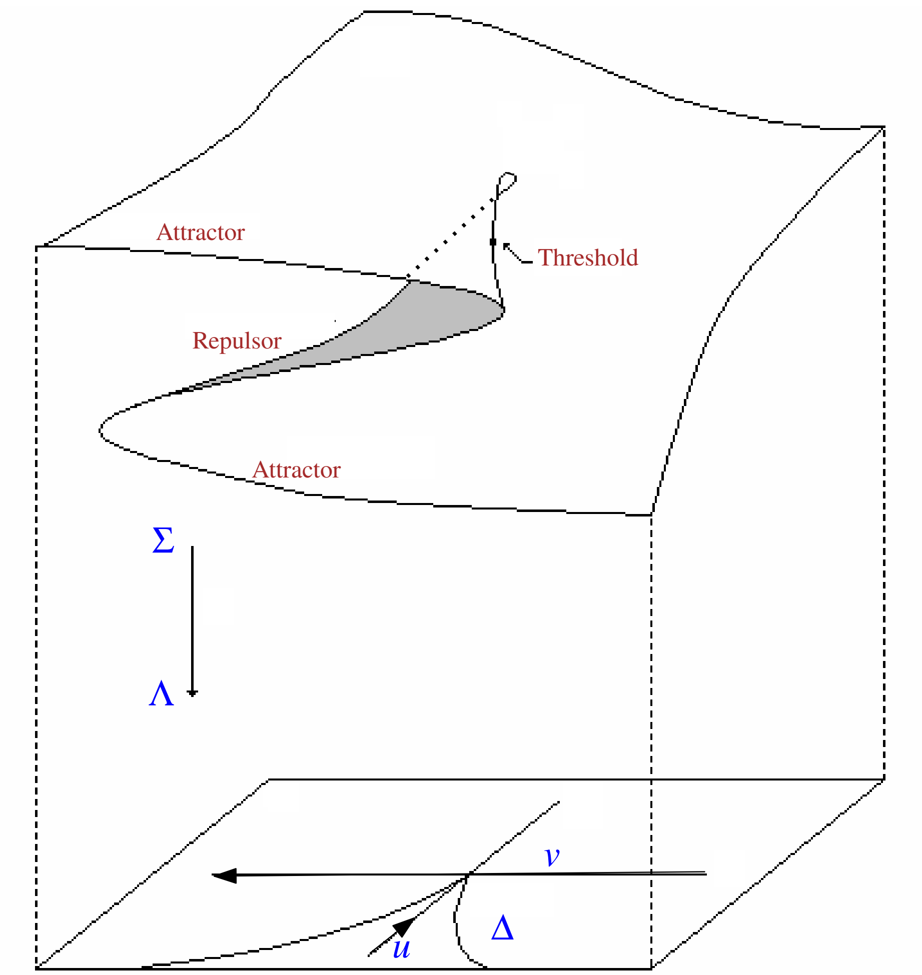


Figure 4.4: Cusp

Another sub-groupoid could be also useful (see figure 4.4): consider the gathered surface Σ in $\Lambda \times \mathbb{R}$ of equation $z^3 + uz + v = 0$; let Δ_3 be the natural lifting of Δ along the folding lines of Σ over Λ, the complement $\Sigma^\star$ of Δ_3 in Σ can be canonically embedded in the complex universal covering $\Lambda^{\sim}_{\star}$, based in the real contractile region Λ_0 inside the real cusp, by taking, for each $(u,v) = \lambda$ in Λ_0 the points λ_+ and λ_- respectively given by the paths $\sigma^+ = \sigma_1$ and $\sigma^- = \sigma_2$, which make simple turn over the branches of the cusp. When λ approaches one of these branches, the corresponding point collides with it on

Δ_3, but the other point continues to be isolated then the construction gives an embedding of $\Sigma^\star$. Therefore we can define the full sub-groupoid of $\mathcal{B}_3$ which has as objects the points of $\Sigma^\star$, and name it $\mathcal{B}_3^r$ or Π_r.

Remark. The groupoid Π_r can be further simplified, by taking one point in each region of interest: one point outside the preimage of the cusp Δ, and three points in each region over the interior of the cusp.

Remark. These four points correspond to the four real structures of Looijenga in the complex kaleidoscope [Loo78].

The groupoid $\mathcal{B}_3^r$ is naturally equipped with a covering (surjective) functor π to the groupoid $\mathcal{B}_3(\mathbb{R})$ of real points.
The interest of $\mathcal{B}_3^r$ with respect to $\mathcal{B}_3(\mathbb{R})$ is that it distinguishes between the stable minimum and the unstable one in the regime $u < 0$. But the interest of $\mathcal{B}_3(\mathbb{R})$ with respect to $\mathcal{B}_3^r$ is that it speaks only of computable quantities u, v without ambiguity, putting all the ambiguities in the group B_3.

All these groupoids are connected, the two first ones, $\mathcal{B}_3(\mathbb{R})$ and $\mathcal{B}_3^r$ because they are full subcategories of the connected groupoid $\mathcal{B}_3$, the other ones in virtue of the definition of a quotient (to the right) of a groupoid by a normal sub-group H of its fundamental group G: it has the same objects, and two arrows f, g from a to b are equivalent if they differ by an element of H. This is meaningful because in Aut_a (*resp.* Aut_b) the sub-group H_a (*resp.* H_b) is well defined, being normal, and moreover $f^{-1}g \in H_a$ is equivalent to $gf^{-1} \in H_b$.

Cardan formulas expresses the roots by using square roots and cubic roots. They give explicit formulas for the differences of roots $z_2 - z_1, z_3 - z_1$. They can be seen directly in the surface Σ.

Remarks. These formulas correspond to the simplest non trivial case of a map of period:

(i) integral classes of the homology $H_0(P_{u,v}^{-1}(0)$ are transported along paths;

(ii) the holomorphic form dz is integrated on the integral classes.

This gives a linear representation of B_3, which factorizes through $\mathfrak{S}_3$.

Augment the variable z by a variable y, the roots can be completed by the levels $Z_{u,v}$ over $(u, v) \in \Lambda$, which are the elliptic curves

$$P_{u,v}(z, y) = z^3 + y^2 + uz + v = 0, \tag{4.31}$$

the 2-form $\omega = dz \wedge dy$ can be factorized as follows

$$\omega = -\frac{1}{2} dP \wedge \frac{dz}{y}; \tag{4.32}$$

the integral of dx/y over the curve $Z_{u,v}$ is an elliptic integral, its periods over integral cycles, gives a linear representation of B_3 which factorizes through $SL_2(\mathbb{Z})$.
Every stabilization of z^3 by a quadratic form gives rise to the representation of the first case in odd dimension and of the second case in even dimension.

Natural groupoids smaller than $\mathcal{B}_3$ are given by quotienting the morphisms, replacing B_3 by $\mathfrak{S}_3$ or $SL_2(\mathbb{Z})$ or its projective version $PSL_2(\mathbb{Z})$ made by homographies.

4.5 Pre-semantics

The natural languages have many functions, from everyday life to poetry and science, or politics and law, however all of them rely on cognitive operations about meanings and shapes, as they appear in the many language-games of Wittgenstein or the action/perception dimensions of Austin. Cf. [Wit53], [Aus61].

The linguist Antoine Culioli, having studied in depth a great variety of natural languages, tried to characterize some of these operations in meta-linguistic, for instance the generic structure and dynamics of a *notional domain*. The notion here can be "dog" or "cat" or "good" or "absent" or anything which has a meaning for most peoples, or specialists in some field. To have a meaning must involve in general several occurrences and disappearances of the notion, a knowledge of its possible properties and individuations, in a language and in the world (data for instance, relations between them and classifications).
A good reference is the book *Cognition and Representation in Linguistic Theory*, A. Culioli, Benjamins, [CLS95].

The notional domain has an interior I where the properties of the notion are sure, an exterior E where the properties are false, and a boundary B, where things are more uncertain. A path through the boundary goes from "truly P" to "truly not P", through an uncertain region where "non-really P, non really not P" can be said. In the center of I are one or several prototypes of the notion. A kind of gradient vector leads the mind to these archetypes, that Culioli named attracting centers, or attractors; however he wrote in 1989 (upcit.) the following important precision: "Now the term attractor cannot be interpreted as an attainable last point (...) but as the representation of the imaginary absolute value of the property (the predicate) which organizes an aggregate of occurrences into a structured notional domain." Culioli also used the term of organizing center, but as we shall see this would conflict with another use.

The division I, B, E takes all its sense when interrogative mode is involved, or negation and double negation, or intero-negative mode. In negation you go out of the interior, in interro-negation you come back inside from E. "Is your brother really here" (it means that "I do not expect that your brother is here".) "Now that, that is not a dog!" (you place yourself in front of proposition P, or inside the notion I, you know what is a dog, then goes to E); "Shall I still call that a dog?" "I do not refuse to help"; here come back in I of "help" after a turn in its exterior E. All these circumstances involve an imaginary place IE, where the regions are not separated, this is like the cuspidal point before the separation of the branches I and E of the cusp.

Mathematically this corresponds precisely to the creation of the external (*resp.* internal) critical point of $z^3 + uz + v$, on the curve Δ. Example: "he could not have left the window open", the meaning mobilizes the place IE of indetermination, the maximum of ambiguity, where the two actions, "left" and "not to left" are possible, then one of them is forbidden, and "not having left" is retained by the negation. In the terminology of Thom, the place IE is the *organizing center*, the function z^3 itself, the most degenerate one in the stable family, giving birth to the unfolding.

To describe the mechanisms beyond these paths, Culioli used the model of the *cam*: "the movement travels from one place to another, only to return to the initial plane". Example: start from IE, then make a half-turn around I which passes by E then come to I by another half-turn. "This book is only slightly interesting." Here the meaning only appears if you imagine the place where interesting and not interesting are not yet separated, then go to not interesting and finally temperate the judgment by going to the boundary, near I; the compete turn leads you in another place, over the same point, thus the meaning is greatly in the path, as an enclosed area. "This book is not uninteresting" means that it is more than interesting. The paths here are well represented on the gathered real surface Σ, of equation

$$z^3 + uz + v = 0, \tag{4.33}$$

but they can also be made in the complement of Δ in Λ in a complexified domain. It seems that only the homotopy class is important, not the metric, however we cannot neglect a weakly quantitative aspect, on the way of discretization in the nuances of the language. Consequently, the convenient representation of the moves of Culioli is in the groupoid $\mathcal{B}_3^r$, that we propose to name the *Culioli groupoid*.

Remind that $LSTM$ and the other memory cells are mostly used in chains, to translate texts.
It is natural to make a rapprochement between their structural and dynamical properties and the meta-linguistic description of Culioli. In many aspects René Thom was closed to Culioli in his own approach of semantics, see his book *Mathematical Models of Morphogenesis* [Tho83], which is a translation of a French book published by Bourgois in 1980. The original theory was exposed in [Tho72]. In this approach, all the elementary catastrophes having a universal unfolding of dimension less than 4 are used, through their sections and projections, for understanding in particular the valencies of the verbs, from the semantic point of view, according to Peirce, Tesnière, Allerton: impersonal, "it rains", intransitive "she sleeps", transitive "he kicks the ball", triadic "she gives him a ball", quadratic "she ties the goat to a tree with a rope".
The list of organizing centers is as follows:

$$y = x^2, \quad y = x^3, \quad y = x^4, \quad y = x^5, \quad y = x^6,$$
$$y = x_1^3 - x_2^2 x_1, \quad y = x_1^3 + x_2^3 \quad (or \quad y = x_1^3 + x_2^2 x_1), \quad y = x_1^4 + x_2^2 x_1; \tag{4.34}$$

respectively named: *well*, *fold*, *cusp*, *swallowtail*, *butterfly*, *elliptic umbilic*, *hyperbolic umbilic* and *parabolic umbilic*, or with respect to the group which generalizes the Galois group $\mathfrak{S}_3$ for the fold, respectively: A_1, A_2, A_3, A_4, A_5, $D_4^+ = D_4^- = D_4$ and D_5. The A_n are the symmetric groups $\mathfrak{S}_{n+1}$ and the D_n index two subgroups of the symmetry groups of the hypercubes I^n [Ben86].

It is not difficult to construct networks, on the model of $MLSTM$, such that the dynamics of neurons obey to the unfolding of these singular functions. The various actors of a verb in a sentence could be separated input data, for different coordinates on the unfolding parameters. The efficiency of these cells should be tested in translation.

Coming back to the memory cell (4.21), the critical parameters x_t over Δ can be interpreted as boarders between regions of notional domains.

The precise learned $2mn$ weights w_x for the coefficients u^a and v^a, for $a = 1, \ldots, m$, together with the weights in the forms α^a for h_{t-1} gives vectors (or more accurately matrices), which are like readers of the words x in entry, taking in account the contexts from the other words through h. Remember Frege: a word has a meaning only in the context od a sentence. This is a citation of Wittgenstein, after he said that "Naming is not yet a move in a language-game" [Wit53, p. 49].

To get "meanings", the names, necessarily embedded in sentences, must resonate with other contexts and experiences, and must be situated with respect to the discriminant, along a path, thus we suggest that the vector spaces of "readers" W, and the vector spaces of states X are local systems A over a fibered category $\mathcal{F}$ in groupoids $\mathcal{B}_3^r$ over the network's category C.
In some circumstances, the groupoid $\mathcal{B}_3^r$ can be replaced by the quotient over objects $\mathcal{B}_3(\mathbb{R})$, or a quotient over morphisms giving SL_2 or $\mathfrak{S}_3$.

The case of z^3 corresponds to A_2. It is tempting to consider the case of D_4, i.e. the elliptic and hyperbolic umbilics, because their formulas are very closed to $MGU2$ as mentioned at the end of the preceding section.

This would allow the direct coding and translation of sentences by using three actant.

$$\eta = z^3 \mp zw^2 + uz + vw + x(z^2 + w^2) + y. \tag{4.35}$$

5 A natural 3-category of deep networks

In this chapter, we introduce a natural 3-category for representing the morphisms, deformations and surgeries of semantic functioning of $DNNs$ based on various sites and various stacks, which have connected models in their fibers.

Grothendieck's derivators will appear at two successive levels:

1. formalizing internal aspects of this 3-category;
2. defining potential invariants of information over the objects of this 3-category. Therefore we can expect that the interesting relations (for the theory and for its applications) appear at the level of a kind of "composition of derivators", and are analog to the spectral sequences of [Gro57].

5.1 Attention moduli and relation moduli

In addition to the chains of $LSTM$, another network's component is now recognized as essential for most of the tasks in linguistic: to translate, to complete a sentence, to determine a context and to take into account a context for finding the meaning of a word or sentence. This modulus has its origin in the *attention operator*, introduced by Bahdanau et al. [BCB16], for machine translation of texts. The extended form that is the most used today was defined in the same context by Vaswani et al. 2017 [VSP+17], under the common name of *transformer* or simply *decoder*.

Let us describe the steps of the algorithm: the input contains vectors Y representing memories or hidden variables like contexts, and external input data X also in vectorial form.

1) Three sets of linear operators are applied:

$$
\begin{aligned}
Q &= W^Q[Y], \\
K &= W^K[Y, X], \\
V &= W^V[Y];
\end{aligned}
$$

where the W's are matrices of weights, to be learned. The vectors Q, K, V are respectively called *queries*, *keys* and *values*, from names used in Computer Science; they are supposed to be indexed by "heads" $i \in I$, representing individuals in the input, and by other indices $a \in A$, representing for instance different instant times, or aspects, to be integrated together. Then we have vectors Q_i^a, K_i^a, V_i^a.

2) The inner products $E_i^a = k(Q_i^a | K_i^a)$ are computed (implying that Q and K have the same dimension), and the soft-max function is applied to them, giving a probability law, from the Boltzmann weights of energy E_i^a
$$p_i^a = \frac{1}{Z_i^a} e^{E_i^a}, \tag{5.1}$$

3) a sum of product is computed
$$V_i' = \sum_a p_i^a V_i^a. \tag{5.2}$$

4) A new matrix is applied in order to mix the heads
$$A_j = \sum_i w_j^i V_i'. \tag{5.3}$$

All that is summarized in the formula:
$$A_j(Y,X) = \sum_i \sum_a w_j^i \text{softmax}\left[k(W^Q(Y)_i^a | W^K(Y,X)_i^a)\right] W^V(Y)_i^a. \tag{5.4}$$

A remarkable point is that, as it is the case for $MGU2$ or $LSTM$ and GRU cells, the transformer corresponds to a mapping of degree 3, made by multiplying a linear form of Y with non-linear function of a bilinear form of Y. Strictly speaking the degree 3 is only valid in a region of the parameters. In other regions, some saturation decreases the degree.

Chains of $LSTM$ were first used for language translations, and were later on used for image description helped by sentences predictions, as in [KL14] or [MXY+15], where they proved to outperform other methods for detection of objects and their relations.

In the same manner, the concatenation of attention cells has been proven to be very beneficial in this context [ZRS+18], then it was extended to develop reasoning about the the relations between objects in images and videos [RSB+17], [BHS+18], [BHS+18], [SRB+17], or [DHSB20].

In the $MHDPA$ (multi-head dot product attention) algorithm [SFR+18], the inputs X are either words, questions and features of objects and their relations, coded into vectors, the inputs Y combine hidden and external memories, the outputs A are new memories, new relations and new questions.

Remark. Interestingly, the method combines fully supervised learning with unsupervised learning (or adaptation) by maximization of a learned functional of the above variables.

In particular, the memories or hidden variables issued from the transformer were re-introduced in the $LSTM$ chain; giving the following symbolic formulas:

$$c_t = c_{t-1} \odot \sigma_f(x_t, h_{t-1}) \oplus \sigma_i(x_t, m_t) \odot \tau_h(x_t, h_{t-1}); \tag{5.5}$$

where m_t results of transformer applied to the antecedent sequence of h_s, c_s and x_s ; and

$$h_t = \sigma_o(x_t, h_{t-1}) \odot \tanh c_t. \tag{5.6}$$

Geometrically, this can be seen as a *concatenation of folds*, as proposed by Thom *Esquisse d'une Sémiophysique* [Tho88], to explain many kinds of organized systems in biology and cognition. From this point of view, the concatenation of folds, giving the possibility of coincidence of cofolds [Arg78], is a necessary condition for representing the emergence of a meaningful structure and oriented dynamic in a living system.
Note that, in the unsaturated regimes, h_t has a degree 5 in h_{t-1}, then its natural groupoid can be embedded in a braids groupoid of type $\mathcal{B}_5$. This augmentation, from the fold to the so called *swallowtail*, could explain the greatest syntactic power of the $MHDPA$ with respect to $LSTM$. However the concrete use of more memories in times s before t makes the cells much more complex than a simple mapping from $t-1$ to t.

The above algorithm can be composed with other cells for detecting relations. For instance, Raposo et al. [RSB+17] have defined a *relation operator*: having produced contexts H or questions Q concerning two objects o_i, o_j by a chain of $LSTM$ (that can be helped by external memories and attention cells) the answer is taken from a formula:

$$A = f\left(\sum_{i,j} g\left(o_i, o_l; Q, H\right)\right), \tag{5.7}$$

where f and g are parameterized functions, and $o_i : i \in I$ are vectors representing objects with their characteristics.
The authors insisted on the important invariance of this operator by the permutation group $\mathfrak{S}_n$ of the objects.

More generally, composed networks were introduced in 2016 by Andreas et al. [ARDK16] for question answering about images. The reasoning architecture MAC, defined by Hudson and Manning, [HM18], is composed of three attention operators named *control*, *write* and *read*, in a DNN, inspired from the architecture of computers.

This leads us to consider the evolution of architectures and internal fibers of stacks and languages, in relation to the problems to be solved in semantic analysis.

5.2 The 2-category of a network

For representing languages in DNNs, we have associated to a small category C the class $\mathcal{A}_C = \mathsf{Grpd}_C^\wedge$ of presheaves over the category of fibrations in groupoids over C. The objects of $\mathcal{A}_C$ were described in terms of presheaves A_U on the fibers $\mathcal{F}_U$ for $U \in C$ satisfying gluing conditions, cf. sections 3 and 4.

Remark. Other categories than groupoids, for instance posets or fibrations in groupoids over posets, can replace the groupoids in this section, and are useful in the applications, as we mentioned before, and as we will show in the forthcoming article on semantic communication.

Natural morphisms between objects $(\mathcal{F}, A)$ and $(\mathcal{F}', A')$ of $\mathcal{A}_C$ are defined by a family of functors $F_U : \mathcal{F}_U \to \mathcal{F}'_U$, such that for any morphism $\alpha : U \to U'$ in C,

$$F'_\alpha \circ F_{U'} = F_U \circ F_\alpha; \tag{5.8}$$

and by a family of natural transformations $\varphi_U : A_U \to F_U^\star(A'_U) = A'_U \circ F_U$, such that for any morphism $\alpha : U \to U'$ in C,

$$F_{U'}^\star(A'_\alpha) \circ \varphi_{U'} = F_\alpha^\star(\varphi_U) \circ A_\alpha, \tag{5.9}$$

from $A_{U'}$ to $F_\alpha^\star(F_U^\star A'_U) = F_{U'}^\star\left((F'_\alpha)^\star A'_U\right)$.
Note that the family $\{F_U; U \in C\}$ is equivalent to a C-functor $F : \mathcal{F} \to \mathcal{F}'$ of fibered categories in groupoids, and the family φ_U is equivalent to a morphism φ in the topos $\mathcal{E}_\mathcal{F}$ from the object A to the object $F^\star(A')$.

Remark. These morphisms include the morphisms already defined for the individual classifying topos $\mathcal{E}_\mathcal{F}$. But, even for one fibration $\mathcal{F}$ and its topos $\mathcal{E}$, we can consider non-identity end-functor from $\mathcal{F}$ to itself, which give new morphisms in $\mathcal{A}_C$.

The composition of $(F_U, \varphi_U); U \in C$ with (G_U, ψ_U) from $(\mathcal{G}, B)$ to $(\mathcal{F}, \mathcal{A})$ is defined by the ordinary composition of functors $F_U \circ G_U$, and the twisted composition of natural transformation

$$(\varphi \circ \psi)_U = G_U^\star(\varphi_U) \circ \psi_U : B_U \to (F_U \circ G_U)^\star A'_U. \tag{5.10}$$

This rule gives a structure of category to $\mathcal{A}_C$.

In addition, the natural transformations between functors give the vertical arrows in $\mathsf{Hom}_\mathcal{A}(\mathcal{F}, A : \mathcal{F}', A')$, that form categories:
a morphism from (F, φ) to (G, ψ) is a natural transformations $\lambda : F \to G$, which in this case with groupoids, is an homotopy in the nerve, plus a morphism $a : A \to A$, such that

$$A'(\lambda) \circ \varphi = \psi \circ a : A \to G^\star A'. \tag{5.11}$$

For a better understanding of this relation, we can introduce the points (U,ξ) in $\mathcal{F}$ over C, and read

$$A'_U(\lambda_U(\xi)) \circ \varphi_U(\xi) = \psi_U(\xi) \circ a_U(\xi) : A_U(\xi) \to A'_U(G_U(\xi)). \tag{5.12}$$

This can be understood geometrically, as a lifting of the deformation λ to a deformation of the presheaves.

Vertical composition is defined by usual composition for the deformations λ and ordinary composition in $\mathsf{End}(A)$ for a. Horizontal compositions are for $\mathcal{F} \to \mathcal{F}' \to \mathcal{F}$".

Horizontal arrows and vertical arrows satisfy the axioms of a 2-category [Gir71], [Mac71].
This structure encodes the relations between several semantics over the same network.

The relations between several networks, for instance moduli inside a network, or networks that are augmented by external links, belong to a 3-category, whose objects are the above semantic triples, and the 1-morphism are lifting of functors between sites $u : C \to C'$.

[Gir71, Theorem 2.3.2] tells us that, as for ordinary presheaves, there exist natural right and left adjoints $u_\star$ and $u_!$ respectively of the pullback $u^\star$ from the 2-category $\mathsf{Cat}_{C'}$ of fibrations over C' to the 2-category Cat_C of fibrations over C. They are natural 2-functors, adjoint in the extended sense. These 2-functors define adjoint 2-functors between the above 2-categories of classifying toposes $\mathcal{A}_C$ and $\mathcal{A}_{C'}$, by using the natural constructions of $SGA4$ for the categories of presheaves. They can be seen as substitutions of stacks and languages induced by functors u.

The construction of $\mathcal{A}_C$ from C is a particular case of Grothendieck's derivators [Cis03].

5.3 Grothendieck derivators and semantic information

For $\mathcal{M}$ a closed model category, the map $C \mapsto \mathcal{M}_C$, or $\mathcal{M}_C^\wedge$ (see section 2.4), is an example of *derivator* in the sense of Grothendieck. References are [Gro83], [Gro90], the three articles of Cisinski [Cis03], and the book of Maltsiniotis on the homotopy theory of Grothendieck [Mal05].

A derivator generalizes the passage from a category to its topos of presheaves, in order to develop homotopy theory, as topos were made to develop cohomology theory. It is a 2-functor $\mathbb{D}$ from the category Cat (or a special sub-category of diagrams, for instance Poset) to the 2-category CAT, satisfying four axioms.

a) The first one tells us that $\mathbb{D}$ transforms sums of categories into products,

b) The second one that isomorphisms of images can be tested on objects,

c) the third one that there exists, for any functor $u : C \to C'$, a right adjoint $u_\star$ (defining homotopy limit) and a left adjoint $u_!$ (defining homotopy colimit) of the functor $u^\star = \mathbb{D}(u)$;

d) the fourth axiom requires that these adjoints are defined locally; for instance, if $X' \in C'$, and $F \in \mathbb{D}(C)$, therefore $u_\star F \in \mathbb{D}(C)'$, the fourth axiom tells us that

$$(u_\star F)_{X'} \cong p_\star j^\star F; \tag{5.13}$$

where j is the canonical map from $C|X'$ to C, and p the unique morphism from $C|X'$ to $\star$.

Another formula that expresses the same thing is

$$(u_\star F)_{X'} \cong H^\star\left(C|X'; F|_{C|X'}\right), \tag{5.14}$$

abstract version of a Kan extension formula.
In general, the cohomology is defined by

$$H^\star(C;F) = (p_C)_\star F \in \mathbb{D}(\star). \tag{5.15}$$

A first example of derivator is given by an Abelian category Ab, like commutative groups or real vector spaces, and it is defined by the derived category of differential complexes, where quasi-isomorphisms (isomorphisms in cohomology) are formally inverted,

$$\mathbb{D}(I) = \mathsf{Der}(\mathsf{Hom}(I^{\mathrm{op}}, \mathsf{Ab})). \tag{5.16}$$

Another kind of example is a *representable derivator*

$$\mathbb{D}_{\mathcal{M}}(I) = \mathsf{Funct}(I^{\mathrm{op}}, \mathcal{M}), \tag{5.17}$$

where $\mathcal{M}$ is a closed model category. This can be seen as a non-Abelian generalization of the above first example.

A third kind of examples is given by the topos of sheaves over a representable derivator $\mathcal{M}_C^\wedge$.

Then representable derivators allow to compare the elements of semantic functioning between several networks, for instance a network with a sub-network of this network, playing the role of a module in computation.

Consider the sub-categories Θ_P, over the languages $\mathcal{A}_\lambda, \lambda \in \mathcal{F}_U$, made by the theories that exclude a rigid proposition $P = !\Gamma$, in the sense they contain $P \Rightarrow \Delta$, for a given chosen Δ, (see appendix E). The right slice category $P|\mathcal{A}_\lambda$ acts on Θ_P. The information spaces F define an object of $\mathcal{M}_{\Theta_P}$, its cohomology allow us to generalize the cat's manifolds, that we defined below with the connected components of the category $\mathcal{D}$, in the following way: the dynamical object $\mathbb{X}$ is assumed to be defined over the stack $\mathcal{F}$, then the dynamical space $g\mathbb{X}$ is defined over the nerve of $\mathcal{F}$, and the semantic functioning gives a simplicial map $gS : g\mathbb{X} \to gI^\bullet$ from $g\mathbb{X}$ space to the equipped theories, then we can consider the inverse image of Θ_P over the functioning network. Composing with F we obtain a parameterized object M_P in $\mathcal{M}$, defining a local system over the category associated to $g\mathbb{X}$, which depends on Γ, Δ. This represents the semantic information in $\mathbb{X}$ about the problem of (rigidly) excluding P when considering that Δ is (thought to be) false. Seen as an element of $\mathbb{D}(g\mathbb{X})$, its cohomology is an homotopical invariant of the information.

In this text, we have defined information quantities, or information spaces, by applying cohomology or homotopy limits, over the category $\mathcal{D}$ which expresses a triple $C, \mathcal{F}, \mathcal{A}$, made by a language over a pre-semantic over a site. The Abelian situation was studied through the bar-complex of cochains of the module of functions Φ on the fibration $\mathcal{T}$ of theories Θ over the category $\mathcal{D}$. A non-Abelian tentative, for defining spaces of information, was also proposed at this level, using (in the non-homogeneous form) the functors F from Θ_{loc} to a model category $\mathcal{M}$ (see section 3.5). Therefore information spaces were defined at the level of $\mathcal{M}_{\mathcal{T}}$, not at a level $\mathcal{M}_C$.

Information spaces belong to $\mathbb{D}_{\mathcal{M}}(\mathcal{T})$. To compare spaces of information flows in two theoretical semantic networks, we have at disposition the adjoint functors $\varphi_\star, \varphi_!$ of the functors $\varphi^\star = \mathbb{D}(\varphi)$ associated to $\varphi : \mathcal{T} \to \mathcal{T}'$, between categories of theories. Those functors φ can be associated to changes of languages $\mathcal{A}$, changes of stacks $\mathcal{F}$ and/or changes of basic architecture C.

An important problem to address, for constructing networks and applying deep learning efficiently to them, is the realization of information relations or correspondences, by relations or correspondences between the underlying invariance structures. For instance, to realize a family of homotopy equivalences (*resp.* fibration, *resp.* cofibration) in $\mathcal{M}$, by transformations of languages, stacks or sites having some properties, like enlargement of internal symmetries.

The analog problem for presheaves (set valued) is to realize a correspondence (or relation) between the topos $\mathcal{I}^\wedge$ and $(\mathcal{I}')^\wedge$ from a correspondence between convenient sites for them.

For toposes morphisms this is a classical result (see [AGV63, 4.9.4] or the Stacks project [Sta, 7.16n 2.29]) that any geometric morphism $f_\star : \mathsf{Sh}(I) \to \mathsf{Sh}(J)$ comes from a morphism of sites up to topos equivalence between I and I'. More precisely, there exists a site I' and a cocontinuous and continuous functor $v : I \to I'$ giving an equivalence $v_! : \mathsf{Sh}(I) \to \mathsf{Sh}(I')$ extending v, and a site morphism $J \to I'$, given by a continuous functor $u : I' \to J$ such that $f_\star = u_\star \circ v_!$.

From [Shu12], a geometric morphism between $\mathsf{Sh}(I)$ and $\mathsf{Sh}(J)$ comes from a morphism of site if and only if it is compatible with the Yoneda embeddings.

5.4 Stacks homotopy of DNNs

The characterization of fibrant and cofibrant objects in $\mathcal{M}_C$ was the main result of chapter 2. All objects of $\mathcal{M}_C$ are cofibrant and the fibrant objects are described by theorem 2.2; we saw that they correspond to ideal semantic flows, where the condition $\pi^\star\pi_\star = \mathsf{Id}$ holds. They also correspond to the contexts and the types of a natural $M - L$ theory. The objects of $Ho(\mathcal{M}_C)$, [Qui67], are these fibrant and cofibrant objects of $\mathcal{M}_C$, the Ho morphisms being the homotopy classes of morphisms in $\mathcal{M}_C$, generated by inverting formally zigzags similar to the above ones. Thus we get a direct access to the homotopy category $Ho\mathcal{M}_C$. The Ho morphisms are the homotopy equivalences classes of the substitutions of variables in the $M - L$ theory.

From the point of view of semantic information, we just saw that homotopy is pertinent at the next level: looking first at languages over the stacks, then at some functors from the posets of theories to a test model category $\mathcal{M}'$, then going to $Ho(\mathcal{M}')$. However, the fact that we restrict to theories over fibrant objects and fibrations between them, implies that the homotopy of semantic information only depends on the images of these theories over the category $Ho(\mathcal{M}_C)$. How to use this fact for functioning networks?

Appendices

A Locallic topos and Fuzzy identities

Definitions. *let* Ω *be a complete Heyting algebra; a* set over Ω, (X,δ), *also named an* Ω-set, *is a set* X *equipped with a map* $\delta : X \times X \to \Omega$, *which is symmetric and transitive, in the sense that for any triple* x, y, z, *we have* $\delta(x,y) = \delta(y,x)$ *and*

$$\delta(x,y) \wedge \delta(y,z) \leq \delta(x,z). \tag{18}$$

Note that $\delta(x,x)$ *can be different from* $\top$.
But we always have $\delta(x,y) = \delta(x,y) \cap \delta(y,x) \leq \delta(x,x)$, *and* $\delta(x,y) \leq \delta(y,y)$.
As Ω *is made for fixing a notion of relative values of truth,* δ *is interpreted as* fuzzy equality *in X; it generalizes the characteristic function of the diagonal when* Ω *is boolean. In our context of DNN, it can be understood as the progressive decision about the outputs on the trees of layers rooted in a given layer.*
A morphism *from* (X,δ) *to* (X',δ') *is an application* $f : X \times X' \to \Omega$, *such that, for every,* x, x', y, y'

$$\delta(x,y) \wedge f(x,x') \leq f(y,x'), \tag{19}$$

$$f(x,x') \wedge \delta'(x',y') \leq f(x,y'); \tag{20}$$

$$f(x,x') \wedge f(x,y') \leq \delta'(x',y'). \tag{21}$$

Moreover

$$\delta(x,x) = \bigvee_{x' \in X'} f(x,x'). \tag{22}$$

Which generalizes the usual properties of the characteristic function of the graph of a function in the boolean case.
The composition *of a map* $f : X \times X' \to \Omega$ *with a map* $f' : X' \times X" \to \Omega$ *is given by*

$$(f' \circ f)(x,x") = \bigvee_{x' \in X'} f(x,x') \wedge f(x',x"). \tag{23}$$

And the identity morphism *is defined by*

$$\mathsf{Id}_{X,\delta} = \delta. \tag{24}$$

This gives the category Set_Ω *of sets over* Ω*, also named* Ω-sets.

The Heyting algebra Ω of a topos $\mathcal{E}$ is made by the subobjects of the final object $\mathbf{1}$; the elements of Ω are named the open sets of $\mathcal{E}$. In fact, there exists an object $\mathbf{\Omega}$ in $\mathbf{E}$, the Lawvere object, such that for every object $X \in \mathcal{E}$, the set of subobjects of X is naturally identified with the set of morphisms $\mathbf{\Omega}^X$. When $\mathcal{E} = \mathsf{Sh}(\mathbf{X})$ is a Grothendieck topos, $\mathbf{\Omega}$ is the sheaf over X, which is defined by $\mathbf{\Omega}(x) = \Omega(\mathcal{E}|x)$, the subobjects of $\mathbf{1}|x$. In the Alexandrov case, $\mathbf{\Omega}(x)$ is the set of open sets for the Alexandrov topology contained in Λ_x.

According to Bell, [Bel08], a localic topos, as the one of a DNN, is naturally equivalent to the category Set_Ω of Ω-sets, i.e. sets equipped with fuzzy identities with values in Ω. We now give a direct explicit construction of this equivalence, because it offers a view of the relation between the network layers directly connected to the intuitionist logic of the topos.
Let us mention the PhD thesis of Johan Lindberg [Lin20, part III], developing this point of view, and studying in details the naturalness of the geometric morphism of topos induced by a morphism of locale.

Definition A.1. *On the poset* $(\Omega, \leq)$*, the canonical Grothendieck topology* K *is defined by the coverings by open subsets of the open sets.*
In the localic case, where we are, the topos is isomorphic to the Grothendieck topos $\mathcal{E} = Sh(\Omega, K)$*.*
We assume that this is the case in the following exposition.

In the particular case $\mathcal{E} = \mathbf{X}^\wedge$, where $\mathbf{X}$ is a poset, Ω is the poset of lower Alexandrov open sets and the isomorphism with $Sh(\Omega, K)$ is given explicitly by proposition 1.2.

Let X be an object of $\mathcal{E}$; we associate to it the set $X^{\mathbf{\Omega}}$ of natural transformation from $\mathbf{\Omega}$ to X. For two elements x, y of $X^{\mathbf{\Omega}}$, we define $\delta_X(x, y) \in \Omega$ as the largest open set over which x and y coincide.
An element u of $X^{\mathbf{\Omega}}$ is nothing else than a sub-singleton in X, its domain ω_u is $\delta_X(u, u)$. In other terms, in the localic case, u is a section of the presheaf X over an open subset ω_u in Ω.
Then, if u, v and w are three elements of $X^{\mathbf{\Omega}}$, the maximal open set where $u = w$ contains the intersection of the open sets where $u = v$ and $v = w$. Thus $X^{\mathbf{\Omega}}$ is a set over Ω.

In the same manner, suppose we have a morphism $f : X \to Y$ in $\mathcal{E}$, if we take $x \in X^{\mathbf{\Omega}}$ and $y \in Y^{\mathbf{\Omega}}$ we define $f(x, y) \in \Omega$ as the largest open set of $\boldsymbol{X}$ where y coincides with $f_\star x$. This gives a morphism of Ω-sets.
All that defines a functor from $\mathcal{E}$ to Set_Ω.

A canonical functor from Set_Ω to $\mathcal{E}$ is given by a similar construction:
for $U \in \Omega$, $\Omega_U = \mathbf{\Omega}(U)$ is an Ω-set, with the fuzzy equality defined by the internal equality

$$\delta_U(\alpha, \alpha') = (\alpha \asymp \alpha'), \tag{25}$$

that is the restriction of the characteristic map of the diagonal subset: $\Delta : \mathbf{\Omega} \hookrightarrow \mathbf{\Omega}\times\mathbf{\Omega}$. The set Ω_U can be identified with the Ω-set $U^{\mathbf{\Omega}}$ associated to the Yoneda presheaf defined by U. More concretely, an element ω of Ω_U is an open subset of U, and its domain $\delta(\omega,\omega)$ is ω itself.
Now, for any Ω-set (X,δ), and for any element $U \in \Omega$, we define the set (see (19),(22)),

$$X_\Omega(U) = \mathsf{Hom}_{\mathsf{Set}_\Omega}(\Omega_U, X) = \{f : \Omega_U \times X \to \Omega\}. \tag{26}$$

In what follows, we sometimes write $X_\Omega = X$, when the notation does not introduce too much ambiguity. If $V \leq W$, the formula $f(\omega_V,\omega_W) = \omega_V \cap \omega_W$ defines a Ω-morphism from Ω_V to Ω_W, which gives a map from $X(W)$ to $X(V)$. Then X_Ω is a presheaf over Ω.

Proposition A.1. *A morphism of Ω-set $f : X\times Y \to \Omega$ gives by composition a natural transformation $f_\Omega : X_\Omega \to Y_\Omega$ of presheaves over Ω.*

Proof. Consider $f_U \in X(U)$; the axiom (22) tells that for every open set $V \subset U$, the family of open sets $f_U(V,u); u \in X$ is an open covering f_U^V of V.
The first axiom of (19), which represents the substitution of the first variable, tells that on $V\cap W$ the two coverings f_U^V and f_U^W coincide. Therefore, for every $u \in X$, the value $f_U(u) = f(U,u)$ of f_U on the maximal element U determines by intersection all the values $f_U(V,u)$ for $V \subset U$.

For $f_U \in X(U)$ and $V \leq U$, the functorial image f_V of f_U in $X(V)$ is the trace on V:

$$\forall u \in X, \quad f_V(u) = \rho_{VU} f_U(u) = f_U(u)\cap V. \tag{27}$$

This implies that X_Ω is a sheaf: consider a covering $\mathcal{U}$ of U, (1) for two elements f_U, g_U of $X(U)$, if the families of restrictions $f_U \cap V; V \in \mathcal{U}$, $g_U \cap V; V \in \mathcal{U}$, then $f_U = g_U$; (2) if a family of coverings $f_V; V \in \mathcal{U}$ is given, such that for any intersection $W = V\cap V'$, the restriction $f_V|W$ and $f_{V'}|W$ coincide, as open coverings, we can define an element f_U of $X(U)$ by taking for each $u \in X$ the open set $f_U(u)$ which is the reunion of all the $f_V(u)$ for $V \in \mathcal{U}$. The union of the sets $f_V(u)$ over $u \in X$ is V, and the union of the sets V is U, then the union of the $f_U(u)$ when u describes X is U. ∎

The second axiom of substitution tells that for any $u,v \in X$, $\delta(u,v)\cap f(u) = \delta(u,v)\cap f(v)$. The third axiom of (19), which expresses the functional character of f, tells that for any $u,v \in X$, $\delta(u,v) \supseteq f(u)\cap f(v)$.

Consequently, the elements of $X(\alpha)$ can be identified with the open coverings $f_U(u); u \in X$ of the open set U, such that, in Ω, we have

$$\forall u,v \in X, \quad f_U(u)\cap f_U(v) \subseteq \delta(u,v) \subseteq (f_U(u) \Leftrightarrow f_U(v)); \tag{28}$$

where $\Leftrightarrow$ denotes the internal equivalence $\Leftarrow \wedge \Rightarrow$ in Ω.
Remind that $\alpha \Rightarrow \beta$ is the largest element $\gamma \in \Omega$ such that $\gamma\wedge\alpha \leq \beta$, and in our topological setting $\Omega = \mathcal{U}(\mathbf{X})$ it is the union of the open sets V such that $V\cap\alpha \subseteq \beta$, therefore $f(u) \Leftrightarrow f(v)$ is the union of the elements V of Ω such that $V\cap f(u) = V\cap f(v)$.

Proposition A.2. *Let Ω be any complete Heyting algebra (i.e. a locale); the two functors $F : (X,\delta) \mapsto (U \mapsto X(U) = \mathsf{Hom}_\Omega(\Omega_U, X)$ and $G : X \mapsto (X^{\mathbf{\Omega}}, \delta_X) = \mathsf{Hom}_{\mathcal{E}}(\mathbf{\Omega}, \mathbf{X})$ define an equivalence of category between* Set_Ω *and* $\mathcal{E} = Sh(\Omega, K)$.

Proof. The composition $F \circ G$ sends a sheaf $X(U); U \in \Omega$ to the sheaf $X^{\mathbf{\Omega}}(U); U \in \Omega$ made by the open coverings of U by sets indexed by the sub-singletons u of X satisfying the two inclusions (28).

Consider an element $s_U \in X(U)$, identified with a section of X over U. For each sub-singleton $v \in X^{\mathbf{\Omega}}$, we define the open set $f(v) = f_U^s(v)$ by the largest open set in U where $v = s_U$. As the sub-singletons generate X, this forms an open covering of U. It satisfies (28) for any pair (u,v): $\delta(u,v)$ is the largest open set where u coincides with v, then the first inclusion is evident, for the second one, consider the intersection $\delta(u,v) \cap f(u)$, on it we have $u = v$ and $u = s$, then it is included in $\delta(u,v) \cap f(v)$. If $V \subset U$ and $s_V = s_U|V$, the open covering of V defined by s_V is the trace of the open covering defined by s_U. Moreover, a morphism $\phi : X \to Y$ in $\mathcal{E}$ sends sub-singletons to sub-singletons and induces injections of the maximal domain of extension; therefore the above construction defines a natural transformation $\eta_{\mathcal{E}}$ from $Id_{\mathcal{E}}$ to $F \circ G$.
This transformation is invertible: take an element f of $X^{\mathbf{\Omega}}(U)$, and for every $U \in \Omega$, consider the set $S(f,U)$ of sub-singletons u of X such that $f_U(u) \neq \emptyset$. If u and v belong to this set, the first inequality of (28) implies that $u = v$ on the intersection $f_U(u) \cap f_U(v)$, then, by the sheaf property 3, $S(f,U)$ defines a unique element $u_U \in X(U)$.

In the other direction, the composition $G \circ F$ associates to a Ω-set (X,δ) the Ω-set $(X^{\mathbf{\Omega}}, \delta_{X,\mathbf{\Omega}})$ made by the sub-singletons of the presheaf X_Ω, i.e. the families (f,U) of compatible coverings $f_V(v), v \in X$ of $V; V \subset U$. We have $\delta((f,U),(f,U)) = U$; therefore, for simplifying the notations, we denote the singleton by f, and U is $\delta(f,f)$.
We saw that, for two elements f, $(f'$, the open set $\delta(f,f')$ is the maximal open subset of $U \cap U'$ where the coverings $f_V(u)$ and $f'_V(u)$ coincide for every $u \in X$ and $V \subset U$.
For a pair (u,f), of $u \in X$ and $(f \in X^{\mathbf{\Omega}}$, we define $H(u,f) \in \Omega$ as the unions of the open sets $f_V(u)$, over $V \subset \delta(f,f) \cap \delta(u,u)$.
The formula (27) implies that $H(u,f)$ is also the union of open sets α such that $\alpha \subset f_\alpha(u)$, i.e. $f_\alpha(u) = \alpha$.
We verify that H is a morphism of Ω-sets: the first axiom

$$\delta(u,v) \wedge H(u,f) \leq H(v,f) \tag{29}$$

results from

$$\delta(u,v) \wedge f_\alpha(u) \leq f_\alpha(v) \tag{30}$$

for every $\alpha \in \Omega$.
The second axiom

$$H(u,f) \wedge \delta(f,f') \leq H(u,f') \tag{31}$$

comes from the definition of $\delta(f,f')$ as an open set where the induced coverings coincide.

For the third axiom,

$$H(u,f) \wedge H(u,f') \leq \delta(f,f'); \tag{32}$$

if α is included in the intersection we have $f_\alpha(u) = \alpha = f'_\alpha(u)$, then $\alpha \leq \delta(f,f')$.
From (28), we have $f_\alpha(u) \subset \delta(u,u)$, then

$$H(u,f) \subset \delta(u,u) \tag{33}$$

And for every $\alpha \leq \delta(u,u)$, we can define a special covering f^u_α by

$$f^u_\alpha(u) = \alpha, \quad f^u_\alpha(v) = \alpha \wedge \delta(u,v); \tag{34}$$

it satisfies (28). Then

$$\delta(u,u) = \bigvee_{f \in X(U)} H(u,f) \tag{35}$$

The Ω-map H is natural in $X \in \mathsf{Set}_\Omega$. To terminate the proof of proposition A.2, we have to show that H is invertible, that is to find a Ω-map $H' : X^{\mathbf{\Omega}}_\Omega \times X \to \Omega$, such that $H' \circ H = \delta_X$ and $H \circ H' = \delta_{X_\Omega,\mathbf{\Omega}}$. We note the first fuzzy identity by δ and the second one by δ'.

In fact $H'(f,u) = H(u,f)$ works; in other terms H is an involution of Ω-sets. let us verify this fact: by definition of the composition

$$H' \circ H(u,v) = \bigvee_f H(u,f) \wedge H'(f,v) \tag{36}$$

is the reunion of the $\alpha \in \Omega$ such that there exists f with $\alpha = f_\alpha(u) = f_\alpha(v)$, then by the first inequality in (28) it is included in $\delta(u,v)$. Now consider $\alpha \leq \delta(u,v) \subseteq \delta(u,u)$, and define a covering of α by $f^u_\alpha(w) = \alpha \cap \delta(u,w)$ for any $w \in X$, this gives $\alpha \leq f^u_\alpha(v)$ then $\alpha \subseteq H(v,f^u)$, then $\alpha \subset H(u,f^u) \wedge H'(f^u,v)$.
On the other side,

$$H \circ H'(g,f) = \bigvee_u H(g,u) \wedge H(u,f), \tag{37}$$

is the reunion of the $\alpha \in \Omega$ such that there exists u with $\alpha = f_\alpha(u) = g_\alpha(u)$. In this case, we consider the set $S(f,\alpha)$ of elements $v \in X$ such that $f_\alpha(v) \neq \emptyset$. If v and w belong to this set, the first inequality of (28) implies that $v = w$ on the intersection $f_\alpha(v) \cap f_z(w)$, then, by the sheaf property, $S(f,\alpha)$ defines a unique element $u_\alpha \in X$. This element must be equal to u. The same thing being true for g, this implies that $f_\alpha(v) = g_\alpha(v)$ for all the elements v of X, some of them giving α the other giving the empty set. Consequently, $H \circ H'(g,f) \subseteq \delta'(f,g)$.
The other inclusion $\delta'(f,g) \subseteq H \circ H'(g,f)$ being obvious, this terminates the proof of the proposition. ■

This proposition generalizes to the localic Grothendieck topos the construction of the sheaf space (espace étalé in French) associated to a usual topological sheaf. However the accent in Ω-sets is put more on the gluing of sections than on a well defined set of germs of sections, as in the sheaf space. In some sense, the more general Ω-sets give also a more global approach, as in the original case of Riemann surfaces. Replacing a dynamics for instance by its solutions, pairs of domains and functions on them, with the relation of prolongation over sub-domains. This seems to be well adapted to the understanding of a DNN, on sub-trees of its architectural graph $\mathbf{\Gamma}$.

The localic Grothendieck topos $\mathcal{E}_\Omega$ are the "elementary topos" which are sub-extensional (generated by sub-singletons) and defined over Set [Bel08, p. 207].
Particular cases are characterized by special properties of the lattice structure of the locale Ω [Bel08, pp. 208-210]:

- we say that two elements U, V in Ω are separated by another element $\alpha \in \Omega$ when one of them is smaller than α but not the other one.

$\mathcal{E}_\Omega$ is the topos of sheaves over a topological space $\mathbf{X}$ if and only if Ω is *spatial*, which means by definition, that any pair of elements of Ω is separated by a *large* element, i.e. an element α such that $\beta \wedge \gamma \leq \alpha$ implies $\beta \leq \alpha$ or $\gamma \leq \alpha$.
Moreover, in this case, Ω is the poset of open sets of $\mathbf{X}$, and the large elements are the complement of the closures of points of $\mathbf{X}$.
The topological space is not unique, only the *sober* quotient is unique. A topological space is sober when every irreducible closed set is the closure of one and only one point.

$\mathcal{E}_\Omega$ is the topos of presheaves over a poset $C_\mathbf{X}$ if and only if Ω is an Alexandrov lattice, i.e. any pair of elements of Ω is separated by a *huge* (very large) element, i.e. an element α such that $\bigwedge_{i\in I} \beta_i \leq \alpha$ implies that $\exists i \in I, \beta_i \leq \alpha$.
In this case Ω is the set of lower open sets for the Alexandrov topology on the poset.
If Ω is finite, large and huge coincide, then spatial is the same as Alexandrov.

B Topos of DNNs and spectra of commutative rings

A finite poset with the Alexandrov topology is sober. This is a particular case of Scott's topology. Then it is also a particular case of spectral spaces [Hoc69], [Pri94], that are (prime) spectra of a commutative ring with the Zariski topology.

From the point of view of spectrum, a tree in the direction described in theorem 1.2, corresponds to a ring with a unique maximal ideal, i.e., by definition a local ring.
The minimal points correspond to minimal primes. The gluing of two posets along an ending vertex corresponds to the fiber product of the two rings over the simple ring with only one prime ideal [Ted16].
A ring with a unique prime ideal is a field, in this case the maximal ideal is $\{0\}$. This gives the following result:

Proposition B.1. *The canonical (i.e. sober) topological space of a DNN is the Zariski spectrum of a commutative ring which is the fiber product of a finite set of local rings over a product of fields.*

The construction of a local rings for a given finite poset can be made by recurrence over the number of primes, by successive application of two operations: gluing a poset along an open subset of another poset, and joining several maximal points; this method is due to Lewis 1973 [Ted16].

Examples. I − The topos of Shadoks [Pro08] corresponds to the poset $\beta < \alpha$ with two points; this is the spectrum of any *discrete valuation ring* only containing the ideal $\{0\}$ and a non-zero maximal ideal. Such a ring is the subset of a commutative field $\mathbb{K}$ with a valuation v valued in $\mathbb{Z}$, defined by $\{a \in \mathbb{K} | v(a) \geq 0\}$. An example is $\mathbb{K}((x))$ the field of fractions of the formal series $\mathbb{K}[[x]]$, with the valuation given by the smallest power of x (and ∞) for $a = 0$. The valuation ring is $\mathbb{K}[[x]]$, also noted $K\{x\}$, its maximal ideal is $\mathfrak{m}_x = x\mathbb{K}[[x]]$.

II − Consider the poset of length three: $\gamma < \beta < \alpha$. Apply the gluing construction to the ring $A = \mathbb{K}\{x\}$ embedded in $\mathbb{K}((x))$ and the ring $B = \mathbb{K}((x))\{y\}$ projecting to $\mathbb{K}((x))$; this gives the following local ring:

$$D = \mathbb{K}\{x\} \times_{\mathbb{K}((x))} \mathbb{K}((x))\{y\} \cong \{d = a + yb | a \in A, b \in B\} \subset B. \tag{38}$$

The sequence of prime ideals is

$$\{0\} \subset yB \subset \mathfrak{m}_x + yB. \tag{39}$$

III − Continuing this process, we get a natural local ring which spectral space is the chain of length $n+1$, $\alpha_n < \ldots < \alpha_0$ or simplest DNNs. There is one such ring for any commutative field $\mathbb{K}$:

$$\begin{aligned} D_n = \{d = a_n + x_{n-1}b_{n-1} + \ldots + x_1 b_1 \in \mathbb{K}((x_1, x_2, \ldots, x_n))| \\ a_n \in \mathbb{K}\{x_n\}, b_{n-1} \in \mathbb{K}((x_n))\{x_{n-1}\}, \ldots, b_1 \in \mathbb{K}((x_2, \ldots, x_n))\{x_1\}. \end{aligned} \tag{40}$$

The sequence of prime ideals is

$$\begin{aligned} \{0\} \subset x_1\mathbb{K}((x_2, \ldots, x_n))\{x_1\} \subset \\ x_1\mathbb{K}((x_2, \ldots, x_n))\{x_1\} + x_2\mathbb{K}((x_3, \ldots, x_n))\{x_2\} \subset \\ \ldots \subset x_1\mathbb{K}((x_2, \ldots, x_n))\{x_1\} + \ldots + x_n\mathbb{K}\{x_n\}. \end{aligned} \tag{41}$$

C Classifying objects of groupoids

Proposition C.1. *There exists an equivalence of category between any connected groupoid $\mathcal{G}$ and its fundamental group G.*

Proof. let us choose an object O in $\mathcal{G}$, the group G is represented by the group G_O of automorphisms of O. The inclusion gives a natural functor $J : G \to \mathcal{G}$ which is full and faithful. In the other direction, we choose for any object x of $\mathcal{G}$, a morphism (path) γ_x from x to O, we choose $\gamma_O = id_O$, and we define a functor R from $\mathcal{G}$ to G by sending any object to O and any arrow $\gamma : x \to y$ to the endomorphism $\gamma_y \circ \gamma \circ \gamma_x^{-1}$ of O. The rule of composition follows by cancellation. A natural isomorphism between $R \circ J$ and Id_G is the identity. A natural transformation T from $J \circ R$ to $Id_{\mathcal{G}}$ is given by sending $x \in \mathcal{G}$ to γ_x, which is invertible for each x. The fact that it is natural results from the definition of R: for every morphism $\gamma : x \to y$, we have

$$T(y) \circ Id(\gamma) = \gamma_y \circ \gamma = (\gamma_y \circ \gamma) \circ \gamma_x^{-1} \circ \gamma_x = JR(\gamma) \circ T(x). \tag{42}$$

What is not natural in general (except if $\mathcal{G} = G = \{1\}$) is the choice of R. This makes groupoids strictly richer than groups, but not from the point of view of homotopy equivalence. Every functor between two groupoids that induces an isomorphism of π_0, the set of connected components, and of π_1, the fundamental group, is an equivalence of category. ■

One manner to present the topos $\mathcal{E} = \mathcal{E}_{\mathcal{G}}$ of presheaves over a small groupoid $\mathcal{G}$ (up to category equivalence) is to decompose $\mathcal{G}$ in connected components $\mathcal{G}_a; a \in A$, then $\mathcal{E}$ will be product of the topos $\mathcal{E}_a; a \in A$ of presheaves over each component. For each $a \in A$, the topos $\mathcal{E}_a$ is the category of G_a-sets, where G_a denotes the group of auto-morphisms of any object in $\mathcal{G}_a$.
The classifying object $\Omega = \Omega_{\mathcal{G}}$ is the boolean algebra of the subsets of A.

In the applications, we are frequently interested by the subobjects of a fixed object $X = \{X_a; a \in A\}$. The algebra of subobjects Ω^X, has for elements all the subsets that are preserved by G_a for each component $a \in A$ independently.
Thus we can consider what happens for a given a. Every element $Y_a \in \Omega^{X_a}$ has a complement $Y_a^c = \neg Y_a$, which is also invariant by G_a, and we have $\neg\neg = Id$. Here the relation of negation $\leq$ is the set-theoretic one. It is also true for the operations $\wedge$ (intersection of sets), $\vee$ (union of sets), and the internal implication $p \Rightarrow q$, which is defined in this case by $(p \wedge q) \vee \neg p$.

All the elements Y_a of Ω^{X_a} are reunions of orbits $Z_i; i \in K(X_a)$ of the group G_a in the G_a-set X_a. On each orbit, G_a acts transitively.
Each subobject of X is a product of subobjects of the X_a for $a \in A$. The product over a of the $K(X_a)$ is a set $K = K(X)$.

The algebra Ω^X is the Boolean algebra of the subsets of the set of elements $\{Z_i; i \in K\}$, that we can note simply Ω_K.

The arrows in this category, $p \to q$, correspond to the pre-order $\leq$, or equivalently to the inclusion of sets, and can be understood as implication of propositions. This is the implication in the external sense, if p is true then q is true, not in the internal sense q^p, also denoted $p \Rightarrow q$, that is also the maximal element x such that $x \wedge p \leq q$).
On this category, there exists a natural Grothendieck topology, named the canonical topology, which is the largest (or the finest) Grothendieck topology such that, for any $p \in \Omega$, the presheaf $x \mapsto \mathsf{Hom}(x, p)$ is a sheaf. For any $p \in \Omega$, the set of coverings $J_K(p)$ is the set of collections of subsets q of p whose

reunion is p. In particular $J_K(\emptyset)$ contains the empty family; this is a singleton.

Proposition C.2. *The topos $\mathcal{E}$ is isomorphic to the topos* $\mathsf{Sh}(\Omega;K)$ *of sheaves for this topology J_K (see for instance* Bell, Toposes and local set theories [Bel08]*).*

Proof. For all p, any covering of p has for refinement the covering made by the disjoint singletons Z_i that belong to p, seen as a set; then, for every sheaf F over Ω, the restriction maps give a canonical isomorphism from $F(p)$ with the product of the sets $F(Z_i)$ over p itself.
In particular, any sheaf has for value in $\perp = \emptyset$ a singleton. ∎

D Non-Boolean information functions

This is the case of chains and injective presheaves on them.

The site S_n is the poset $0 \to 1 \to \ldots \to n$. A finite object E is chosen in the topos of presheaves $S_n^\wedge$, such that each map $E_i \to E_{i-1}$ is an injection, and we consider the Heyting algebra Ω^E, that is made by the subobjects of E. The inclusion, the intersection and the union of subobjects are evident. The only non-trivial internal operations are the exponential, or internal implication $Q \Rightarrow T$, and the negation $\neg Q$, that is a particular case $Q \Rightarrow \emptyset$.

Lemma D.1. *Let $T_n \subset T_{n-1} \subset \ldots \subset T_0$ and $Q_n \subset Q_{n-1} \subset \ldots \subset Q_0$ be two elements of Ω^E, then the implication $U = (Q \Rightarrow T)$ is inductively defined by the following formulas:*

$$
\begin{aligned}
&U_0 = T_0 \vee (E_0 \backslash Q_0),\\
&U_1 = U_0 \wedge (T_1 \vee (E_1 \backslash Q_1),\\
&\ldots\\
&U_k = U_{k-1} \wedge (T_k \vee (E_k \backslash Q_k),\\
&\ldots
\end{aligned}
$$

Proof. By recurrence. For $n = 0$ this is the well known boolean formula. Let us assume the result for $n = N - 1$, and prove it for $n = N$. The set U_N must belong to U_{N-1} and must be the union of all the sets $V \subset E_N \cap U_{N-1}$ such that $V \wedge Q_N \subset T_N$, then it is the union of $T_N \cap U_{N-1}$ and $(E_N \backslash Q_N) \cap U_{N-1}$.
In particular the complement $\neg Q$ is made by the sequence

$$\bigcap_{k=0}^{n}(E_k \backslash Q_k) \subset \bigcap_{k=0}^{n-1}(E_k \backslash Q_k) \subset \ldots \subset E_0 \backslash Q_0. \tag{43}$$

∎

Definition D.1. *We choose freely a strictly positive function μ on E_0; for any subset F of E_0, we note $\mu(F)$ the sum of the numbers $\mu(x)$ for $x \in F$.*
In practice μ is the constant function equal to 1*, or to* $|F|^{-1}$.

Definition D.2. *Consider a strictly decreasing sequence $[\delta]$ of strictly positive real numbers $\delta_0 > \delta_1 > ... > \delta_n$; the function $\psi_\delta : \Omega^E \to \mathbb{R}$ is defined by the formula*

$$\psi_\delta(T_n \subset T_{n-1} \subset ... \subset T_0) = \Sigma_{k=0}^{n} \delta_k \mu(T_k). \tag{44}$$

Lemma D.2. *The function ψ_δ is strictly increasing.*

This is because index by index, T'_k contains T_k.

Definition D.3. *A function $\varphi : \Omega^E \to \mathbb{R}$ is* concave *(*resp. *strictly concave), if for any pair of subsets $T \leq T'$ and any proposition Q, the following expression is positive (*resp. *strictly positive),*

$$\Delta\varphi(Q;T,T') = \varphi(Q \Rightarrow T) - \varphi(T) - \varphi(Q \Rightarrow T') + \varphi(T'). \tag{45}$$

Hypothesis on δ: for each k, $n \geq k \geq 0$, we assume that $\delta_k > \delta_{k+1} + ... + \delta_n$.
This hypothesis is satisfied for instance for $\delta_0 = 1, \delta_1 = 1/2, ..., \delta_k = 1/2^k, ...$

Proposition D.1. *Under this hypothesis, the function ψ_δ is concave.*

Proof. Let $T \leq T'$ in Ω^E. We define inductively an increasing sequence $T^{(k)}$ of S_n-sets by taking $T^{(0)} = T$ and, for $k > 0$, $T^{(k)}_j$ equal to $T^{(k-1)}_j$ for $j < k$ or $j > k$, but equal to T'_j for $j = k$. In other terms, the sequence is formed by enlarging T_k to T'_k, index after index. Let us prove that $\Delta\psi_\delta(Q;T^{(k-1)},T^{(k)})$ is positive, and strictly positive when at the index k, T_k is strictly included in T'_k. The theorem follows by telescopic cancellations.
The only difference between $T^{(k-1)}$ and $T^{(k)}$ is the enlargement of T_k to T'_k, and this generates a difference between $T^{(k-1)}_j|Q$ and $T^{(k)}_j|Q$ only for the indices $j > k$. This allows us to simplify the notations by assuming $k = 0$.
The contribution of the index 0 to the double difference $\Delta\psi_\delta$ is the difference between the sum of $\delta_0\mu$ over the points in $E_0\backslash Q_0$ that do not belong to T_0 and the sum of $\delta_0\mu$ over the points in $E_0\backslash Q_0$ that do not belong to T'_0, then it is the sum of $\delta 0\mu$ over the points in $E_0\backslash Q_0$ that belong to $T'_0\backslash T_0$.
As in lemma D.1, let us write $U_0 = T_0 \vee (E_0\backslash Q_0)$ and $U'_0 = T'_0 \vee (E_0\backslash Q_0)$. And for $k \geq 1$, let us write $V_k = T_k \vee (E_k\backslash Q_k)$, and $W_k = V_1 \cap ... \cap V_k$.
From the lemma 1, the contribution of the index 1 to the double difference $\Delta\psi_\delta$, is the simple difference between the sum of $\delta_k\mu$ over the points in $U_0 \cap W_k$ and its sum over the points in $U'_0 \cap W_k$, then it is equal to the opposite of the sum of $\delta_k\mu$ over the points in $(T'_0\backslash T_0) \cap (E_0\backslash Q_0) \cap W_k$. The hypothesis on the sequence δ implies that the sum over k of these sums is smaller than the difference given by the index 0. ∎

Remark. In general the function ψ_δ, whatever being the sequence δ, is not strictly concave, because it can happen that T_0' is strictly larger than T_0, and the intersection of $T_0'\backslash T_0$ with $E_0\backslash Q_0$ is empty. Therefore, to get a strictly concave function, we take the logarithm, or another function from $\mathbb{R}_+^\star$ to $\mathbb{R}$ that transforms strictly positive strictly increasing concave functions to strictly increasing strictly concave functions. This property for the logarithm comes from the formulas

$$(\ln\varphi)" = [\frac{\varphi'}{\varphi}]' = \frac{\varphi\varphi" - (\varphi')^2}{\varphi^2} < 0. \tag{46}$$

In what follows we take $\psi = \ln\psi_\delta$ as the fundamental function of precision.
By normalizing μ and taking $\delta_0 = 1$, we get $0 < \psi_\delta \leq 1$, $-\infty < \psi \leq 0$.

Remark. Lemmas D.1, D.2 and proposition D.1 can easily be extended to the case where the basic site $\mathcal{S}$ is a rooted (inverse) tree, i.e. the poset that comes from an oriented graph with several initial vertices and a unique terminal vertex. The computation with intersections works in the same manner. The hypothesis on δ concerns only the descending branches to the terminal vertex.

Now, remember that the poset of a DNN is obtained by gluing such trees on some of their initial vertices, interpreted as tips (of forks) or output layers. The maximal points correspond to tanks (of forks) of input layers. Therefore it is natural to expect that the existence of ψ holds true for any site of a DNN.

E Closer to natural languages: linear semantic information

Several attempts were made by logicians and computer scientists, since Frege and Russel, Tarski and Carnap, to approach the properties of human natural languages by formal languages and processes. In particular, a computational grammar was proposed by Lambek [Lam58]: a syntactic category is defined with sentences as objects and applications of grammatical rules as arrows, a second category is defined, that contains products and exponentials, for instance a topos, and semantic is seen as some functor from the first category to the second one. This is the first place where semantic is defined as interpretations of types and propositions in a topos. Precursors of the kind of grammar considered by Lambek were Adjukiewicz in 1935 [Adj35] and Bar-Hillel in 1953 [BH53].

Then a decisive contribution was made by Montague in 1970, [Mon70], who developed in particular a formal treatment of pieces of English [Par75]. Also in this approach, semantics appears as a transformation from a syntactic algebraic structure, having lexis and multiple operations, to a coarser structure. In the nineties mathematicians and linguists observed that the categorical point of view, as in Lambek, gives a good framework for developing further Montague's theory [vB90].

The next step used intensional type theories, like Martin-Löf's theory [ML80], named modern TT by Luo [Luo14], or rich TT by Cooper et al. [CDLL15]. New types were introduced, corresponding to the many structural notions of linguistic, e.g. noun, verb, adjective, and so on. Also modalities like interrogative, performative, can be introduced (see Brunot [Bru36] for the complexity of the enterprise in

French). Recent experiment with programming languages have shown that many properties of languages can be captured by extending TT. For instance, in Martin-Löf TT it is possible to construct ZFT theories but also alternative Non-well-founded set theories, like in [Acz88], taking into account paradoxical vicious circles as natural languages do [Lin89]. Even more powerful is the homotopical type theory (HoTT) of Voevodski, Awodey, Kapulkin, Lumsdaine, Shulman, ..., [KLV12]. Also see Gylterud and Bonnevier [GB20] for the inclusion of non-well-founded sets theories.

These formal theories do not give a true definition of what is *meaning*, (see the fundamental objections of Austin [Aus61]), but they give an insight of the various ways the meanings can be combined and how they are related to grammar, compatible with the intuition we have of human interpretations. We do not suggest that the categorial presentation defines the natural languages, but here also we think that its capture something of toys languages, an some languages games that can help the understanding of semantic functioning in networks, including properties of natural semantics of human peoples.

In what follows, we consider that a given category $\mathcal{A}$ represents the semantic for a given language, or some language game [Wit53], and reflects properties of a language, not the abstract rules, as in the algebra $\Omega^{\mathbb{L}}$ before. The objects of $\mathcal{A}$ represent interpretations of sentences, or images, corresponding to the "types as propositions" (Curry-Howard) in a given grammar, and its arrows represent the evocations, significations, or deductions, corresponding to proofs or application of rules in grammar. Oriented cycles are *a priori* admitted.

We simply assume that $\mathcal{A}$ is a *closed monoidal category* [EK66] that connects with linear logic and linear type theory as in Mellies, "Categorical Semantics of Linear Logic" [Mel09].
In such a category, a bifunctor $(X,Y) \mapsto X \otimes Y$ is given, that is associative up to natural transformation, with a neutral element $\star$ also up to linear transformation, satisfying conditions of coherence. This product representing aggregation of sentences. Moreover there exists classifiers objects of morphisms, i.e. objects A^Y defined for any pair of objects A, Y, such that for any X, there exist natural isomorphisms

$$\mathsf{Hom}(X \otimes Y, A) \simeq \mathsf{Hom}(X, A^Y). \tag{47}$$

The functor $X \mapsto X \otimes Y$ has for right-adjoint the functor $A \mapsto A^Y$.

For us, this defines the semantic conditioning, the effect on the interpretation A that Y is taken into account, when A is evoked by a composition with Y. Thus we also denote A^Y by $Y \Rightarrow A$ or $A|Y$.

When A is given, and if $Y' \to Y$ we get $A|Y \to A|Y'$.
From $X \otimes \star \cong X$, it follows that canonically $A^\star \cong A$. We make the supplementary hypothesis that $\star$ is a final object, then we get a canonical arrow $A \to A|Y$, for any object Y. This represents the internal constants.

Remark. In the product $X \otimes Y$, the ordering plays a role, and in linguistic, in the spirit of Montague, two functors can appear, the one we just said $Y \mapsto X \otimes Y$ and the other one $X \mapsto X \otimes Y$. If both have a left adjoint, we get two exponentials: $A^Y = A|Y$ and ${}^X A = X\ A$; the natural axiomatic becomes the *bi-closed category* of Eilenberg and Kelly [EK66]. Dougherty [Dou92] gave a clear exposition of part of the Lambek calculus in the Montague grammar in terms of this structure (same in [Lam88]). A theory of semantic information should benefit of this possibility, where composition depends on the ordering, but in what follows, to begin, we assume that $\mathcal{A}$ is *symmetric*: there exist natural isomorphisms exchanging the two factors of the product.

All that can be localized in a context $\Gamma \in \mathcal{A}$ by considering the category $\Gamma\backslash\mathcal{A}$ of morphisms $\Gamma \to A$, where A describes $\mathcal{A}$, with morphisms given by the commutative triangles. For $\Gamma \to A$, and $Y \in \mathcal{A}$, we get a morphism $\Gamma \to A|Y$ by composition with the canonical morphism $A \to A|Y$. This extends the conditioning. We will discuss the existence of a restricted tensor product later on; it asks restrictions on Γ.

The analog of a theory, that we will also name theory here, is a collection S of propositions A, that is stable by morphisms to the right, i.e. $A \in S$ and $A \to B$ implies $B \in S$. This can be seen as the consequences of a discourse. A theory S' is said weaker than a theory S if it is contained in it, noted $S \leq S'$. Then the analog of the conditioning of S by Y is the collection of the objects A^Y for A in S. The collection of theories is partially ordered.
We have $S|Y' \leq S|Y$ when there exists $Y' \to Y$. In particular $S|Y \leq S$, as it was the case in simple type theory.

When a context is given, it defines restricted theories, because it introduces a constraint of commutativity for $A \to B$, to define a morphism from $\Gamma \to A$ to $\Gamma \to B$.

The monoidal category $\mathcal{A}$ acts on the set of functions from the theories to a fixed commutative group, for instance the real numbers.
We will later discuss how the context Γ can be included in a category generalizing the category $\mathcal{D}$ of sections 3.4 and 3.5, to obtain the analog of the classical ordinary logical case with the propositions P excluded. This needs a notion of negation, which, we will see, are many.

Remark. The model should be more complete if we introduce a syntactic type theory, as in Montague 1970, such that $\mathcal{A}$ is an interpretation of part of the types, compatible with products and exponentials. Then some of the arrows can interpret transformation rules in the grammar. The introduction of syntaxes will be necessary for communication between networks.

Let us use the notations of chapter 2. Between two layers $\alpha : U \to U'$ lifted by h to $\mathcal{F}$, we assume the existence of a functor $\pi_\star\alpha, h$ from $\mathcal{A}_{U,\xi}$ to $\mathcal{A}_{U',\xi'}$, with a left adjoint $\pi^\star_{\alpha,h}$, such that $\pi^\star\pi_\star = \mathsf{Id}$, in such a manner that $\mathcal{A}$ becomes a pre-cosheaf over $\mathcal{F}$ for $\pi_\star$ and the sets of theories Θ form a presheaf for $\pi^\star$.

The information quantities are defined as before, by the natural bar-complex associated to the action of $\mathcal{A}$ on the pre-cosheaf Φ' of functions on the functor Θ.

The passage to a network gives a dynamic to the semantic, and the consideration of weights gives a model of learning semantic. Even if they are caricature of the natural ones, we hope this will help to capture some interesting aspects of them.

A big difference with the ordinary logical case, is the absence of "false", then in general, the absence of the negation operation. This can make the cohomology of information non-trivial.

Another big difference is that the category $\mathcal{A}$ is not supposed to be a poset, the sets Hom can be more complex than $\emptyset$ and $\star$, and they can contain isomorphisms. In particular loops can be present.

Consider for instance any function ψ on the collection of theories; and suppose that there exist arrows from A to B and from B to A; then the function ψ must take the same value on the theories generated by A and B. This tells in particular that they contain the same information.

The homotopy construction of a bi-simplicial set $g\Theta$ can be made as before, representing the propagation feed-forward of theories and propagation backward of the propositions, and the information can be defined by a natural increasing and concave map F with values in a closed model category $\mathcal{M}$ of Quillen (see chapter 2).

The semantic functioning becomes a simplicial map $gS : g\mathbb{X} \to g\Theta$, and the semantic spaces are given by the composition $F \circ gS$.

Here is another interest of this generalization: we can assume that a measure of complexity K is attributed to the objects, seen as expressions in a language, and that this complexity is additive in the product, i.e. $K(X \otimes Y) = K(X) + K(Y)$, and related to the combinatorics of the syntax, and the complexity of the lexicon, and the grammatical rules of formation. In this framework, we could compare the values of K in the category, and define the *compression* as the ratio F/K of information by complexity.

Remark. It is amazing and happy that the bar-complex for the information cocycles and the homotopy limit, can also be defined for the bi-closed generalization. The two exponentials ${}^{X}A$ and A^{Y} an action of the monoid $\mathcal{A}$ to the right and to the left that commute on the functions of theories, and on the bi-simplicial set $g\Theta$. Then we can apply the work of MacLane, Beck on bi-modules and the work of Schulman on enriched categories.
Taking into account the network, we get a tri-simplicial set $\Theta_{\star}^{\bullet\bullet}$ of information elements, or tensors, giving rise to a bi-simplicial space of histories of theories, with multiple left and right conditioning, $gI^{\bullet\bullet}$, that is the geometrical analog of the bar-complex of semantic information.

Links with Linear Logic (intuitionist) and negations.

The generalized framework corresponds to a fragment of an intuitionist Linear Logic (see Bierman and de Paiva [BdP00], Mellies [Mel09]). The arrows $A \to B$ in the category are the expression of the assertions of consequence $A \vdash B$, and the product expresses the joint of the elements of the left members of consequences, in the sense that a deduction $A_1,...,A_n \vdash B$ corresponds to an arrow $A_1 \otimes ... \otimes A_n \to B$. There is no necessarily a "or" for the right side, but there is an internal implication $A \multimap B$ which satisfies all the axioms of the above implication $A \Rightarrow B$, right adjoint of the tensor product. The existence of the final element corresponds to the existence of (multiplicative) truth $\mathbf{1} = \star$. To be more complete, we should suppose that all the finite products exist in the category $\mathcal{A}$. Then the (categorial) product of two corresponds to an additive disjunction $\oplus$, then a "or", that can generate the right side of sequents $B_1,...,B_m$ in $A_1,...,A_n/B_1,...,B_m$; however, a neutral element for $\oplus$ could be absent, even if it is always present in the full theory of Girard [Gir87]. No right adjoint is required for $\oplus$. And in what follows we do not assume the data $\oplus$.

One of the main ideas of [Gir87] was to incorporate the fact that in real life the proposition A that is used in a consequence $A \multimap B$ does not remain unchanged after the event, however it is important to give a special status for propositions that continue to hold after the event. For that purpose Girard introduced an operator on the formulas, named a *linear exponential*, and written !. It is named "of course" and has the meaning of a reaffirmation, something stable. The functor ! is required to be naturally equivalent to !!, then a projector in the sense of categories, such that, in a natural manner, the objects $!A$ and the morphisms $!f$ between them satisfy the Gentzen rules of weakening and contraction, respectively $(\Gamma \vdash \Delta)/(\Gamma, !A \vdash \Delta)$ and $(\Gamma, A, A \vdash \Delta)/(\Gamma, A \vdash \Delta)$. (This corresponds to the traditional assertions $A \wedge B \leq A$ and $A \leq A \wedge A$.) Further axioms state, when translated in categorical terms, that ! is a monoidal functor equipped with two natural transformations $\epsilon_A : !A \to A$ and $\delta_A : !A \to !!A$, that are monoidal transformations, satisfying the coherence rules of a comonad, and with natural transformations $e_A : !A \to 1$ (useful when 1 is not assumed final) and $d_A : !A \to !A \otimes !A$, that is a diagonal operator, also satisfying coherence axioms telling that each $!A$ is a commutative comonoid, and each $!f$ a morphism of commutative comonoid. From all these axioms, it is proved that under ! the monoidal product becomes a usual categorial product in the category $!\mathcal{A} := \mathcal{A}^!$,

$$!(A \otimes B) \cong !A \otimes !B \cong !(A \times B); \tag{48}$$

and the category $\mathcal{A}^!$, named the Kleisli category of $(\mathcal{A}, !)$, is cartesian closed. More precisely, under ! the multiplicative exponential becomes the usual exponential:

$$!(A \multimap B) \cong !B^{!A}. \tag{49}$$

Remind that a *comonad* in a category is a functor T of this category to itself, equipped with two natural transformations $T \to T \circ T$ and $\varepsilon : T \to \mathsf{Id}$, satisfying coassociativity and counity axioms. This the dual of a *monad*, $T \circ T \to T$ and $\mathsf{Id} \to T$, that is the generalization of monoids to categories. The functor ! is an example of comonad [Mac71].

The axioms of a closed symmetric monoidal category, plus the existence of finite products, plus the functor !, give the largest part of the Gentzen rules, as they were generalized by Jean-Yves Girard in 1987

[Gir87].

Proposition E.1. *The linear exponential ! allows to localize the product at a given proposition, in the sense that the slice category to the right $\Gamma|\mathcal{A}$ is closed by products of linear exponential objects as soon as Γ belongs to $\mathcal{A}^!$.*

Proof. If we restrict us to the arrows $!\Gamma \to Q$, then the product $!\Gamma \to Q \otimes Q'$ is obtained by composing the diagonal $d_{!\Gamma} : !\Gamma \to !\Gamma \otimes !\Gamma$ with the tensor product $!\Gamma \otimes !\Gamma \to Q \otimes Q'$.
Its right adjoint is given by $!\Gamma \to (Q \multimap R)$, obtained by composing $!\Gamma \to Q$ with the natural map $Q \to Q|R$. ∎

To localize the theories themselves at P, for instance at a $!\Gamma$, we used, in the Heyting case, a notion of negation. To exclude a given proposition was the only coherent choice from the point of view of information, and this was also in accord with the experiments of spontaneous logics in small networks [BBG21a].

In the initial work of Girard, negation was a fundamental operator, verifying the hypothesis of involution $\neg\neg = \mathsf{Id}$, thus giving a duality. That explains that the initial theory is considered as a classical Linear Logic; it generalizes the usual Boolean logic in another direction than intuitionism. In a linear intuitionist theory, the negation is not necessary, but it is also not forbidden, and axioms were discussed in the nineties.

We follow here the exposition of Paul-André Melliès in [Mel09] and of his article with Nicolas Tabareau [MT10]. The authors work directly in a monoidal category $\mathcal{A}$, without assuming that it is closed, and define negation as a functor $\neg : \mathcal{A} \to \mathcal{A}^{\mathrm{op}}$, such that the opposite functor $\neg^{\mathrm{op}}$ from $\mathcal{A}^{\mathrm{op}}$ to $\mathcal{A}$, also denoted by $\neg$, is the left-adjoint of $\neg$, giving a unit $\eta : \mathsf{Id} \to \neg\neg$ and a counit $\epsilon : \neg\neg \to \mathsf{Id}$, that are not equivalence in general. Then there exist for any objects A, B a canonical bijection bijection between $\mathsf{Hom}_{\mathcal{A}}(\neg A, B)$ and $\mathsf{Hom}_{\mathcal{A}}(\neg B, A)$. Note that in this case ε and η coincide, because the morphisms in $\mathcal{A}^{\mathrm{op}}$ are the morphisms in $\mathcal{A}$ written in the reverse order.
The double negation $T = \neg^{\mathrm{op}}\neg$ forms a monad whose η is the unit; the multiplication $\mu : \neg\neg\neg\neg \to \neg\neg$ is obtained by composing $\mathsf{Id}_{\neg}$ with $\neg(\eta)$, to the left or to the right, that is $\mu_A = \neg(\eta_A) \circ \mathsf{Id}_{\neg A} = \mathsf{Id}_{\neg\neg\neg A} \circ \neg(\eta_A)$.
In theoretical computer science, T is called the *continuation monad*, and plays an important role in computation and games logics as in the works of Kock, Moggi, Mellies, Tabareau.

In the case of the Heyting algebra of a topos (elementary), this continuation defines a topology, named after Lawvere and Tierney, which defines the unique subtopos that is Boolean and dense (i.e. contains the initial object $\emptyset$ [Car12]).

The second important axiom tells how the (multiplicative) product $\otimes$ is transformed : it is required that for any objects B, C the object $\neg(B \otimes C)$ represents the functor $A \mapsto \mathsf{Hom}(A \otimes B, \neg C) \cong \mathsf{Hom}(C, \neg(A \otimes B)$;

that is

$$\mathsf{Hom}(A \otimes B, \neg C) \cong \mathsf{Hom}(A, \neg(B \otimes C)). \tag{50}$$

This bijection being natural in the three argument and coherent with the associativity and unit for the product $\otimes$.
For instance all the sets $\mathsf{Hom}(ABC, \neg D)$, $\mathsf{Hom}(AB, \neg(CD)$, $\mathsf{Hom}(A, \neg(BCD))$, are identified with $\mathsf{Hom}(ABCD, \neg\mathbf{1})$.
Mellies and Tabareau [MT10] called such a structure a *tensorial negation*, and named the monoidal category $\mathcal{A}$, equipped with $\neg$, a *dialogue category*.

The special object $\neg\mathbf{1}$ is canonically associated to the chosen negation; it is named the *pole* and frequently denoted by $\perp$. It has no reason in general to be an initial object of $\mathcal{A}$.

A monoidal structure of (multiplicative) disjunction is deduced from the tensor product by duality:

$$A \wp B = \neg(\neg A \otimes \neg B). \tag{51}$$

Its neutral element is the pole of $\neg$.
This implies that the notion of "or" is parameterized by the variety of negations, that we will see equivalent to $\mathcal{A}$ itself.

In the same manner an additive conjonction is defined by

$$A \& B = \neg(\neg A \oplus \neg B). \tag{52}$$

Its neutral element is $\top = \neg\emptyset$, when an initial element $\emptyset$ exists, that is the additive "false".

An operator ? was introduced by Girard in classical linear logic, that satisfies

$$?\neg A = \neg !A, \quad \neg ?A = !\neg A \tag{53}$$

For us, just these relations are not sufficient to define it, because $\neg$ is not a bijection.

The Girard operator ? means "why not?", as the operator ! means "of course"; they are examples of modalities, and correspond to the modalities more frequently denoted $\square$ and $\diamond$ in modal logics.

However, Hasegawa [Has03], Moggi [Mog91], Mellies and Tabareau [MT10] have remarked that more convenient tensorial negations must satisfy a further axiom. Note that this story started with Kock [Koc70] inspired by Eilenberg and Kelly [EK66].

Lemma E.1. *From the second axiom of a tensorial negation it results two natural transformations*

$$\neg\neg A \otimes B \to \neg\neg(A \otimes B); \tag{54}$$

$$A \otimes \neg\neg B \to \neg\neg(A \otimes B). \tag{55}$$

A monad where such maps exist in a monoidal category, is named a strong monad *[Koc70] and [Mog91]. The first transformation is named the* strength *of the monad* $T = \neg\neg$*, the second one its* costrength.

Proof. Let us start with the Identity morphism of $\neg(A \otimes B)$; by the axiom, it can be interpreted as a morphism $B \otimes \neg(A \otimes B) \to \neg A$, then applying the functor $\neg$, we get a morphism

$$\neg\neg A \to \neg[B \otimes \neg(A \otimes B)]; \tag{56}$$

then, applying the axiom again, we obtain a natural transformation

$$\neg\neg A \otimes B \to \neg\neg(A \otimes B). \tag{57}$$

Exchanging the roles of A and B gives the other transformation.
Said in other terms, we have natural bijections given by the tensorial axiom, applied two times,

$$\begin{aligned}\mathsf{Hom}(\neg(A \otimes B), \neg(A \otimes B)) &\cong \mathsf{Hom}(\neg(A \otimes B) \otimes B, \neg A)\\ &\cong \mathsf{Hom}(A, \neg[B \otimes \neg(A \otimes B)] \cong \mathsf{Hom}(A \otimes B, \neg\neg(A \otimes B));\end{aligned} \tag{58}$$

and also natural bijections, obtained in the same manner,

$$\begin{aligned}\mathsf{Hom}(\neg(A \otimes B), \neg(A \otimes B)) &\cong \mathsf{Hom}(\neg(A \otimes B) \otimes A, \neg B)\\ &\cong \mathsf{Hom}(B, \neg[A \otimes \neg(A \otimes B)] \cong \mathsf{Hom}(A \otimes B, \neg\neg(A \otimes B));\end{aligned} \tag{59}$$

The identity of $\neg(A \otimes B)$ in the first term gives a natural marked point, that is also identifiable with $\eta_{A\otimes B}$ in the last term.
On the set $\mathsf{Hom}((\neg(A \otimes B) \otimes B, \neg A)$ (*resp.* $\mathsf{Hom}(A \otimes \neg(A \otimes B), \neg B)$) we can apply the functor $\neg$; this gives a map to $\mathsf{Hom}(\neg\neg A, \neg[B \otimes \neg(A \otimes B)])$ (*resp.* $\mathsf{Hom}(\neg\neg B, \neg[A \otimes \neg(A \otimes B)])$), then the strength (*resp.* the costrength) after applying the second axiom. ∎

The strength and costrength taken together give two *a priori* different transformations $TA \otimes TB \to T(A \otimes B)$ (see *n lab cafe*, Kock, Moggi, Hazegawa).
The first one is the composition starting with the costrength of TA followed by the strength of B, then ending with the product:

$$TA \otimes TB \to T(TA \otimes B) \to TT(A \otimes B) \to T(A \otimes B); \tag{60}$$

the other one starts with the strength, then uses the costrength, and ends with the product

$$TA \otimes TB \to T(A \otimes TB) \to TT(A \otimes B) \to T(A \otimes B). \tag{61}$$

Then a third axiom was suggested by Kock in general for strong monads, and reconsidered by Hazegawa, Moggi, Mellies and Tabareau, it consist to require that these two morphisms coincide. This is named, since Kock, a *commutative monad*, or a *monoidal monad*. We will say that the negation itself is monoidal.

According to Mellies and Tabareau, Hasegawa observed that $T = \neg\neg$ is commutative, if and only if η gives an isomorphism $\neg \cong \neg\neg$ on the objects of $\neg\mathcal{A}$, if and only if μ gives an isomorphism on the objects of $\mathcal{A}$.

Proposition E.2. *A necessary and sufficient condition for having $\neg$ monoidal is that for each object A, the transformation $\eta_{\neg A}$ is an equivalence from $\neg A$ and $\neg\neg\neg A$ in the category $\mathcal{A}$.*

Corollary. *Define $\mathcal{A}^{\eta}$ as the collection of objects A' of $\mathcal{A}$, such that $\eta_{A'}$ is an isomporphism; in the commutative case, $\neg\mathcal{A}$ is a sub-category $\neg$ induces an equivalence of the full subcategory $\mathcal{A}^{\eta}$ of $\mathcal{A}$ with its opposite [Bel08, Proposition 1.31].*

Thus we recover most of the usual properties of negation, without having a notion of false.

Now assume that $\mathcal{A}$ is symmetric monoidal and closed; we get natural isomorphisms

$$\neg(A \otimes B) \approx A \Rightarrow \neg B \approx B \Rightarrow \neg A. \tag{62}$$

And using the neutral element $\mathbf{1} = \star$ for C, and denoting $\neg\mathbf{1}$ by P, we obtain that $\neg B = B \multimap P$.

Proposition E.3. *For any object $P \in \mathcal{A}$, the functor $A \mapsto (A \multimap P) = P|A$ is a tensor negation whose pole is P.*

Proof. First, this is a contravariant functor in A.
Secondly, for any pair A, B in $\mathcal{A}$, using the symmetry hypothesis, we get natural bijections

$$\mathsf{Hom}(B, A \multimap P) \cong \mathsf{Hom}(B \otimes A, P) \cong \mathsf{Hom}(A, B \multimap P). \tag{63}$$

This gives the basic adjunction.
Third, for any triple A, B, C in $\mathcal{A}$, the associativity gives

$$\mathsf{Hom}(A \otimes B, C \multimap P) \cong \mathsf{Hom}(A \otimes B \otimes C, P) \cong \mathsf{Hom}(A, (B \otimes C) \multimap P). \tag{64}$$

This gives the tensorial condition. ■

The transformation η is given by the Yoneda lemma, from the following natural map

$$\mathsf{Hom}(X, A) \to \mathsf{Hom}^{\mathrm{op}}(\neg X, \neg A) \cong \mathsf{Hom}(X, \neg\neg A). \tag{65}$$

There is no reason for asserting that this negation is commutative.
From proposition E.1, the necessary and sufficient condition is that, for any object A, the following map is an isomorphism

$$\eta_{A\Rightarrow P} : (A \Rightarrow P) \to (((A \Rightarrow P) \Rightarrow P) \Rightarrow P). \tag{66}$$

Even for $A = 1$ this is a non-trivial condition: $P \approx ((P \Rightarrow P) \Rightarrow P)$.
The fact that $1 \Rightarrow P \equiv P$ being obvious.

Choose an arbitrary object Δ and define $\neg Q$ as $Q \multimap \Delta$. This Δ will play the role of "false".
We say that a theory $\mathbb{T}$ excludes P if it contains $P \multimap \Delta$. This is equivalent to say that there exists R in $\mathbb{T}$ such that $R \to (P \multimap \Delta)$, i.e. $R \otimes P \to \Delta$, that is by symmetry: there exists $P \to (R \multimap \Delta)$. In particular, if $P \to R$, we obtain such a map by composition with $R \to (R \multimap \Delta)$.

To localize the action of the proposition at P, we have to prove the following lemma:

Lemma E.2. *Conditioning by Q such that $P \to P \otimes Q$ is non-empty, sends a theory $\mathbb{T}$ that excludes P into a theory $\mathbb{T}$ that also excludes P.*

Proof. From the hypothesis we have a morphism $\neg(P \times Q) \to \neg P$, but $\neg(P \times Q)$ is isomorphic to $Q \Rightarrow (P \Rightarrow \Delta) = (\neg P)|Q$. ■

This is analog to the statement of Proposition 3.2 in section 3.3, because in this case $P \leq Q$ is equivalent to $P = P \wedge Q$ and to $P \leq P \wedge Q$. The proof does not use that P is a linear exponential object.

Now assume that P belongs to the category $\mathcal{A}^!$, i.e. $P = !\Gamma$ for a given object $\Gamma \in \mathcal{A}$; we saw that the set $\mathcal{A}_P$ of Q such that $P \to Q$ forms a closed monoidal category, and by the above lemma, it acts on the set of theories excluding P. That is because $P \to Q$ implies $P \to P \otimes P \to P \otimes Q$

Therefore, all the ingredients of the information topology of chapter 2 are present in this situation.

Bibliography

[AB11] Samson Abramsky and Adam Brandenburger. The sheaf-theoretic structure of non-locality and contextuality. *New Journal of Physics*, 13(11):113036, 2011.

[Acz88] Peter Aczel. *Non-Well-Founded Sets*. Stanford University, Center for the Study of Language and Information, 1988.

[Adj35] Kazimierz Adjukiewicz. Die Syntaktische Konnexität. *Studia Philosophica*, 1:1–27, 1935.

[AGV63] Michael Artin, Alexander Grothendieck, and Jean-Louis Verdier. Théorie des topos et cohomologie étale des schémas. *SGA4, IHES*, 1963.

[AGZV12a] Vladimir Igorevitch Arnold, Sabir Medgidovich Gusein-Zade, and Alexander Nikolaevich Varchenko. *Singularities of Differentiable Maps, Volume 1: Classification of Critical Points, Caustics and Wave Fronts*. Modern Birkhäuser Classics. Birkhäuser Boston, 2012.

[AGZV12b] Vladimir Igorevitch Arnold, Sabir Medgidovich Gusein-Zade, and Alexander Nikolaevich Varchenko. *Singularities of Differentiable Maps, Volume 2: Monodromy and Asymptotics of Integrals*. Modern Birkhäuser Classics. Birkhäuser Boston, 2012.

[AK11] Peter Arndt and Krzysztof Kapulkin. Homotopy-theoretic models of type theory. In *Proceedings of the 10th International Conference on Typed Lambda Calculi and Applications*, TLCA'11, Berlin, Heidelberg, 2011. Springer-Verlag.

[ARDK16] Jacob Andreas, Marcus Rohrbach, Trevor Darrell, and Dan Klein. Learning to compose neural networks for question answering. *CoRR*, abs/1601.01705, 2016.

[Arg78] Jose Argemi. Approche qualitative d'un problème de perturbations singulières dans $\mathbb{R}^4$. Equadiff. 78, Conv. int. su equazioni differenziali ordinarie ed equazioni funzionali, Firenze 1978, 333-340, 1978.

[Arn73] Vladimir Igorevitch Arnold. Normal forms for functions near degenerate critical points, the Weyl groups of A_k, D_k, E_k and Lagrangian singularities. *Funct. Anal. Appl.*, 6:254–272, 1973.

[Aus61] John Langshow Austin. *Philosophical Papers*. Oxford University Press, 1961.

[AW09] Steve Awodey and Michael A. Warren. Homotopy theoretic models of identity types. *Mathematical Proceedings of the Cambridge Philosophical Society*, 146(1), Jan 2009.

[BB15] Pierre Baudot and Daniel Bennequin. The Homological Nature of Entropy. *Entropy*, pages 3253–3318, 2015.

[BB22] Jean-Claude Belfiore and Daniel Bennequin. A search of semantic spaces. Internal technical report, Huawei, 2022.

[BBCV21] Michael M. Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges, 2021.

[BBD+11] Jie Bao, Prithwish Basu, Mike Dean, Craig Partridge, Ananthram Swami, Will Leland, and James A. Hendler. Towards a theory of semantic communication. In *2011 IEEE Network Science Workshop*, pages 110–117, 2011.

[BBDH14] Prithwish Basu, Jie Bao, Mike Dean, and James A. Hendler. Preserving quality of information by using semantic relationships. *Pervasive and Mobile Computing*, 11:188 – 202, 2014.

[BBG20] Jean-Claude Belfiore, Daniel Bennequin, and Xavier Giraud. Logico-probabilistic information. Internal technical report, Huawei, 2020. Available upon request.

[BBG21a] Jean-Claude Belfiore, Daniel Bennequin, and Xavier Giraud. Logical Information Cells I. arXiv, 2021. 2108.04751.

[BBG21b] Jean-Claude Belfiore, Daniel Bennequin, and Xavier Giraud. Logical Information Cells, Part II. Internal technical report, Huawei, 2021. Available upon request.

[BCB16] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *ArXiv*, abs/1409.0473v7, 2016.

[BdP00] G. M. Bierman and V. C. V. de Paiva. On an intuitionistic modal logic. *Studia Logica: An International Journal for Symbolic Logic*, 65(3):383–416, 2000.

[Bel08] John L. Bell. *Toposes and Local Set Theories*. Dover, 2008.

[Ben86] Daniel Bennequin. Caustique mystique. In *Séminaire Bourbaki : volume 1984/85, exposés 633-650*, number 133-134 in Astérisque. Société mathématique de France, 1986. talk:634.

[BFL11] John C Baez, Tobias Fritz, and Tom Leinster. A characterization of entropy in terms of information loss. *Entropy*, 13(11):1945–1957, 2011.

[BH53] Yehoshua Bar-Hillel. A quasi-arithmetical notation for syntactic description. *Language*, 29(1):47–58, 1953.

[BHS+18] David Barrett, Felix Hill, Adam Santoro, Ari Morcos, and Timothy Lillicrap. Measuring abstract reasoning in neural networks. In Jennifer Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 511–520, Stockholmsmässan, Stockholm Sweden, 10–15 Jul 2018. PMLR.

[BK72] A. K. Bousfield and D. M. Kan. *Homotopy Limits, Completions and Localizations*. Springer, 1972.

[BPSPV20] Daniel Bennequin, Olivier Peltre, Grégoire Sergeant-Perthuis, and Juan Pablo Vigneaux. Extra-fine sheaves and interaction decompositions. `https://doi.org/10.48550/arXiv.2009.12646`, 2020.

[Bro73] Kenneth Brown. Abstract homotopy theory and generalized sheaf cohomology. *Transaction of the American Mathematical Society*, 186:419–458, 1973.

[Bru36] Ferdinand Brunot. *La pensée et la langue*. Masson et compagnie, 1936.

[BTBG19] Pierre Baudot, Monica Tapia, Daniel Bennequin, and Jean-Marc Goaillard. Topological information data analysis. *Entropy*, 21(9):869, Sep 2019.

[BW21] Roberto Bondesan and Max Welling. The hintons in your neural network: a quantum field theory view of deep learning, 2021.

[Car50] Rudolf Carnap. *Logical Foundations of Probability*. Chicago Press, University of Chicago, 1950.

[Car09] Olivia Caramello. *The Duality Between Grothendieck Toposes and Geometric Theories*. University of Cambridge, 2009.

[Car12] Olivia Caramello. Universal models and definability. *Mathematical Proceedings of the Cambridge Philosophical Society*, 152(2), 2012.

[Car18] Olivia Caramello. *Theories, sites, toposes : relating and studying mathematical theories through topos-theoretic 'bridges'*. Oxford University Press, Oxford, 2018.

[CBH52] Rudolf Carnap and Jehoshua Bar-Hillel. An Outline of a Theory of Semantic Information. Technical report, Research Laboratory of Electronics, MIT, 1952.

[CDLL15] R. Cooper, Simon Dobnik, Staffan Larsson, and Shalom Lappin. Probabilistic type theory and natural language semantics. *Linguistic Issues in Language Technology*, 10, 2015.

[CGCB14] Junyoung Chung, Çaglar Gülçehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. *CoRR*, abs/1412.3555, 2014.

[CGW20] Taco Cohen, Mario Geiger, and Maurice Weiler. A general theory of equivariant CNNs on homogeneous spaces, 2020.

[Cis03] Denis-Charles Cisinski. Images directes cohomologiques dans les catégories de modèles. *Annales Mathématiques Blaise Pascal*, 10(2):195–244, 2003.

[Cis06] Denis-Charles Cisinski. *Les Préfaisceaux Comme Modèles Des Types d'Homotopie*. Astérisque (Société Mathématique de France). Société Mathématique de France, 2006.

[Cis19] Denis-Charles Cisinski. *Higher Categories and Homotopical Algebra*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2019.

[CLS95] Antoine Culioli, Michel Liddle, and John T. Stonham. *Cognition and Representation in Linguistic Theory*. Benjamins, John Publishing Company, 1995.

[Cur13] Justin Curry. Sheaves, cosheaves and applications. *arXiv preprint arXiv:1303.3255*, 2013.

[Cur17] Carina Curto. What can topology tell us about the neural code? *Bulletin of the American Mathematical Society*, 54(1):63–78, 2017.

[CvMBB14] KyungHyun Cho, Bart van Merrienboer, Dzmitry Bahdanau, and Yoshua Bengio. On the properties of neural machine translation: Encoder-decoder approaches. *CoRR*, abs/1409.1259, 2014.

[CWKW19] Taco S. Cohen, Maurice Weiler, Berkay Kicanaoglu, and Max Welling. Gauge equivariant convolutional networks and the icosahedral CNN, 2019.

[CZ21] Olivia Caramello and Riccardo Zanfa. Relative Topos Theory via Stacks. `https://arxiv.org/abs/2107.04417`, 2021.

[CZCG05] Gunnar Carlsson, Afra Zomorodian, Anne Collins, and Leonidas J Guibas. Persistence barcodes for shapes. *International Journal of Shape Modeling*, 11(02):149–187, 2005.

[DHKS04] William G. Dwyer, Philip S. Hirschhorn, Daniel M. Kan, and Jeffrey H. Smith. *Homotopy Limit Functors on Model Categories and Homotopical Categories*. AMS Mathematical Surveys and Monographs, 2004.

[DHSB20] David Ding, Felix Hill, Adam Santoro, and Matt M. Botvinick. Object-based attention for spatio-temporal reasoning: Outperforming neuro-symbolic models with flexible distributed architectures. *CoRR*, abs/2012.08508, 2020.

[Dou92] Daniel J. Dougherty. Closed categories and categorical grammar. *Notre Dame Journal of Formal Logic*, 34(1):36 – 49, 1992.

[Dug08] Daniel Dugger. A primer on homotopy colimits. `https://pages.uoregon.edu/ddugger/hocolim.pdf`, 2008.

[EK66] Samuel Eilenberg and G. Max Kelly. Closed categories. In S. Eilenberg, D. K. Harrison, S. MacLane, and H. Röhrl, editors, *Proceedings of the Conference on Categorical Algebra*, pages 421–562, Berlin, Heidelberg, 1966. Springer Berlin Heidelberg.

[EM45] Samuel Eilenberg and Saunders MacLane. General theory of natural equivalences. *Transactions of the American Mathematical Society*, 58(2):231–294, 1945.

[Fri05] Joel Friedmann. Cohomology in Grothendieck topologies and lower bounds in Boolean complexity. *ArXiv*, 2005.

[FS18] Brendan Fong and David I Spivak. Seven sketches in compositionality: An invitation to applied category theory, 2018.

[FST19] Brendan Fong, David I. Spivak, and Rémy Tuyéras. Backprop as functor: a compositional perspective on supervised learning. *ArXiv*, 2019.

[GB20] Håkon Robbestad Gylterud and Elisabeth Bonnevier. Non-wellfounded sets in homotopy type theory, 2020.

[GH11] R. Ghrist and Y. Hiraoka. Applications of sheaf cohomology and exact sequences on network codings. *Research Institute for Mathematical Sciences*, 1752:31–40, 2011.

[Gir64] Jean Giraud. *Méthode de la descente*. Number 2 in Mémoires de la Société Mathématique de France. Société mathématique de France, 1964.

[Gir71] Jean Giraud. *Cohomologie non abélienne*. Springer, 1971.

[Gir72] Jean Giraud. Classifying Topos. Toposes, algebraic Geometry and Logic, Dalhousie Univ. Halifax 1971, Lect. Notes Math. 274, 43-56 (1972)., 1972.

[Gir87] Jean-Yves Girard. Linear logic. *Theoretical Computer Science*, 50(1):1–101, 1987.

[GLH+20] Anirudh Goyal, Alex Lamb, Jordan Hoffmann, Shagun Sodhani, Sergey Levine, Yoshua Bengio, and Bernhard Schölkopf. Recurrent independent mechanisms, 2020.

[Gro57] Alexander Grothendieck. Sur quelques points d'algèbre homologique. *Tohoku Mathematica Journal*, 9:119–221, 1957.

[Gro83] Alexander Grothendieck. Pursuing Stacks. `https://thescrivener.github.io/PursuingStac` 1983.

[Gro90] Alexandre Grothendieck. Les Dérivateurs. `https://webusers.imj-prg.fr/~georges.maltsi` 1990.

[GWDPL76] C.G. Gibson, K. Wirthmuller, A.A. Du Plessis, and Eduard Looijenga. *Topological Stability of Smooth Mappings*, volume 552 of *Lecture Notes in Mathematics*. Springer, 1976.

[GZ67] Pierre Gabriel and Michel Zisman. *Calculus of Fractions and Homotopy Theory*. Ergebnisse der Mathematik und ihrer Grenzgebiete. Island Press, 1967.

[Has03] Masahito Hasegawa. Coherence of the double negation in linear logic. In *Algebra, Logic and Geometry in Informatics*, 2003.

[HB20] R. Devon Hjelm and Philip Bachman. Representation learning with video deep infomax, 2020.

[Hir03] Philip S. Hirschhorn. *Model Categories and their Localizations*, volume 99 of *Mathematical Surviews and Monographs*. AMS, 2003.

[HM18] Drew Hudson and Christopher Manning. Compositional attention networks for machine reasoning. `https://arxiv.org/pdf/1803.03067.pdf`, March 2018.

[HM21] Liao Heng and Bill McColl, editors. *Mathematics for Future Computing and Communications*. Cambridge University Press, 2021.

[Hoc69] Melvin Hochster. Prime ideal structure in commutative rings. *Transactions of the American Mathematical Society*, 142:43–60, 1969.

[Hol01] Sharon Hollander. A homotopy theory for stacks. *Israel Journal of Mathematics*, 163:93–124, 2001.

[Hol08] Sharon Hollander. A homotopy theory for stacks. *Israel Journal of Mathematics*, 2008.

[HS97] Sepp Hochreiter and Jurgen Schmidhuber. Long-Short Term Memory. *Neural Computation*, 9, 1997.

[HS98] Martin Hofmann and Thomas Streicher. The Groupoid interpretation of type theory. In *Twenty-five years of constructive type theory (Venice, 1995)*, volume 36 of *Oxford Logic Guides*, pages 83–111. Oxford Univ. Press, New York, 1998.

[HS17] Joel C. Heck and Fathi M. Salem. Simplified minimal gated unit variations for recurrent neural networks. *CoRR*, abs/1701.03452, 2017.

[Ill14] Luc Illusie. Travaux de Gabber sur l'uniformisation locale et la cohomologie étale des schémas excellents. In Luc Illusie, Yves Laszlo, and Fabrice Orgogozo, editors, *Logical Aspects of Computational Linguistics*, pages 213–234. Asterisque, 2014.

[Jar09] John Frederick Jardine. Cocycle categories. In Nils Baas, Eric M. Friedlander, Björn Jahren, and Paul Arne Østvær, editors, *Algebraic Topology: The Abel Symposium 2007*, pages 185–218, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg.

[KL14] Andrej Karpathy and Fei-Fei Li. Deep visual-semantic alignments for generating image descriptions. *CoRR*, abs/1412.2306, 2014.

[KLV12] Chris Kapulkin, Peter Lumsdaine, and Vladimir Voevodsky. The simplicial model of univalent foundations. *arxiv*, 11 2012.

[Koc70] Anders Kock. Monads on symmetric monoidal closed categories. *Archiv der Mathematik*, 21:1–10, Jan. 1970.

[Kon18] Risi Kondor. $N-$body Networks: a Covariant Hierarchical Neural Network Architecture for learning atomic potentials. *ArXiv*, abs/1803.01588, 2018.

[Lam58] Joachim Lambek. The mathematics of sentence structure. *The American Mathematical Monthly*, 65(3):154–170, 1958.

[Lam88] Joachim Lambek. *Categorial and Categorical Grammars*, pages 297–317. Springer Netherlands, Dordrecht, 1988.

[Lin89] Ingrid Lindström. A construction of Non-Well-Founded Sets within Martin-Löf's Type Theory. *The Journal of Symbolic Logic*, 54(1):57–64, 1989.

[Lin20] Johan Lindberg. *Localic Categories of Models and Categorical Aspects of Intuitionistic Ramified Type Theory*. PhD thesis, Stockholm University, Department of Mathematics, 2020.

[Loo78] Eduard Looijenga. The discriminant of a real simple singularity. *Compositio Mathematica*, 37(1):51–62, 1978.

[Low15] Zhen Lin Low. Cocycles in categories of fibrant objects. `https://arxiv.org/abs/1502.03925`, 2015. arXiv.

[LS81] Joachim Lambek and Philip J. Scott. Intuitionist type theory and foundations. *Journal of Philosophical Logic*, 10:101–115, 1981.

[LS88] Joachim Lambek and Philip J Scott. *Introduction to higher-order categorical logic*, volume 7. Cambridge University Press, 1988.

[Luo14] Zhaohui Luo. Formal semantics in modern type theories: Is it model-theoretic, proof-theoretic, or both? In Nicholas Asher and Sergei Soloviev, editors, *Logical Aspects of Computational Linguistics*, pages 177–188, Berlin, Heidelberg, 2014. Springer Berlin Heidelberg.

[Lur09] Jacob Lurie. *Higher Topos Theory (AM-170)*. Academic Search Complete. Princeton University Press, 2009.

[Mac71] Saunders MacLane. *Categories for the Working Mathematician*. Springer-Verlag, New York, 1971. Graduate Texts in Mathematics, Vol. 5.

[Mac12] Saunders MacLane. *Homology*. Classics in Mathematics. Springer Berlin Heidelberg, 2012.

[Mal05] Georges Maltsiniotis. *La théorie de l'homotopie de Grothendieck*. Number 301 in Astérisque. Société mathématique de France, 2005.

[Mar82] Jean Martinet. *Singularities of smooth functions and maps. Transl. from the French by Carl P. Simon*, volume 58. Cambridge University Press, Cambridge. London Mathematical Society, London, 1982.

[MBHSL19] Haggai Maron, Heli Ben-Hamu, Nadav Shamir, and Yaron Lipman. Invariant and equivariant graph networks, 2019.

[McG54] William McGill. Multivariate information transmission. *Psychometrika*, 19:97–116, 1954.

[Mel09] Paul-André Melliès. *Categorical Semantics of Linear logic*. SMF, 2009.

[Mil62] John Milnor. On axiomatic homology theory. *Pacific Journal of Mathematics*, 12(1):337 – 341, 1962.

[ML80] Per Martin-Löf. *Intuitionistic Type Theory*. Lectures given at University of Padova, 1980.

[MLM92] Saunders Mac Lane and Ieke Mœrdijk. *Sheaves in geometry and logic: a first introduction to topos theory.* Universitext. New York etc.: Springer-Verlag., 1992.

[MM20] Yuri Manin and Matilde Marcolli. Homotopy theoretic and categorical models of neural information networks. *ArXiv*, 2020.

[Mog91] Eugenio Moggi. Notions of computation and monads. *Information and Computation*, 93(1):55–92, 1991.

[Mon70] Richard Montague. Universal grammar. *Theoria*, 36, 1970.

[MT10] Paul-André Melliès and Nicolas Tabareau. Resource modalities in tensor logic. *Annals of Pure and Applied Logic*, 161(5):632–653, February 2010.

[MXY+15] Junhua Mao, Wei Xu, Yi Yang, Jiang Wang, and Zhiheng Huangand Alan Yuille. Deep captioning with multimodal recurrent neural networks (m−RNN). *ArXiv*, abs/1412.6632v5, 2015.

[Par75] Barbara Partee. Montague grammar and transformational grammar. *Linguistic Inquiry*, 6(2):203–300, 1975.

[Pea88] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.

[Pel20] Olivier Peltre. *A Homological Approach to Belief Propagation and Bethe Approximations*. PhD thesis, Dept of Mathematics, Université Paris Diderot, 2020.

[PKM19] Alexander Port, Taelin Karidi, and Matilde Marcolli. Topological analysis of syntactic structures. *arXiv preprint arXiv:1903.05181*, 2019.

[Pri94] Hillary Priestley. Spectral sets. *Journal of Pure and Applied Algebra*, 94:101–114, 1994.

[Pro08] Alain Prouté. La théorie des ensembles selon les Shadoks. Séminaire général de Logique de l'université Paris Diderot, 2008.

[Pro19] Alain Prouté. Introduction à la logique catégorique. MSc Course, Université Paris Diderot, 2019.

[Qui67] Daniel G. Quillen. *Homotopical Algebra*. Lecture notes in mathematics. Springer-Verlag, 1967.

[Rap10] George Raptis. Homotopy theory of posets. *Homology, Homotopy and Applications*, 12:211–230, 2010.

[Rob17] Michael Robinson. Sheaves are the canonical data structure for sensor integration. *Information Fusion*, 36:208–224, 2017.

[rR09] Jiří Rosický. On combinatorial model categories. *Applied Categorical Structures*, 17:303–316, 2009.

[RSB+17] David Raposo, Adam Santoro, David G. T. Barrett, Razvan Pascanu, Timothy P. Lillicrap, and Peter W. Battaglia. Discovering objects and their relations from entangled scene representations. *CoRR*, abs/1702.05068, 2017.

[See84] R. A. G. Seely. Locally cartesian closed categories and type theory. *Mathematical Proceedings of the Cambridge Philosophical Society*, 95(1):33–48, 1984.

[SFR+18] Adam Santoro, Ryan Faulkner, David Raposo, Jack W. Rae, Mike Chrzanowski, Théophane Weber, Daan Wierstra, Oriol Vinyals, Razvan Pascanu, and Timothy P. Lillicrap. Relational recurrent neural networks. In *NeurIPS*, 2018.

[SGW21] Dan Shiebler, Bruno Gavranović, and Paul W. Wilson. Category theory in machine learning. *CoRR*, abs/2106.07032, 2021.

[Shu10] Michael A. Shulman. Stack semantics and the comparison of material and structural set theories, 2010.

[Shu12] Michael Shulman. Exact completions and small sheaves. *Theory and Applications of Categories*, 27:97–173, 2012.

[Shu19] Michael Shulman. Comparing material and structural set theories. *Annals of Pure and Applied Logic*, 170(4), Apr 2019.

[SP12] Chris Schommer-Pries. The canonical model structure on Cat. `https://sbseminar.wordpress.com/2012/11/`, 2012. Blog.

[SP21] Grégoire Sergeant-Perthuis. *Intersection property, interaction decomposition, regionalized optimization and applications*. PhD thesis, University of Paris Diderot, March 2021.

[SRB+17] Adam Santoro, David Raposo, David G. T. Barrett, Mateusz Malinowski, Razvan Pascanu, Peter W. Battaglia, and Timothy P. Lillicrap. A simple neural network module for relational reasoning. *CoRR*, abs/1706.01427, 2017.

[Sta] Stacks. The Stacks Project. `https://stacks.math.columbia.edu/`.

[Sta14] Alexandru E. Stanculescu. Stacks and sheaves of categories as fibrant objects, I and II. *Arxiv*, 2014.

[SW49] Claude E. Shannon and Warren Weaver. *The Mathematical Theory of Communication*. University of Illinois Press, Urbana and Chicago, 1949.

[Ted16] Christopher Francis Tedd. *Ring constructions on spectral spaces*. PhD thesis, University of Manchester, 2016.

[Tho72] René Thom. *Stabilité structurelle et morphogénése : Essai d'une théorie générale des modèles*. Benjamin, 1972.

[Tho80] Robert Wayne Thomason. Cat as a closed model category. *Cahiers de Topologie et Géométrie Différentielle Catégoriques*, 21(3):305–324, 1980.

[Tho83] René Thom. *Mathematical Models of Morphogenesis*. Ellis Horwood Series in Mathematics and its applications. Ellis Horwood, 1983. translated by W.M. Brookes and D. Rand.

[Tho88] René Thom. *Esquisse d'une sémiophysique : Physique aristotélicienne et théorie des catastrophes*. Dunod, 1988.

[Tin62] Hu Kuo Ting. On the amount of information. *Theory of Probability and its Applications*, 1962.

[Uem17] Taichi Uemura. Fibred fibration categories. *2017 32nd Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*, Jun 2017.

[vB90] Johan van Benthem. Categorial grammar and type theory. *Journal of Philosophical Logic*, 19(2):115–168, 1990.

[Vig19] Juan-Pablo Vigneaux. *Topology of statistical systems : a cohomological approach to information theory*. PhD thesis, University of Paris Diderot, 2019.

[Vig20] Juan Pablo Vigneaux. Information structures and their cohomology. *Theory Appl. Categ.*, 35:1476–1529, 2020.

[VSP+17] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *CoRR*, abs/1706.03762, 2017.

[Wit53] Ludwig Wittgenstein. *Philosophical Investigations*. Oxford, Blackwell, 1953.

[XQLJ20] Huiqiang Xie, Zhijin Qin, Geoffrey Ye Li, and Biing-Hwang Juang. Deep learning enabled semantic communication systems. *ArXiv*, 2020.

[YFW01] Jonathan S Yedidia, William Freeman, and Yair Weiss. Generalized belief propagation. In T. Leen, T. Dietterich, and V. Tresp, editors, *Advances in Neural Information Processing Systems*, volume 13. MIT Press, 2001.

[ZCZ+19] Tao Zhuo, Zhiyong Cheng, Peng Zhang, Yongkang Wong, and Mohan Kankanhalli. Explainable video action reasoning via prior knowledge and state transitions. In *Proceedings of the 27th ACM International Conference on Multimedia*, MM '19, New York, NY, USA, 2019. Association for Computing Machinery.

[ZRS+18] Vinícius Flores Zambaldi, David Raposo, Adam Santoro, Victor Bapst, Yujia Li, Igor Babuschkin, Karl Tuyls, David P. Reichert, Timothy P. Lillicrap, Edward Lockhart, Murray Shanahan, Victoria Langston, Razvan Pascanu, Matthew Botvinick, Oriol Vinyals, and Peter W. Battaglia. Relational deep reinforcement learning. *CoRR*, abs/1806.01830, 2018.

[ZWZZ16] Guo-Bing Zhou, Jianxin Wu, Chen-Lin Zhang, and Zhi-Hua Zhou. Minimal Gated Unit for Recurrent Neural Networks. *CoRR*, abs/1603.09420, 2016.