information

MDPI

Article

On the Functional Nature of Cognitive Systems

Vincenzo Manca

Department of Computer Science, University of Verona, 37134 Verona, Italy; vincenzo.manca@univr.it

Abstract: The functional nature of cognitive systems is outlined as a general conceptual model where typical notions of cognition are analyzed apart from the physical realization (biological or artificial) of such systems. The notion of function, one of the main logical bases of mathematics, logic, linguistics, physics, and computer science, is shown to be a unifying concept in analyzing cognition components: learning, meaning, comprehension, language, knowledge, and consciousness are related to increasing levels in the functional organization of cognition.

Keywords: neural networks; cognitive systems; learning; machine learning; large language models; transformers; meaning; comprehension; knowledge

1. Introduction

The idea of a neural network was first introduced in the seminal paper of Warren McCulloch and Walter Pitts [1]. Neurons were considered containers of discrete values, and synapses connecting neurons were seen as functions sending values between connected neurons, depending on the values contained in the neurons. These networks were shown to be equivalent to finite automata [2]. An important change of perspective in this approach emerged after the seminal book of Donald Hebb [3,4], where the intrinsic functional perspective of neural networks shifted toward a dual vision of these systems. Neurons became places containing functions, and the synapses expressed arrows to which real values, **weights** were associated to denote connection strengths. In the following years, seminal papers with different perspectives but with a common emphasis on the connectionist nature of these systems enforced this vision, giving more relevance to the Hebbian viewpoint focused on the organization of behavior as a result of synaptic activity [5–10].

The mathematical notion of function is strongly related to that of a dependent variable, which assumes values depending on the values taken by an independent variable. In *Introductio in Analysin in Infinitorum* (1748) [11], Leonhard Euler gives the first account of a function related to the notion of algebraic expression. In the 1920s, Alonzo Church introduced λ-notation, which, more rigorously, expresses functions independently from the used variables. Moreover, in 1936, Church proved the equivalence between *lambda* functions and Turing machines by showing that Euler's functional notation essentially provides all the computable functions.

Surely, the recent success of artificial neural networks is related to the specific structures of their internal organization, the architecture of these structures, the learning algorithms, the computational power of electronic devices, and the availability of big data [12–20]. However, a more abstract descriptive theory based on functions can enlighten a logic of cognition organization within a very general perspective, where cognitive levels correspond to functional types and computational concepts become relevant to understanding neurological phenomena and vice versa.

Of course, neural networks are functional networks, but what is new is that other cognition levels can be expressed in terms of other functional networks integrating a basic one. Namely, the general approach of this paper is to represent cognition in functional terms, where learning, comprehension, and knowledge correspond with functional integration, transformation, and abstraction.

Citation: Manca, V. On the Functional Nature of Cognitive Systems. *Information* **2024**, *15*, 807. https://doi.org/10.3390/info15120807

Academic Editor: Birgitta Dresp-Langley

Received: 9 November 2024
Revised: 9 December 2024
Accepted: 11 December 2024
Published: 16 December 2024

In the following, I will show that the abstract notion of functional networks (FNs) allows for a uniform description of many cognitive aspects. First, a direct consequence of functional composition is the "holomorphy" of functional networks. Which is the basis of many phenomena of "cognitive reflexivity".

Any cognition competence is a suitable FN, and the capability of cognitive systems to acquire new competencies for adapting, developing, and evolving is associated with FNs. Namely, I show that learning can be expressed by integrating an FN with a dual FN acting on it (see Proposition 4). Moreover, this integration can be transformed into a resulting FN of the same kind as the original one. Analogously, cognitive concepts such as meaning and knowledge, often used without a clear and precise explanation, result in functional mechanisms of information transformation, where further levels of functional organizations intervene, which maintain a complete homogeneity with the previous functional levels.

The key concept of an FN controlling another FN is that of a meta-function, which changes a function's weights. I will show that meta-functions can be represented with suitable FNs integrating the FN on which they act.

Synapse plasticity was identified as the basis of learning processes and memory [4]. Synapses are very complex structures expressing the levels of connection between neurons. These levels encode many aspects of the relational system that the neuronal tissue expresses. As Eric Kandel has shown [21] in his investigations on Aplysia (a marine slug), an entire history of experiences can drive the state of a synapse, by originating seeds of associative memories that are the basis of many emergent aspects of neurons. The function of a single neuron is a real function with only one exit, but minimal aggregates of connected neurons can realize an elementary competence, for example, in the case of Aplysia, a circuit of four neurons is responsible for the gill-withdrawal reflex, in which the slug will withdraw its gill when its adjacent water siphon is touched. This phenomenon unravels the molecular mechanisms for learning and memory. In this system, synapses exhibiting long-term potentiation provide the acquisition of a stable strength in synapse connections, which are responsible for the observed functional acquisition.

This evidence shows the biological relevance of weight synapses as the true encoding of competence acquisition, and functional moduli exhibit the same structure as the biggest units of the whole network they belong to. In this model, functional modules correspond to groups of neurons with a functional identity, which, mathematically, are functions from real hyperspaces to real hyperspaces. The homogeneity between functional moduli and the whole system is a kind of reflexivity, corresponding with holomorphy, which is probably related to the scaling-up of cognitive competencies [22]. Computational approaches based on ANNs *capsule*, neural hyper-networks, neural meta-networks, and compositional ANNs [23–26] develop similar ideas, where certain kinds of self-application are the main ingredient of their architectures (for the role of reflexivity in mathematics, see [27]). In [28], *synapsembles* were introduced, which play a similar role [29].

Synapse modification through experiences is the basis of the open character of cognitive systems, where the updated connection strengths improve behaviors, enforcing new possibilities. This kind of reflexivity is a form of circularity that captures the capability of exhibiting competencies to reconfigure themselves continuously.

Cognitive systems are the typical focus of theoretical psychology and related disciplines from both the neurological or behavioral sides. Artificial intelligence investigation aims to define mathematical and computational models of typical aspects of cognition. However, a mathematical analysis of cognition in itself, abstracting from particular methods of biology or medicine, is necessary for having rigorous definitions, which are now missing, of main cognitive concepts, such as learning, comprehension, and knowledge. On the other hand, the success of computational models, ultimately rooted in mathematical concepts, suggests that mathematical models of cognition could improve the understanding of natural intelligence by suggesting new ideas for artificial intelligence as well.

The following words of the recent Nobel laureate Geoffrey Hinton (Physics Nobel Prize 2024) seem to be very appropriate for supporting the trend that this paper wants to promote, with a reflection on the centrality that mathematical functions could play in cognition.

But sooner or later computational studies of learning in artificial neural networks will converge on the methods discovered by evolution. When that happens a lot of diverse empirical data about the brain will finally make sense and many new applications of artificial neural networks will become feasible [13].

2. Results

2.1. Cognitive Systems

A cognitive system acquires, elaborates, memorizes, and provides information through specific competencies defining its behavior due to its internal structure. However, exhibiting its competencies, it continuously reconfigures itself by changing its structure and competencies accordingly. In this sense, a cognitive system is not only a system elaborating information but an open reflexive informational system elaborating new possibilities of elaborating information.

In the following, I give more details and formal support to this initial intuition of a cognitive system in terms of **functional networks** (FNs).

An FN is a weighted network of real functions belonging to a limited class (polynomial, sigmoid, hyperbolic tangent, zero-linear functions). Any function of an FN has entering or input variables and one exiting or output variable. The entering variables are weighted and summed, and the function is applied to the sum; the exiting variable is equal to the result of the application. Input variables correspond to the output variables of other functions or are global input variables to the FN. Output variables from a node correspond to global output variables of the FN or to input variables to other nodes. No cycles are allowed in the input–output relation. The FN is completely expressed by a system of equations representing a composition of basic functions.

The composition is based on **weights**. Namely, assume that a function f_i has entering arrows $x_1, x_2, \ldots, x_k$ and y_j is the result arrow. Real numbers, called weights, are associated with the entering arrows in such a way that

$$y_j = f(w_1 x_1 + w_2 x_2 + \ldots + w_k x_k)$$

and the argument of f is called its **weighted input**.

All the basic ingredients of mathematics are in such a system: functions, equations, variables, and numbers.

Under general hypotheses, the following proposition holds.

Proposition 1. *Any continuous m-valued real function of n arguments can be approximated (at any given approximation level) by a suitable FN [30–33].*

The last proposition has a strong significance because it entails that acquiring a given competence means acquiring a vector of a real hyperspace. In mathematics, many results ensure the approximate representation of continuous functions using special classes of functions (polynomials, trigonometric polynomials, Bernstein polynomials. . .). The FN approximation result is significant because the class of basic functions is limited and can be reduced to only one type of function (for example, the Rectified Linear Unit given by $\lambda x.max\{0, x\}$), and the level of approximation can be reduced by using many levels in the composition of functions. These two characteristics are essential for the machine learning approach to cognitive systems.

In the following FN, functions are simple algebraic expressions; the variables x and y are global input variables of the FN and input variables for the first two functions; v_1 and z_1 are outputs for the first two functions and input variables for the second pair of functions; v_2 and z_2 are the outputs of the second pair of functions and inputs for the third pair of functions; and u and w are output variables of the third pair of functions and global output as well:

$$\begin{aligned} v_1 &= 2x \times 5y \\ z_1 &= 3x + 2y \\ v_2 &= 2v_1 - z_1 \\ z_2 &= 5v_1^2 + 7z_1 \\ u &= 5z_2 + 10 \\ w &= 3v_2 \end{aligned}$$

Therefore, equations are organized at different levels. At the first level, we have the input variables (x, y), which are weighted arguments of some functions that provide second-level variables (v_1, z_1). In general, variables of level i are weighted arguments of functions giving the variables of level $i + 1$. Variables of the maximum level are the output variables of the system (u, w). In other words, an FN is realized by a system of equations organized in levels; at each level, some functions transform variables of a given level into variables of the next level, avoiding jumps of more than one level (jumps can be avoided by adding identity functions that pass values from one level to the next without altering them).

The equation system above expresses a function from $\mathbb{R}^2 \to \mathbb{R}^2$. The Figure 1 below is the graph representation of the above system of equations ($i = \lambda x; q = \lambda x, y.x^2 + y$).

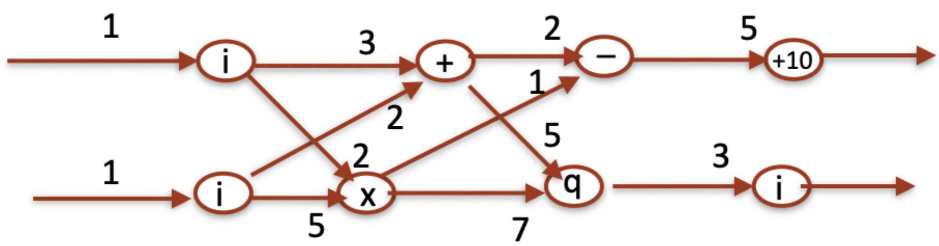

Figure 1. A graph representation of the above FN expressed by a system of equations.

Any part of an FN where we can distinguish n input variables and m output variables represents a function between real hyperspaces. The system above is of type $FN(2, 2)$ (two input variables and two output variables) or, more precisely, of type $FN(2, 2, 2, 2)$ if we also want to explicitly obtain the number of variables of the internal levels. Of course, the terminology extends naturally to a type $F(n, m)$, for any positive n and m, and a type $FN(n_1, n_2, \ldots, n_k)$ refers to FNs with k levels with n_1 input variables, n_k output variables, and $n_2, \ldots, n_{k-1}$ variables of internal levels. Figure 2 visualizes the holomorphy of FNs.

Consider a simple example of behavior competence: *crossing the road in a city*. The crossing persons can perform a safe crossing task when they evaluate the distances and speeds of vehicles approaching the crossing zone and consequently are provided with information about times, speeds, and trajectories of motion, with possible points and times of interruption. All this information is encoded in input and output values, whose association expresses competence. It is coordinated with other sensorimotor competencies and translated into muscle forces, rotations, and tractions at many levels, involved in the movements necessary for the whole process. In general, we can state the following proposition.

Proposition 2. *Any Cognition Competence identifies a real function* $F : \mathbb{R}^n \to \mathbb{R}^m$.

Therefore, the capability to adequately perform a process is ultimately represented by a function or several coordinated functions, giving the right correspondences between arguments and results that a given ability requires through a suitable weighted composition of elementary functions.

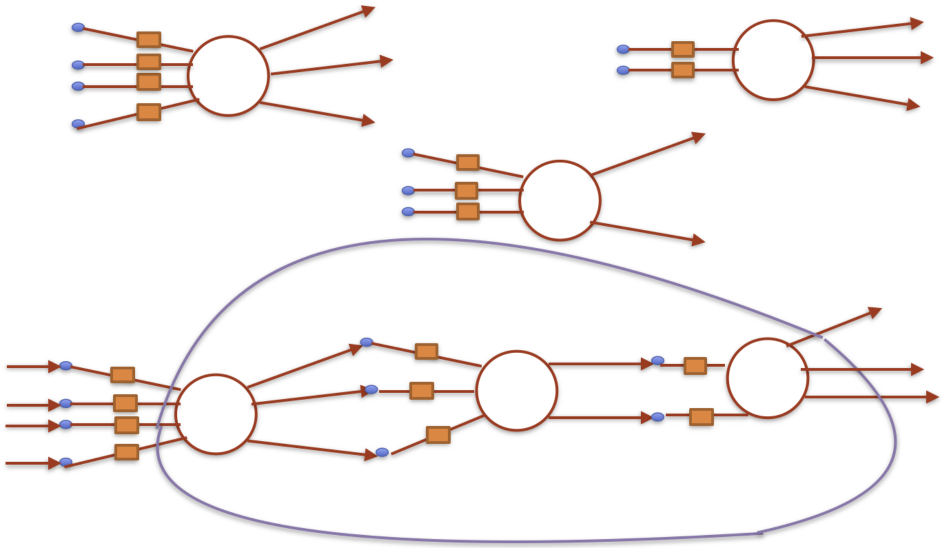

Figure 2. FN holomorphy. Top: three weighted functions (bullets receive inputs and rectangles represent weights). Bottom: an FN is obtained by connecting the functions on the top, which provides a weighted function of the same kind as a single connected elements.

Given an FN, the real function it computes is completely identified by the weights of its connections; this means that F is given by a vector in the real hyperspace $\mathbb{R}^S$, where S is the number of FN weights.

A short digression on the neuronal systems of animals is useful for appreciating the biological meaning of the picture given above. An animal neuron is a special cell and a neuronal tissue is an aggregate of connected neurons exchanging signals. However, signals are represented by numerical variables with a complex physical and biochemical nature. In particular, there is a strong difference between afferent and efferent signals. The neuron exit signal has an electrical nature due to a change of cellular electrical potential, and this potential, through branches of central axons, called dendrites, propagates outside the neuron (toward other neurons, or sensorimotor cells that are not neurons). On the other hand, synapses are cell sites receiving stimuli (internal or external to neuronal tissue) where signals activate biochemical processes mediated by appropriate kinds of molecules that interact with the neuron cell dynamics at many levels of its internal metabolism and its informational DNA-RNA-based system.

2.2. Attention and Learning

In our initial definition of a cognitive system, two main aspects are present: (1) a cognitive system performs some competencies; (2) it continuously reconfigures its structure to extend and enrich its competencies.

This second aspect implies a continuous activity of control over behavior, which suggests a dual nature of a cognitive system, where some parts have to control other parts. An attention FN is a functional module of an FN that controls other parts of it. Since the behavior of an FN is completely encoded by its weights, the only way to alter it is to act on weights. We call the **learning network** a functional module that receives "correction stimuli" as inputs and provides outputs that change the weights for "improving" the behavior of an FN, following some criteria of adequacy.

The passage from computer science to artificial intelligence is an epochal step synthesized by the expression *from computing machines to machine learning*. A computing machine is a device that is able to compute a function expressed by an internal program, providing the result of the function in correspondence to any argument of its domain. Conversely, a machine-learning system from some triples (argument, result, error) can discover the function approximating, within a given error threshold, the functional rule underlying all the pairs (argument, result) received in a training process.

I will now explain in more detail the possibility of describing learning in terms of functional moduli within a cognitive system.

A weight adjustment mechanism, coupled with a functional network, is a necessary condition for learning. In this sense, learning can be seen as a meta-competence, giving the ability to acquire competences through a finite number of pairs (input, output) and corresponding errors, evaluated in some way, committed to providing outputs different from those of a target function. The following proposition ensures this possibility.

Proposition 3. *Given a real differentiable function $F : \mathbb{R}^n \to \mathbb{R}^m$, there exist effective and efficient methods for determining, from a finite subset of the graphic of F, the weights of an FN approximating F, at a specified approximation level (see Backpropagation via Descendent Gradient [7,9,10,20,32–35]).*

The proposition above entails the passage from learning via instructions (programs) to learning via training (examples). Cognitive systems are not "programmed" as computers are. Namely, they are "trained" by discovering the internal rule expressing a function through trial and error. The training efficiently converges to a target function by adjusting the weights of an FN until it reaches, with a good approximation, the desired function corresponding to the learned competence. Even this aspect is described in functional terms because it corresponds to the search for a minimum of the function expressing the errors committed in acquiring the target function.

The general schema of training in acquiring competence can be described in functional terms, apart from any specific algorithm of its realization. Let f_R be the function computed by an FN and $F_C(a)$ the function expressing a competence. Given an argument a, the error in computing $F_C(a)$ can be expressed by $|f_R(a) - F_C(a)|$.

Suppose the FN is provided with a sequence of arguments $a_1, a_2, a_3, \ldots$ for training, and correspondingly, gives the errors $e_1, e_2, e_3, \ldots$. For each error, the network changes its weights by reconfiguring itself. Let $R_1, R_2, R_3, \ldots$ be the sequence of networks obtained after reconfigurations. The network has an adequate training strategy when, after a certain number of steps, the errors are under a prefixed error threshold ϵ, by arriving at a step j where $|f_{R_j}(a) - F_C(a)| < \epsilon$.

Natural cognitive systems can learn because they possess learning strategies according to the pattern above. The artificial intelligence revolution of recent years has been based on machine learning, which essentially discovered efficient algorithms for implementing the learning schema outlined above. Moreover, also the learning schema of an FN can be described in terms of a system of functional equations [36] as a dual FN acting on the given FN for reconfiguring its weights toward a target function. In other words, learning is the super competence allowing for the acquisition of competencies. The following proposition holds.

Proposition 4. *Given an FN of type $FN(n_1, n_2, \ldots, n_k)$, a dual* **learning network** *of type $FN(n_n, n_{k-1}, \ldots, n_1)$ exists acting on the first one and directing its learning strategy toward the target function of the competence acquired by its training.*

Proof. Backpropagation is an algorithm by which a network learns to compute a given function F through examples. The following system of equations (see [36], Theorem in Sect. three) represents the backpropagation, where (1) $\boldsymbol{E}$ is the error function (usually expressed by the sum of differences between the target function and the function computed by the FN); (2) Out represents the exit variables, and Int the entering variables of the function f_j; (3) $w_{i,j}$ are the weights of the entering variables u_i to f_j; (4) $\Delta w_{i,j}$ are the **weight variations**; $\eta < 1$ is a chosen fixed parameter, called *the learning rate*; $f'_j(z_j)$ are the derivative of $u_j = f_j(z_j)$ with respect to its weighted input z_j:

$$\boldsymbol{\Delta w_{i,j} = -\eta \delta_j u_i} \tag{1}$$

$$\delta_j = \frac{\partial E}{\partial u_j} f'_j(z_j) \quad \text{if } \boldsymbol{j} \in \boldsymbol{Out}(\boldsymbol{M}) \tag{2}$$

$$\delta_j = f'_j(z_j) \sum_{k:j \to k} \delta_k \, w_{j,k} \quad \text{if } \boldsymbol{j} \in \boldsymbol{Int}(\boldsymbol{M}) \tag{3}$$

The proof follows by integrating the training FN with a trained FN. Figure 3 is a visualization of integrating an FN with a learning FN controlling it. The main idea of integrating the trained FN with the training one (learning FN) is in the right part of the figure. Namely, each arrow of the first original FN is replaced by a node of the second FN and by two arrows, determining a bridge path with the same origin and target of the replaced arrow. The weight update mechanism for a single function is visualized in Figure 4 by node K, which is a meta-function because it acts on a function weight. The transformation of the meta-function K in a usual function L is visualized in Figure 5 on the bottom by a "bridge function" that provides the same effect as the meta-function by transforming the connection G-H into the connection G-L and adding a connection L-H with weight 1. The function L transforms its input variable received from G by realizing the weight variation required by K. By exploiting the bridge mechanism to all the functions of the trained FN, we obtain a second FN, integrated with the original one, that realizes the learning mechanism of the backpropagation mechanism outlined in Figure 3. □

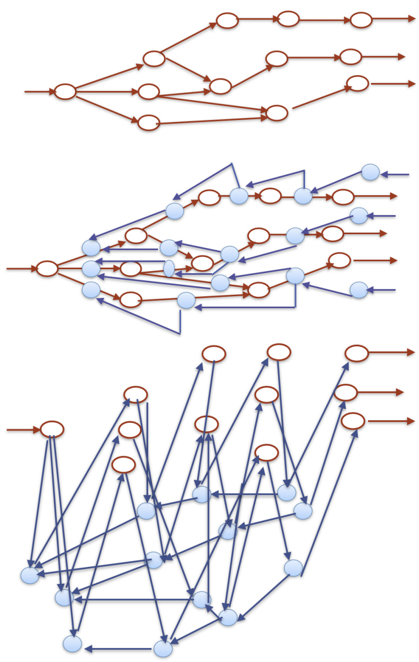

Figure 3. An FN on top and its integration with learning FN on bottom. Integration is represented at two levels, employing a reverse network of nodes that are arrow bridges of original FN.

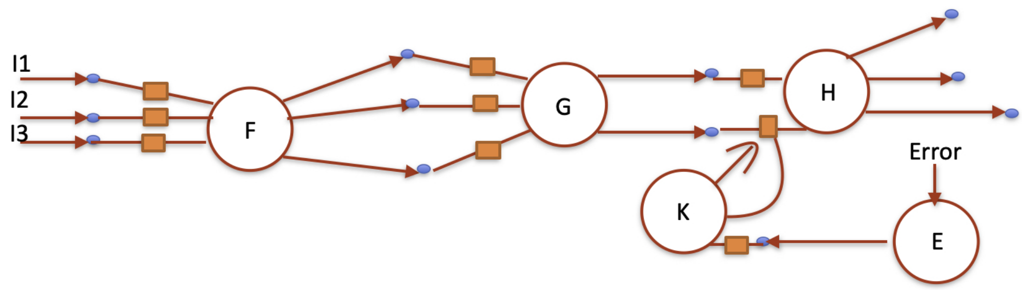

Figure 4. An FN and a meta-function K adjusting a weight according to an input error (between the computed function and a target function).

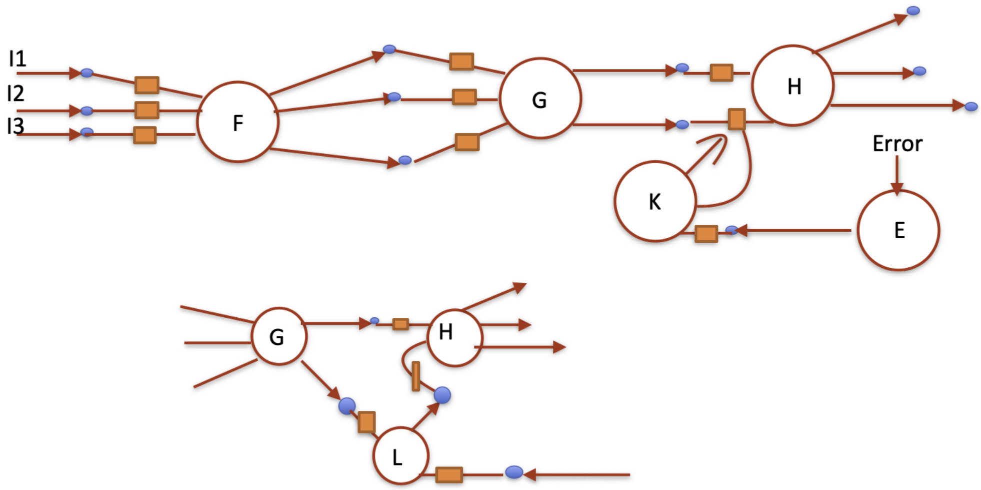

Figure 5. The translation of the meta-function K into a function providing the same effect according to a bridge mechanism.

This idea of meta-functions, realized by usual functions through the bridge mechanism, could be applied to express many kinds of super competencies. Figure 6 shows a memory mechanism based on meta-functions.

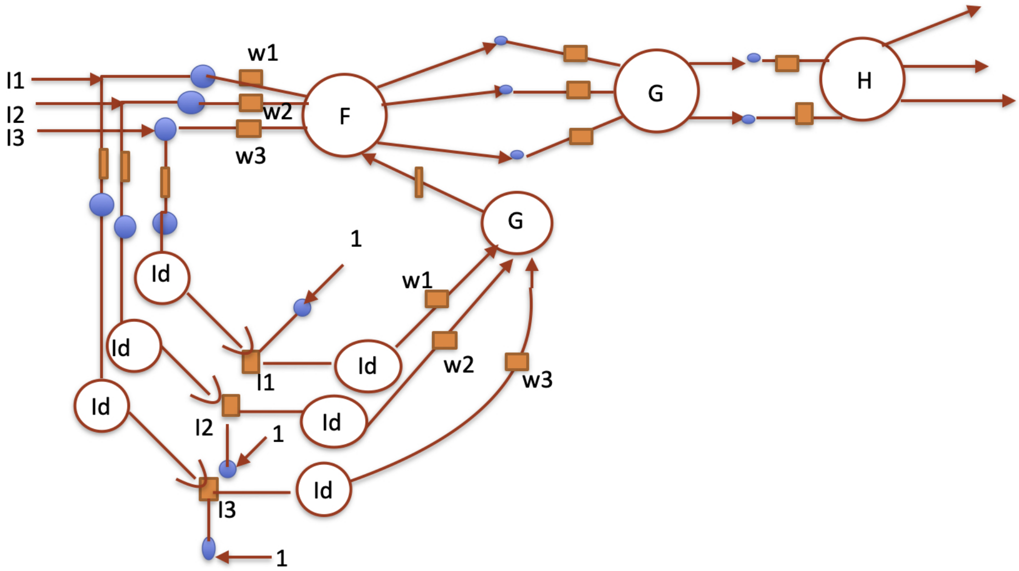

Figure 6. The inputs I1, I2, and I3 of F-G-H are sent to the FN with six Id functions (Id is the identity), where three meta-functions update three weights. When input 1 is given to the synapses on the bottom, G sends to F the same value generated by I1, I2, and I3. In other words, the FN on the bottom memorizes inputs of F-G-H as weights (those between the pairs of Id functions). This representation individuates a memory mechanism transforming input values into weights, where meta-functions are essential (weights indicated by slim rectangles have value 1).

Learning is not a "conscious" activity; it is performed without the learner understanding how it happens. It is similar to a child who can go on a bike without knowing the mechanisms that make it possible for them to do so. Comprehension corresponds to elaborating an internal model of what we do and observe.

2.3. Meaning and Comprehension

The notion of meaning is associated with symbolic systems by which cognitive systems represent the external world and interact with other cognitive systems living in that world. Even in the most primitive cases, symbols represent things and relations between them.

A cognitive system is based on primary data (sensory and perceptive) that the system receives from outside and organizes within its functional networks. These networks represent associations of primary data with "forms" internal to the system (sounds, images, odors, tastes, touches). When symbolic systems evolve toward languages, the associative mechanisms become more complex because symbols express data and data aggregates, but also relations, facts, evaluations, judgments, and previsions. In other words, primary data become the terminal points of conceptual systems that realize higher-order realities, giving an interpretation to the primary data derived from the external reality of observable things and facts. The terms meaning and comprehension refer to the mechanisms linking things and facts from external reality with the conceptual world that cognitive systems have elaborated as a result of a process of internal organization, developed through interactions with the world and other cognitive systems. The conceptual world related to a language is intrinsically based on societies of cognitive systems, that is, groups of individual cognitive systems elaborating a culture and a collective educational system. Every language is developed within a society. Namely, the individual can establish a symbolic system by interacting with someone similar to him but external to him, in the same external world. From this collective perspective, the world is seen by the individual's eyes and through interactions with other eyes seeing it.

The recent success of chatbots is based on transformations of words into numerical vectors, **Embedding Vectors**, of dimensions of several thousands of components. These components express membership to syntactical and semantical categories and features due to the contexts where a word can occur in the text of the given language. These vectors express the correct use of words in discourse [37]. Namely, knowing the meaning of a word means using it correctly. This notion of meaning has a long tradition in linguistics and philosophy [38]. It has become effective with the implementation of this idea in artificial intelligence with language representation systems such as Word2Vec [38] and its developments in the wider context of LLMs (large language models) [19] and recent conversational systems.

In phonemics, any phoneme is defined by the values of pertinent tracts of sounds associated with distinctive features for the phoneme (labial, guttural, velar, nasal, dental, occlusive…). In the case of words, typical features are syntactic categories, semantical aspects, contexts, styles, registers, etc. Given a linguistic corpus, these features are completely derived from all the propositions where words occur, syntactic roles, replaceable words, forbidden or expected words in close positions, etc. Many of these features correspond to numerical values of vector components corresponding to scores in linguistic games based on paraphrases, synonyms, antonyms, comparisons, analogies, metaphors, syntheses, and expansions (similar words successfully pass almost the same tests). This very kind of game was the basis of the *pretraining* phase in *Generative Pretrained Transformer* models of artificial conversational models [39]. However, the basic idea of semantic vectors is that words have *distributional profiles* in the texts of a given language, and these profiles can be encoded by numerical (probabilistic) vectors, from a perspective very close to [40].

According to this perspective, meanings are points of a multidimensional vector space, the *language semantic space*, with two important characteristics: *openness* and *relativity*. The component values of such vectors are not fixed because these values express relations among words, and these relations continuously change during conversational and dialogical activity. Moreover, these values do not have an absolute character because they depend on the textual fonts, the particular strategies for which they are computed, and the history of training and development of a conversational system. Linguistic games provide components of linguistic embedding vectors. However, numerical values of related semantic vector spaces depend on the criteria for measuring features of their word distributional profiles. This means that different individual experiences determine different semantic vectors and related semantic spaces. Then, a natural question arises. How can we conciliate the relativism of meaning with its social communication value? A possible answer is that understanding is always uncertain, approximate, partial, and erroneous. People's communication is generally based on a small percentage of reliability that is inversely proportional to the semantic complexity of the relevant communication and directly proportional to the similarity of the semantic spaces of communicating subjects. Nevertheless, on average, language communication works enough to produce correct and reliable interactions.

Given a semantic space, comprehension corresponds to the localization of linguistic expressions in this space. A proposition of n words provides a $(n+1)$-gon where vertices are the origin and the n vectors exiting from the origin of the space (of many thousands of components). In the most recent transformers, this n-gon is further transformed by using the contextualization matrices Query and Key, which are constructed in the training phase, for considering the influence of the context for the words occurring in the sentence. The scalar products of these matrices give final value vectors and, consequently, a new n-gon. A vector inside this n-gon, and close to its barycenter, provides the meaning of the original sentence as a vector with the same dimensionality as single words. Of course, this mechanism can be iterated at other levels of meaning composition, giving **meaning invariance** for all semantic levels constructed from the basic word level.

This interesting phenomenon is exemplified in the way meanings are represented within the most recent chatbot of the ChatGPT family of OpenAI (since the version ChatGPT.4). Complex meanings are obtained by processes of localization inside semantic spaces in regions delimited by vertices. In these localizations, single meanings lose some specific individual aspects, but final embedding vectors gain the characters of synthesis and semantic amalgamation.

A consequence of the intrinsic homogeneity of meanings is that it allows for extending language semantics to complex conceptual meanings by passing from semantics to knowledge.

2.4. Language and Knowledge

So far, we have considered functional networks and attention mechanisms that control them during learning processes. However, a complex cognitive system is built of many—up to tens of thousands of—functional networks exhibiting basic behaviors and competencies integrated into complex activities. The scaling up in connections is based on the holomorphic structure of FNs. Namely, each FN is representable in the same way single functions are represented. Therefore, the connection of many FNs is a second-order FN (over the first-order FN). This mechanism can be iterated, again and again, by obtaining competencies built on competencies of the previous levels. A logical type is assigned to any function [41]. A function from type τ to type σ has type $(\tau \mapsto \sigma)$ (τ can have a Cartesian product type $\tau_1 \times \tau_2 \times \ldots \times \tau_k$). This typing mechanism extends naturally to FNs. In this sense, we have functional networks and competencies of high orders, where many FNs can be organized, with possible synchronization mechanisms, for effective coordination.

Knowledge is a high-order competence where many functional networks are coordinated and related to the language competence considered in the previous section. Complex concepts are obtained, as outlined, by "barycentric" syntheses of basic meanings. These syntheses correspond to concepts that become arguments of predications, implications, quantifications, abstractions, etc. In other words, the logical structure of ordinary language becomes a conceptual framework for the organization of complex concepts. This viewpoint explains the importance of language, not only for its primary communicative functionalities but also for its composition mechanisms applied to higher levels of representations.

Language logic [42] is based on functions at different logical levels, starting from individuals, truth values, and unary predicates. Predicates over predicates exist, called hyper-predicates. $Good(a)$ is a predication of property $Good$ to the individual a, while $Past(P)$ is the predication of the property $Past$ to a predicate constant P. Predicates are complemented by an operator, which we denote by _. Therefore, $Love\_b$ expresses the property of an individual loving b, whence $Love\_b(a)$ asserts that a loves b. The proposition "John is going home with a bike" can be represented by the following:

$$John(a), \widehat{Go}(P), Pres(P), Progr(P), Home(b), Bike(c), P\_Place(b)\_With(c)(a)$$

I will not go on to present a logical method of sentence representation [42,43]. We will limit ourselves to observing that using 20 logical operators, the meaning of any sentence can be obtained by the meanings of the single words. The same approach can be extended to complex semantic units constructed by functional networks. In this sense, ordinary language provides a logical structure that can be iterated at high-order functional networks.

One logical operator deserves special attention; it is the **predicative abstraction** operator. When it applies to a predicate $Pred$ of type $(ind \mapsto bool)$, where ind is the type of individuals and $bool$ is the type of truth values, it provides $\widehat{Pred}$ of type $((ind \mapsto bool) \mapsto bool)$, that is, a hyper predicate, for $Pred$. Namely, $\widehat{Go}(P)$ does not mean that P is going, but that P has the property of predicates that imply the predicate Go.

Very often, knowledge involves meaningful entities of a purely conceptual nature that shed new light on the whole semantic system of a cognitive system. Let us call these meanings **ideal elements** by extending terminology owed to David Hilbert [44] for the mathematical entities that completely redefined the previous conceptual organization of mathematical knowledge. Examples of these concepts are $\sqrt{2}$, the root square of 2, which is a number different from any fraction; the imaginary unit $i = \sqrt{-1}$, which is a number different from any real number; Euler's number e, which is the natural basis of Neperian logarithms; Newton–Leibniz differentials, on which mathematical analysis is based; or Cantor's transfinite ordinal and cardinal numbers. Analogous concepts in physics are waves, force fields, electrons, photons, subatomic particles, and quanta. Of course, lists of the same relevance are available in biology, medicine, economy, and any field of human knowledge. Any conceptual system changes when these ideal elements are elaborated. These elements are new meanings, making new the previous meanings from which they spring.

Ideal elements highlight two important aspects of cognitive dynamics, which are both related to Jean Piaget's investigation of "Genetic Epistemology" [45]—the Assimilation–Accommodation dialectic in cognition equilibrium and the Individual–Collective duality of knowledge evolution.

When a cognitive system acquires new information that provides a gain of knowledge, the acquiring system reacts by searching the minimum variation of its internal state that is compatible with the process of assimilation that internalizes the received information. When this is not possible, or only partially possible, the system promotes a change involving its internal structure. The former process is called assimilation, while the latter is called accommodation. In the assimilation–accommodation interplay, the best solution is a min–max optimization for which the maximum quantity of information has to enter the system, with the minimum change in its structure. An acquisition that does not alter the internal structure is informationally poor, while an acquisition requiring an inordinately intense change could require an excessive updating cost that prevents an effective realization of the change. Therefore, a cognitive system is subjected to opposite forces. A natural comparison can be performed with the acquisition of food, where the advantage of the ingested food has to be in equilibrium with its digestion cost. In the next section, we outline the role of an emotional system in pushing cognitive systems toward purposes and finalities.

The Individual–Collective duality in knowledge evolution means that individual cognitive systems cannot exploit their potentiality without an external world stimulating them where other cognitive systems similar to them interact with them in a continuous process of reciprocal influence, from which the cognition acquires a collective character within which each single system is a specific instance. This means that individual processes of cognitive maturation are a sort of holographic image of the collective definition of the "culture" of the society to which single systems belong. This vision is inspired by a principle held in biological evolution, according to which *ontogenesis is a recapitulation of filogenesis*.

2.5. Autonomy and Sensoriality

A dynamic is produced by a force orienting and driving it. Informational dynamics is based on data that transform according to rules into new data or new data organization.

In this way, a data sequence provides a computation as an information processing dynamic. The starting point is an initial configuration, and the final point is a configuration where no rule can be applied.

Cognitive natural systems are autonomous; that is, they work without any external intervention. In this case, any elaboration is generated by some internal mechanism. This means that, internally, a kind of disequilibrium has to be produced, with consequent dynamics, to transform the disequilibrium into a new form of equilibrium. Therefore, a cognition system can autonomously generate a process if a system is connected to it, providing disequilibrium conditions and driving it to equilibrium. Such a system is an **emotional system** supporting cognitive dynamics. An angry animal searching for food is a simple case to explain the situation. An angriness instinct produces the necessity for satisfaction, which activates the food search. Instinct satisfaction is the primordial mechanism activating natural cognitive systems. Such mechanisms are based on a sensoriality. Whence, a hierarchy of finalities is defined, going from instincts to more evolved forms of satisfaction and happiness. Finally, a very sophisticated organization of finalities is established, moving behaviors at all levels. Each person aspires to be happy, and this wish is the motor of their life.

This means that behavior autonomy is based on sensoriality and is oriented toward forms of happiness. But happiness implies the satisfaction of finalities in a condition of freedom. This situation is intrinsically the origin of conflicts when the finalities of some individuals are against others. Therefore, the freedom of the cognitive system is a critical aspect involving moral and ethical issues. Since the origin of cybernetics, the risk of thinking machines was considered by the founders of the digital era [46,47]. Today, their speculations are a reality and science cannot decline its responsibilities. In Wiener and von Neumann's time, an analogous debate was active about the nuclear bomb, and the positions of the two scientists were opposite. Artificial intelligence needs a consideration of the relationship between man, nature, and science, in a framework of social finalities and principles that cannot be derogated to the policies of small groups aimed at particular interests.

Emotional events can leave traces in synapse values that can remain even when the inputs producing a particular activation of a functional module are missing. In other words, emotions are related to processes of "tracing" with memories, which are very important for the efficiency of cognitive competencies.

2.6. Consciousness and Mind

At the highest level of cognitive competencies, there is consciousness. Consciousness can be represented as a map, essentially a function producing images of FNs, hence exploiting the competence of **imagination**. This meta-competence allows for evaluating possibilities, and it relates to behavior planning. It is interesting to consider that the notion of function, which was the origin of our discourse on cognition, is also its conclusion by exploiting its image, which, even in this mathematical terminology, provides the essential representational aspect of function. In this case, parts of a cognitive system are mapped in elements by producing a map of the whole system, giving an internal representation of competencies and concepts. This super-competence of a cognitive system provides what is intended as a **mind**, based on the internal perception of its reality as something that exists apart from any external reality by which and for which a cognitive system is designed in its original constitution. The possibility of virtualization is the basis of an autonomous existence motivated by the internal activity of the structure itself.

An important aspect of consciousness is the activity of continuous navigation in itself. Mechanisms of auto-attention are necessary for focusing on some parts of the cognitive system. This kind of navigation is an activity that any thinking person experiments with continuously. A reliable and uniform way of navigating in cognitive space has to be based on random mechanisms, such as Brownian motion or drunk walking. Many artificial cognitive systems use randomness to give effects similar to human behavior. Still, it is a necessary aspect of intelligence.

Intelligence is a complex notion that is difficult to define completely and precisely. For this paper, it is a quality of a cognitive system that possesses many efficient cognitive competencies. But surely, there is a crucial aspect of intelligence that is related to randomness. Namely, an intelligent system cannot be deterministic; otherwise, it is not creative, and creativity is an essential aspect of intelligence. Even in its etymology, it is based on the ability to capture the internal logic of phenomena. This ability requires abstract and general reasoning. However, a very subtle creative aspect of intelligent behavior relates to randomness, according to the following rule: *using random events to realize finalities*. Intelligent cognitive systems must include randomness in their structure to ensure efficient exploration of solution spaces in problem-solving situations.

Even if a cognitive system can manage a map of its competencies, many are unconscious, especially if they are at a high level in the functional hierarchy. This can be explained because knowledge about a functional level requires a higher logical level; therefore, since the cognitive hierarchy is finite, an uncovered portion in the consciousness mapping has to exist. However, the reason for the necessity of unconsciousness is also related to randomness. A complex cognitive system can do more than it is conscious to do. This incompleteness allows for openness and plasticity, which are characteristics very often mentioned as typical of cognition. A comparison occurs naturally with incompleteness in mathematical logic. We can formalize mathematical theories, but when theories are mathematically rich, there are theorems of the theory that we can never prove via the corresponding formal theory [48,49].

3. Conclusions

The functional perspective of cognition elaborated in this article clarifies many basic notions of the intellectual activity. The following short synthesis could help summarize the discourse developed so far. Learning means acquiring the function associated with some cognitive competence. Meanings are represented by semantic vectors in a multidimensional real space, and comprehension reduces data to the language meaning internally available in a cognitive system. Knowledge is a system of concepts that extend language semantics to a system of complex meanings based on primary meanings. A cognitive system can reach autonomy when coupled with an emotional system supporting it and defining a hierarchy of finalities, starting from the sensory level of pleasure–pain. The moral and ethical systems of values based on basic emotions produce internal mechanisms of evaluation and orientation for the whole cognitive system. Memory is a relational system of meanings associated with events that give traces to experiences of emotional importance. Consciousness consists of a (incomplete) cognitive map of the whole cognitive system.

A final lesson emerges from our discourse. Cognition is a functional activity, and artificial cognitive systems shed new light on natural cognition, becoming an instrument for understanding them and providing, at the same time, new possibilities for artificial intelligence. The limitations of the functional perspective provided are intrinsic to the mechanisms of mathematical representation with a reflexive character, where incompleteness and partiality are rooted in mathematical logic and the theory of computable functions [48,49].

Funding: This research received no external funding.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. McCulloch, W.; Pitts, W. A Logical Calculus of Ideas Immanent in Nervous Activity. *Bull. Math. Biophys.* **1943**, *5*, 115–133. [CrossRef]
2. Minsky, M. *Computation. Finite and Infinite Machines*; Prentice-Hall Inc.: Upper Saddle River, NJ, USA, 1967.

3. Brown, R.E.; Donald, O. Hebb and the Organization of Behavior: 17 years in the writing. *Mol. Brain* **2020**, *13*, 55. [CrossRef] [PubMed]
4. Hebb, O. *Organization of Behaviour*; Wiley: New York, NY, USA, 1949.
5. Hinton, G.E. Implementing semantic networks in parallel hardware. In *Parallel Models of Associative Memory*; Hinton, G.E., Anderson, J.A., Eds.; Lawrence Erlbaum Associates. 1981; pp. 191–217. Available online: https://www.taylorfrancis.com/chapters/edit/10.4324/9781315807997-13/implementing-semantic-networks-parallel-hardware-geoffrey-hinton (accessed on 1 December 2024).
6. Hopfield, J.J. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 2554–2558. [CrossRef] [PubMed]
7. Le Cun, Y. Une Procédure d'Apprentissage pour Réseau à Seuil Asymétrique. In *Cognitiva 85: A la Frontiere de l'Intelligence Artificielle des Sciences de la Conaissance des Neurosciences;* Proceedings of Cognitiva 85, Paris, France, 1985; pp. 599–604. Available online: https://www.academia.edu/3073721/Une_procedure_dapprentissage_pour_reseau_a_seuil_asymmetrique_a_Learning_Scheme_for_Asymmetric_Threshold_Networks (accessed on 1 December 2024).
8. Rosenblatt, F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychol. Rev.* **1958**, *65*, 386–408. [CrossRef] [PubMed]
9. Werbos, P. Beyond Regression: New Tools for Prediction and Analysis in Behavior Sciences. Ph.D Thesis, Harvard University, Cambridge, MA, USA, 1974.
10. Werbos, P. Backpropagation Through Time: What It Does and How to Do It. *Proc. IEEE* **1990**, *78*, 1550–1560. [CrossRef]
11. Euler, L. *Introductio in Analysin Infinitorum*, volume 1^o (1748). Euler Archive—All Works 101. Available online: https://scholarlycommons.pacific.edu/euler-works/101 (accessed on 1 December 2024).
12. Brown, T.B.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language Models are Few-Shot Learners. *NEURIPS* **2020**, *33*, 1877–1901.
13. Hinton, G.E. How neural networks learn from experience. *Sci. Am.* **1992**, *267*, 145–151. [CrossRef] [PubMed]
14. Hinton, G.E.; Osindero, S.; Teh, Y. A fast learning algorithm for deep belief nets. *Neural Comput.* **2006**, *18*, 1527–1554. [CrossRef] [PubMed]
15. Hinton, G.E. Learning multiple layers of representation. *Trends Cogn. Sci.* **2007**, *11*, 1527–1554. [CrossRef] [PubMed]
16. Hinton, G.E. To recognize shapes, first learn to generate images. In *Computational Neuroscience: Theoretical Insights into Brain Function*; Drew, T., Cisek, P., Kalaska, J., Eds.; Elsevier: Amsterdam, The Netherlands, 2007; pp. 535–548.
17. Hinton, G.E.; Krizhevsky, A.; Wang, S.D. Transforming auto-encoders. In *International Conference on Artificial Neural Networks Systems;* (ICANN-11); 2011; pp. 1–8. Available online: https://www.cs.toronto.edu/~bonner/courses/2022s/csc2547/papers/capsules/transforming-autoencoders,-hinton,-icann-2011.pdf (accessed on 1 December 2024).
18. Hinton, G.E. Where do features come from? *Cogn. Sci.* **2014**, *38*, 1078–1101. [CrossRef]
19. Kaplan, J.; McCandlish, S.; Henighan, T.; Brown, T.B.; Chess, B.; Child, R.; Gray, S.; Radford, A.; Wu, J.; Amodei, D. Scaling Laws for Neural Language Models. *arXiv* **2020**, arXiv:2001.08361.
20. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.
21. Kandel, E.R. *Search of Memory. The Emergence of a New Science of Mind*; W. W. Norton & Company, Inc.: New York, NY, USA, 2006.
22. Awret, U. Holographic Duality and the Physics of Consciousness. *Front. Syst. Neurosci.* **2022**, *15*, 685699. [CrossRef] [PubMed]
23. Cagnetta, F.; Petrini, L.; Tomasini, U.M.; Wyart, M. How Deep Neural Networks Learn Compositional Data: The Random Hierarchy Model. *arXiv* **2024**, arXiv:2307.02129v5. [CrossRef]
24. Ha, D.; Dai, A.; Le, Q.V. Hypernetworks. *arXiv* **2016**, arXiv:1609.09106v4.
25. Munkhdalai, T.; Yu, H. Meta Networks. *arXiv* **2017**, arXiv:1703.00837v2.
26. Sabour, S.; Frosst, N.; Hinton, G.E. Dynamic Routing Between Capsules. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017.
27. De Giorgi, E. *Selected Papers*; Dal Maso, G., Forti, M., Miranda, M., Spagnolo, S., Eds.; Springer: Berlin/Heidelberg, Germany, 2006.
28. Buzsaki, G. Neural syntax: Cell assemblies, synapsembles, and readers. *Neuron* **2010**, *68*, 362–385. [CrossRef]
29. Tomasello, R.; Garagnani, M.; Wennekers, T.; Pulvermüller, F.A. Neurobiologically constrained cortex model of semantic grounding with spiking neurons and brain-like connectivity. *Front. Comput. Neurosci.* **2018**, *12*, 88. [CrossRef] [PubMed]
30. Cybenko, G. Approximation by superposition of a sigmoid function. *Mat. Control. Signals Syst.* **1989**, *2*, 303–314. [CrossRef]
31. Hornick, K.; Stinchcombe, M.; White, M. Multilayer feedforward networks are universal approximators. *Neural Netw.* **1989**, *2*, 359–366. [CrossRef]
32. Mitchell, T. *Machine Learning*; McGraw Hill: New York, NY, USA, 1997.
33. Nielsen, M. *Neural Networks and Deep Learning;* 2013. Available online: https://jingyuexing.github.io/Ebook/Machine_Learning/Neural%20Networks%20and%20Deep%20Learning-eng.pdf (accessed on 1 December 2024).
34. Kelley, H.J. Gradient Theory of Optimal Flight Paths. In Proceedings of the ARS Semi-Annual Meeting, Los Angeles, CA, USA, 9–12 May 1960.
35. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533–536. [CrossRef]
36. Manca, V. Artificial Neural Network Learning, Attention, and Memory. *Information* **2024**, *15*, 387. [CrossRef]

37. Bahdanau, D.; Cho, K.; Bengio, Y. Neural Machine Translation by Jointly Learning to Align and Translate. *arXiv* **2014**, arXiv:1409.0473.
38. Skansi, S. *Guide to Deep Learning Basics. Logical, Historical, and Philosophical Perspectives*; Springer: Berlin/Heidelberg, Germany, 2020.
39. OpenAI, GPT4-Technical Report, ArXiv Submit/4812508 [cs.CL] 27 Mar (2023). Available online: https://arxiv.org/abs/2303.08774 (accessed on 1 December 2024).
40. Harris, Z.S. Distributional Structure. *WORD* **1954**, *10*, 146–162. [CrossRef]
41. Russell, B.; Whitehead, A.N. *Principia Mathematica*; Cambridge University Press: Cambridge, UK, 1910.
42. Manca, V. Agile Logical Semantics for Natural Languages. *Information* **2024**, *15*, 64. [CrossRef]
43. Manca, V. *Conversazioni Artificiali. Maieutica al Tempo Dei Chatbots*; Carte Amaranto: Pontremoli, Italy, 2024.
44. Hilbert, D. Über das Unendliche. *Math. Ann.* **1926**, *95*, 161–190. [CrossRef]
45. Piaget, J. *L'epistemologie Génétique*; Presses Universitaires de France: Paris, France, 1970.
46. von Neumann, J. *The Computer and the Brain*; Yale University Press: New Haven, CT, USA, 1958.
47. Wiener, N. Science and Society, Methodos. *Science* **1962**, *138*, 651. Available online: https://www.science.org/doi/epdf/10.1126/science.138.3541.651 (accessed on 1 December 2024). [CrossRef]
48. Church, A. *Introduction to Mathematical Logic*; Princeton University Press: Princeton, NJ, USA, 1956.
49. Manca, V. *Python Arithmetic. The Informational Nature of Numbers*; Springer: Berlin/Heidelberg, Germany, 2024.