UNIVERSITÉ DU QUÉBEC À MONTRÉAL

SYNTHETIC NEURO-COGNITION:
AUTOPOIETIC SEMIOTIC NEURON NETWORKS

THESIS

SUBMITTED

AS PARTIAL REQUIREMENT

FOR A Ph.D. IN COGNITIVE INFORMATICS

BY

PIERRE VADNAIS

DECEMBRE 2015

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

NEURO-COGNITION SYNTHÉTIQUE :
LES RÉSEAUX DE NEURONES SÉMIOTIQUES AUTOPOÏÉTIQUES

THÈSE

PRÉSENTÉE

COMME EXIGENCE PARTIELLE

DU DOCTORAT EN INFORMATIQUE COGNITIVE

PAR

PIERRE VADNAIS

DÉCEMBRE 2015

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

FOREWORD

This thesis is a transverse integration of cognitive sciences in search of artificial intelligence. We will talk about intelligence, thinking, mind, cognition as a concept representing simultaneously the common core of all these concepts in their respective domain, but also the specificities of each of these concepts, even if not obvious, outside their respective domain, underlining that they are all physically instantiated by a neuron network called the brain.

The transverse approach allows an intuitive progression jumping from one domain to another to verify the coherence of proposed hypotheses or to sidestep obstacles which cannot be addressed in a given domain. On the other hand, this intuitive progression is not really explanatory because it is too tortuous and often leads to dead ends. The explanation cannot follow multiple directions simultaneously. It then becomes necessary to reframe the explanation in more classical silos where the assertions, sometimes radical at first glance, can only be justified after establishing basic knowledge in other domains.

This thesis is part of a doctoral program in cognitive informatics in the informatics department. As such, it should be considered as an informatics project where philosophical, psychological or neurological contributions are attempts to analyze the system under consideration which happens to be the intelligence or, more concretely, the brain.

If, being computer specialists, some parts seem too philosophical, remember that, when working on the functional analysis of intelligence, the papers written by

psychologists and philosophers are probably the best "use-cases" available and they should be read as such.

If, some philosophers discover philosophically interesting sections, I will be flattered, but, let's be realistic, do not forget that our analysis must lead not only to an understanding of intelligence as an abstract concept, but to the realization of a non-biological model capable of emulating intelligent functionality in a concrete environment.

The functional analysis would of course be incomplete without a comparative study of the implementation materials; on the one hand the brain, a biological material described by neuroscientists and biologists, and on the other hand the computer, a programmable material able to simulate many very complex physical systems.

This transverse analysis will find its meaning only through an entanglement of links revealed by the functional analysis of an existing system and its known physical support, the brain, in order to reproduce it on a computer… like any good computerization project; no more, no less.

I would like to thank Professor Pierre Poirier, department of philosophy, UQÀM, and Professor Mounir Boukadoum, department of informatics, UQÀM, for their involvement as thesis director and co-director. Professor Poirier was especially patient, open and generously available in taking an engineer to the required level of understanding in cognitive sciences. I would also like to acknowledge the supporting contribution of the faculty members, and my fellow students, throughout this very enriching adventure in the Ph.D. program of Cognitive Informatics. The *Institut des Sciences Cognitives de l'UQÀM* was also, by its diversity, a very stimulating forum and a precious source of discoveries.

ABSTRACT

SYNTHETIC NEURO-COGNITION:
AUTOPOIETIC SEMIOTIC NEURON NETWORKS

This thesis was, from the beginning, guided by the interrogation: "Is strong Artificial Intelligence still possible?" We first identified what seem to be the biggest roadblocks in cognitive science, namely: the Symbol Grounding Problem and the Zero Semantic Commitment Condition. Then, we defined the problem through a functional analysis at the system level which took us from cognition to cognitive systems, from intelligence (or mind) to brains. The problem could then be transformed into a typical informatics project where a desired functional specification existing in a given (in this case biological) environment must be reproduced in a digital computer system. We reviewed the most probable required biological mechanisms (spiking neurons, synaptic plasticity, spike-timing dependent plasticity, Bienenstock-Cooper-Munro model, metaplasticity) and integrated them into well-encapsulated algorithms to produce a basic set of cognitive functionality. The resulting autopoietic semiotic network of artificial dynamic analog neurons can be developed into a representational structure following basic propositional logic and offers a framework to investigate Synthetic Neuro-Cognition, a bottom-up approach to empirically study the development of such representational structures and, perhaps, elaborate algorithms to automate it.

KEYWORDS : Artificial Intelligence, cognition, semiotics, autopoiesis, spiking neurons, doubleLIF, synaptic plasticity, metaplasticity, synthetic neuro-cognition.

RÉSUMÉ

NEURO-COGNITION SYNTHÉTIQUE : LES RÉSEAUX DE NEURONES SÉMIOTIQUES AUTOPOÏÉTIQUES

Cette thèse fut, dès le début, guidée par la question : « L'Intelligence Artificielle forte est-elle encore possible? » Nous avons d'abord identifié ce qui nous semblait être les principaux obstacles en science cognitive, soit le problème d'ancrage des symboles la contrainte d'absence de sémantique préalable. Nous avons, ensuite, défini le problème au niveau du système, ce qui nous a forcés à penser systèmes cognitifs plutôt que cognition, cerveaux plutôt qu'intelligence (ou esprit). Le problème s'est donc transformé en projet d'informatique typique où une fonctionnalité désirée, déjà instanciée dans un environnement donné (ici biologique), doit être reproduite dans un système d'ordinateurs numériques. Nous avons identifié les mécanismes biologiques ayant le plus de chance de répondre aux attentes (neurones impulsionnels, plasticité synaptique, plasticité déterminée par le temps d'occurrence des impulsions, modèle de Bienenstock-Cooper-Munro, métaplasticité) nous les avons intégrés dans des algorithmes adéquatement encapsulés pour reproduire un ensemble de fonctions cognitives de base. Le réseau sémiotique autopoïétique de neurones analogues dynamiques artificiels qui en résulte peut être édifié en structure représentationnelle en suivant une logique propositionnelle de base et, ainsi, offrir un encadrement pour l'investigation d'une Neuro-Cognition Synthétique, une approche ascendante pour l'étude empirique du développement de telles structures représentationnelles et, peut-être, l'élaboration d'algorithmes pour automatiser ce développement.

MOTS-CLÉS : Intelligence artificielle, cognition, sémiotique, autopoïèse, neurones impulsionnels, doubleLIF, plasticité synaptique, métaplasticité, neuro-cognition synthétique.

EPISTEMIC PRELUDE

Box 1 | The role of theory in science

Can theory be useful in neuroscience? We know that theory is very useful in the physical sciences and no one doubts the value of hypothesis-driven experiments in the biological sciences. It is when the connection between hypothesis and conclusion requires many steps that mathematical theories show their value. The biological sciences, we are sometimes told, are data-driven and too complex to allow for the effective use of mathematical theories. However, consider pre-Copernican astronomy. Ptolemaic astronomers introduced a variety of devices (including equants, deferents and, most famously, circles moving on circles called epicycles) to account for the positions of the planets against the fixed stars. By the time of Copernicus, astronomers were using up to 80 epicycles to fit vast quantities of data gathered over thousands of years of observation. Could the mediaeval astronomer have foreseen that the complexities of the planetary motions would all follow as a consequence of two postulates, namely Newton's second law of motion and Newton's law of gravitation? Of course success in the physical sciences is no guarantee that theory can succeed in neuroscience. However, it does suggest that large amounts of data do not preclude the possibility or usefulness of theory. Rather, we might say that such quantities of data make theory necessary if we are ever to order and understand them. Experiment winnows the possible hypotheses and theory narrows and focuses the experimental alternatives.

What is a good theory? The usefulness of a theory lies in its concreteness and in the precision with which questions can be formulated. A successful approach is to find the minimum number of assumptions that imply as logical consequences the qualitative features of the system that we are trying to describe. As Einstein is reputed to have said: « Make things as simple as possible, but no simpler. » Of course there are risks in this approach. We may simplify too much or in the wrong way so that we leave out something essential or we may choose to ignore some facets of the data that distinguished scientists have spent their lifetimes elucidating. Nonetheless, the theoretician must first limit the domain of the investigation: that is, introduce a set of assumptions specific enough to give consequences that can be compared with observation. We must be able to see our way from assumptions to conclusions. The next step is experimental: to assess the validity of the underlying assumptions if possible and to test predicted consequences.

A 'correct' theory is not necessarily a good theory. For example, in analysing a system as complicated as a neuron, we must not try to include everything too soon. Theories involving vast numbers of neurons or large numbers of parameters can lead to systems of equations that defy analysis. Their fault is not that what they contain is incorrect, but that they contain too much.

A theory is not a legal document and, in spite of occasional suggestions to the contrary, no scientist is in communication with the Almighty. Theoretical analysis is an ongoing attempt to create a structure — changing it when necessary — that finally arrives at consequences consistent with our experience. Indeed, one characteristic of a good theory is that one can modify the structure and know what the consequences will be. From the point of view of an experimentalist, a good theory provides a structure in to which seemingly incongruous data can be incorporated and that suggests new experiments to assess the validity of this structure. A good theory helps the experimentalist to decide which questions are the most important.

Cooper, L.N. and Bear, M.F. (2012)

INTRODUCTION

Since electronic was invented in mid XXth century, computers' performance incessantly improved in the execution of tasks which, until then, were only accessible to human intelligence. In accounting, engineering, astrophysics, medicine or many other domains, every day new computer applications seem to challenge human supremacy in the solution of more and more complex problems. Thanks to their speed and precision, these machines often exceed human capacities and could pretend to superior intelligence.

Yet, every day also, these "brilliant" machines demonstrate their ineptitude and their clumsiness when times come to face changes, sometimes minimal, in their environment. Why such a paradox?

After careful consideration, it looks like these major computer realisations lie in the extension of human intelligence without proper genesis (i.e. the solution method is generated by humans and its application, as complex as it can be, is left to the machine). Programming allows transposing some human knowledge (declarative knowledge) into the machine which can then use them with speed, perseverance and precision to ever more complex problems. The computer is unquestionably the generalization expert of known solutions to similar problems whatever their number or complexity, but it fails miserably when the problem is new; it is not capable of invention.

To elucidate this paradox, we could attempt to narrow the definition of intelligence from the suggestions of experts in the field, but, cognitive science being highly multidisciplinary, even the use of the word intelligence may seem tendentious. The philosophers prefer to speak of mind to study cognition. The word intelligence is

accepted by psychologists in the study of behavior, but generally refers to different degrees of human intelligence. Even though Descartes' mind-body duality has fewer and fewer supporters, few are willing to identify brain with intelligence or mind, especially as physiological studies of neurons are far from filling the now famous Leibnitzian gap.

> Moreover, it must be confessed that perception and that which depends upon it are inexplicable on mechanical grounds, that is to say, by means of figures and motions. And supposing there were a machine, so constructed as to think, feel, and have perception, it might be conceived as increased in size, while keeping the same proportions, so that one might go into it as into a mill. That being so, we should, on examining its interior, find only parts which work one upon another, and never anything by which to explain a perception. Thus it is in a simple substance, and not in a compound or in a machine, that perception must be sought for. Further, nothing but this (namely, perceptions and their changes) can be found in a simple substance. It is also in this alone that all the internal activities of simple substances can consist.
>
> —Gottfried Leibniz, The Monadology (1698)

Is it no wonder that, nowadays, IT experts (still) hope to replicate in "machines" the mental abilities of the human mind? In fact, everything (re)started when, in 1950, Alan Turing asked the question: « *Can machines think?* (Turing 1950).

The first chapter of this thesis, paraphrasing Turing, will discuss the possibility of (strong) artificial intelligence by studying the various hypotheses inspired by, or implied in, his computability thesis (also known as the Church-Turing thesis). Having defined the conditions necessary to generate this strong artificial intelligence, the second chapter will analyze the functionality of biological neurons to identify the mechanisms needed to support natural intelligence and will propose a model of artificial neurons including equivalent mechanisms. A third chapter will present a series of simulations based on this model of artificial neurons to observe its behavior in an (relatively friendly) environment offering various stimuli. Finally, the fourth

chapter will analyze the relevance and validity of the many assumptions used to justify the model.

It is important to note, before we start, that even if we are talking about (strong)[1] artificial intelligence and Turing, we are not seeking in any way to pass the Turing test, which requires a fully developed intelligence capable of verbal communication at an advanced formal level. Rather we are at a preverbal and preconscious level corresponding to what Piaget identified as the sensorimotor stage (Piaget 1936). We believe that this step is as important for the understanding of artificial intelligence and its development as it is, according to Piaget, to understand the ontogenetic development of human intelligence.

The research project focuses on two complementary theories. The first, under the cognitive sciences, inspired by Newell and Simon states that: *only autopoietic semiotic systems have the necessary and sufficient means for general intelligent action.* The second, neurocomputational, offers a concrete realization of the first using *a model of dynamic, analog and asynchronous neurons, which associated in networks are sufficient to simulate an autopoietic semiotic system.* As we can only verify the intelligence by observing behaviors, we will build an experiment around the psychological corollary resulting from these two theses and stating that *such autonomous networks are capable of seemingly intentional and decisional behaviors.*

---

[1] Strong AI is an expression from Searle (1980) who, by his experience of the Chinese room (more on this later), says, imagining himself in the role of the machine, that a machine capable of reading questions in Chinese and responding in Chinese (an intelligent human's behavior) does not yet understand anything about the conversation. Searle assumes that it is possible to write a "recipe", an algorithm, to read and speak Chinese without understanding Chinese (reading and speaking without understanding, not write the "recipe" without understanding). In other words, a machine is able to reproduce, and even learn, any behavior (even human behaviors) observable, analyzable and algorithmizable (even discover new evidence of complex mathematical theorems). Strong AI should not only be able to learn, but also to understand. Utopia for many, yet it remains an option for the true materialists among us.

CHAPTER I

Is (strong) artificial intelligence possible?

1.1 - Recent history

Recent history of computing took off in the early twentieth century with the convergence of mathematical works such as the theses of Church (1932, 1936a, 1936b), Turing (1936, 1947, 1950) and Post (1936, 1943), the proofs of Gödel (1931) and Kleene (1952) and the algorithms of Markov (1960).

1.1.1 - The Turing Machine

This convergence was initiated by David Hilbert's (1900) program of formalization of mathematics which led to Gödel's incompleteness theorems (1931). First Church (1936ab) attacked the Entscheidungsproblem (the problem of undecidability) using the lambda-calculus based on recursion and confirmed Gödel's theorem whereby, in a symbolic logic system, it is impossible to find an *effective* method for determining whether a proposition P is verifiable in this system. Meanwhile, Turing (1936) reached the same conclusion using a mechanical conception of computability now known as the *Turing Machine*. There are several variations of the definition of a Turing machine (TM) and the original (Turing 1936) is neither the easiest nor the most obvious. Wikipedia[2] provides us with a clear and precise definition of this machine:

---

[2] http://en.wikipedia.org/wiki/Turing_machine

More precisely, a Turing machine consists of:

1. A tape divided into cells, one next to the other. Each cell contains a symbol from some finite alphabet. The alphabet contains a special blank symbol [...] and one or more other symbols. The tape is assumed to be arbitrarily extendable to the left and to the right, i.e., the Turing machine is always supplied with as much tape as it needs for its computation. Cells that have not been written before are assumed to be filled with the blank symbol. In some models the tape has a left end marked with a special symbol; the tape extends or is indefinitely extensible to the right.
2. A head that can read and write symbols on the tape and move the tape left and right one (and only one) cell at a time. In some models the head moves and the tape is stationary.
3. A state register that stores the state of the Turing machine, one of finitely many. Among these is the special start state with which the state register is initialized. These states, writes Turing, replace the "state of mind" a person performing computations would ordinarily be in.
4. A finite table [...] of instructions [...], given the state[...] the machine is currently in and the symbol[...] it is reading on the tape (symbol currently under the head), tells the machine to do the following in sequence [...]: Either erase or write a symbol [...], and then Move the head ([...]'L' for one step left or 'R' for one step right or 'N' for staying in the same place), and then Assume the same or a new state as prescribed [...].

Note that every part of the machine (i.e. its state, symbol-collections, and used tape at any given time) and its actions (such as printing, erasing and tape motion) is finite, discrete and distinguishable; it is the unlimited amount of tape and runtime that gives it an unbounded amount of storage space.

Though still abstract with its infinite tape, the Turing machine is clearly a concrete approach to computability based on mechanisms (like any machine by definition). These physical mechanisms relate to causality rather than to implication as would any formal logical approach like Church's thesis based on recursion. Using mechanisms suggests that the change of state and the writing of symbols are caused by the reading of a symbol in a given state. This set of mechanical rules forms an algorithm.

Considering the equivalence of the two theories, nowadays, we usually refer to the *Church-Turing thesis*, but, here, as we will focus on specific aspects of Turing's approach, we present the computational hypothesis in terms of a Turing machine:

H1 - Computational Hypothesis (or computational axiom)
Any algorithm may be performed by a Universal Turing Machine[3].

With the development of the transistor in the 1940s and the rapid development of computers thereafter, there can be no doubt anymore, the Turing machine is real[4], this is no longer a hypothesis. Computer science is based on this axiom and produces daily increasingly powerful algorithms. Markov (1960) has, in a way, generalized Turing's approach in vectorizing its symbols and atomic states. The markovian state is always finite, but it is multivariate allowing a combinatorial explosion of representations driven by equally multivariate inputs.

1.1.2 - The Turing test

The Turing machine, although very abstract, is not a pure invention of the mind. To develop it, Turing was inspired by a "computor" in the most human sense of the word, that is to say, a person who "computed" with the help of a pencil and a sheet of paper. Therefore we find in the machine very concrete and physical elements such as paper, reading and writing, symbols, etc. His objective was to mechanize the work of the mind of such a "computor".

---

[3] Any set of TM's can be simulated by a single more complex TM. We call Universal TM the TM able to simulate the set of all TM's and therefore any simpler TM. A Universal TM is not necessary for any algorithm taken individually, but it is sufficient for the most complex of them, since it is sufficient for the set of all (ignoring the material and temporal constraints).

[4] "Real" obviously implies that its instantiation in a computer makes the TM prone to hardware constraints (memory) and time constraints (speed of execution).

Being a good mathematician, Turing (1950) did not hesitate to generalize the experience: if the machine can "compute" like a human, it can think like a human. This was enough to awaken old dreams of artificial intelligence, but the algorithm of intelligence had yet to be defined.

Adept at Extreme Programming (Beck 2000) long *avant la lettre*, Turing proposed a functionality test, or more precisely, an acceptance test to determine if the goal had been reached. Drawing on a popular game of his times, the imitation game, Turing (1950, §1.) wrote:

> [The imitation game] is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. He knows them by labels X and Y, and at the end of the game he says either "X is A and Y is B" or "X is B and Y is A." [...] It is A's object in the game to try and cause C to make the wrong identification.
> In order that tones of voice may not help the interrogator the answers should be written, or better still, typewritten. The ideal arrangement is to have a teleprinter communicating between the two rooms. Alternatively the question and answers can be repeated by an intermediary. The object of the game for the third player (B) is to help the interrogator. The best strategy for her is probably to give truthful answers. She can add such things as "I am the woman, don't listen to him!" to her answers, but it will avail nothing as the man can make similar remarks.
> We now ask the question, "What will happen when a machine takes the part of A in this game?" Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman?

Very powerful acceptance test, as Harnad (1992) points out; the Turing test is however of no use when it comes to determine progress in instantiating the functional features because the specificity of the test, which, paradoxically, is its generality, is implicit in the behavior of the interrogator and can only be explicited by fully solving the original problem. In the good old imitation game, the three participants are human

and the three roles are interchangeable. So, any machine capable of passing the Turing test would naturally be able to hold any of these three roles.

Turing's experience, first mathematical, has allowed the elaboration of a machine capable of computation like humans. His computational thesis argued that the machine was able to run any algorithm devised by humans since it was able to perform all computable functions. This was the basis of computing.

His test was going much further; it asserted that the set of all computable functions was sufficient to simulate thought. Were we to give the machine all known human algorithms, it could think like a human. The approach was consistent with the method used by Turing to demonstrate the computability of π although it would have been necessary to show that the set of all computable functions could be regarded as a convergent series. This was the origin of research in artificial intelligence (AI) and the hatching of cognitive science (SC). Both schools were born a few months apart. The pioneers of AI mat Dartmouth College in 1956 at the invitation of McCarthy al. (1955) « to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. »

The hypothesis was clear: « Intelligence can be simulated by a machine ». Behind this hypothesis from the fathers of AI, there was identification, conscious or not, of the ontology of intelligence with that of computing machines. In other words, intelligence « processes data » following « defined rules » in order to g« results ». This was the basis of computationalism.

The description of the « rules » for the processing of « data » by intelligence was not nearly as simple as the 1955 invitation implied. For over a century, psychologists have tried to define a scientific approach to the operation of intelligence. From

Wundt's experimental psychology to John B. Watson's (1913) behaviorism through Freud's psychoanalysis and Pavlov's classical conditioning, the functionality of the intelligence was still far from the precise description required to allow computer simulation (« ... every aspect of learning or any other feature of intelligence can in principle be **so precisely described** that a machine can be made to simulate it »). And yet, behaviorism, which prevailed as the dominant theory in psychology at time of the meeting in Dartmouth, would be challenged by Noam Chomsky (1959) in his review of B.F. Skinner's book « Verbal Behavior ».

Especially that some participants, and even the organizers, came to the conference with a very different approach to simulating the operation of intelligence. As stated in the invitation (McCarthy al., 1955)

> M. L. Minsky, Harvard Junior Fellow in Mathematics and Neurology, ... has built a machine for simulating learning by nerve nets and has written a Princeton PhD thesis in mathematics entitled, 'Neural Nets and the Brain Model Problem' which includes results in learning theory and the theory of random neural nets.

The door was already open to a connectionist approach based on neurology and biology with reference to the work of Donald O. Hebb (Hebbian learning).

Both schools have, since that time, cooperated to achieve converging objectives. Proponents of AI used mathematical sciences and rising computer sciences to produce an artificial intelligence, while experts in Cognitive Sciences tried to apply the emerging paradigm of *mechanical* information processing to unify the philosophical and psychological knowledge. However, it is not these differences in objectives or origins that lead to the greatest conflicts, but rather the two basic assumptions mentioned above regarding: 1) the simulation of the mind and 2) the simulation of the brain.

1.1.3 - Computationalism

By computationalism we mean Putnam's functionalism (1965) and Fodor's representationalism (1975). In cognitive science, functionalism is the general ontological thesis that mental states are functional states. Although it is not included in the functionalist theory, it is generally understood that it is functional states of the brain[5]. When we want to explain this assumption, we say that functionalism was added a token psychoneural identity thesis (Bickle 1998), that is to say that each occurrence of a mental state is a state of the brain. Putnam's functionalism is said "machine functionalism" or "Turing machine functionalism" since, according to him, the functional states that are mental states are computational states described as those of a Turing machine, that is to say with reference to the symbolic inputs and outputs, and to other computational states that a mental state is bound to by the instructions in the instruction table (i.e. the program) in the Turing machine. Since the symbols in a Turing machine are representations when the instructions are interpreted as computations, Putnam's functionalism is already representational. Fodor (1975) takes up and strongly defends this representationalist functionalism[6] and adds the thesis that each representation has roughly the syntactic form of a sentence of a natural language, particularly the predicative form (subject- predicate) of such sentences.

So defined, computationalism includes much of cognitivism. In Putnam's and Fodor's times, cognitivism excluded neuroscience, and it was believed that it was sufficient to understand the program (the instruction table in the Turing machine) since we knew that if there was one physical implementation of the Turing machine

---

[5] It is relevant today to clarify this addition, because various philosophers defend functionalism, but reject the thesis of token psychoneural identity to replace it by a thesis of token psychophysical identity: each brain state is a physical state, but not necessarily a physical state of the brain (could include physical states of the body or of the environment (Clark, 2008).

[6] There are forms of non-representationalist functionalism; see Armstrong 1968.

with this program, there were infinitely many and they did not all have to be brains. So, the basic idea was reversed; rather than having the machine thinking like humans, it would be the humans who would think like machines: the theory of information processing based on inputs, outputs and manipulation of symbols. This reversibility was explicitly stated by Newell and Simon (1976) in their physical symbol systems hypothesis: « A physical symbol system has the necessary and sufficient means for general intelligent action. »

We can therefore summarize the computationalist approach by a psychological hypothesis:

> **H2a - Psychological hypothesis** (which we will not accept)
> Intelligence is directly algorithmizable.

Searle (1980) harshly attacked computationalism and symbolic systems with a thought experiment, now famous under the name of "*Chinese Room*", where he shows that the manipulation of symbols does not imply any understanding of the symbols and even less of the handling itself.

Harnad (1990) redefined the impasse as the symbol grounding problem. In a symbolic system, all symbols can only be defined from other symbols and no symbol is, for the system, grounded in the real-world experience.

It is therefore not surprising that the computationalists (including Fodor, Chomsky, Pinker) have often been innatist or nativist; otherwise, where could the symbols... or grammar come from?

1.1.4 - Connectionism

In parallel, another approach was growing based on a biological hypothesis:

H2b - Biological Hypothesis

The brain is algorithmizable.

Connectionism opted to algorithmize the brain, trying to model the neural mechanisms. Several generations of models have followed since McCulloch and Pitts's binary neurons (1943), to the recent spiking neurons including the rate neurons. It is important to note Rosenblatt's perceptron (1957) which has had some success until Minsky and Pappert (1969) demonstrated the limits of its linearity. The ensuing ardor-cooling greatly favored the psychological hypothesis relatively to the biological hypothesis. Rumelhart, McClelland and the PDP Research Group (1980) revived interest with a research program using multi-layer perceptrons with a learning rule based on error back-propagation. The computational difficulties of perceptrons did not facilitate the acceptance of connectionism, but the reluctance of many mainly resided, and still reside, in the reduction of thought, intelligence, mind, cognition into simple cellular mechanisms.

Connectionism still keeps its neuro-biological inspiration in spite of its mathematical appearance. Maass (1997) speaks of a second and a third generation of models that attempt to capture more accurately the behavior of biological neurons. The inspiration for the second generation goes back to the first steps of neuronal responses to electrical stimulation represented by frequency to current curves (see **Error! Reference source not found.**[7]) showing the spiking frequency (in Hz) of pulses (action potentials) as a function of injected electric current (in mA). This approach helps to understand the neural signal as a continuous analog signal.

---

[7] Figure I.1 taken from Eliasmith and Anderson (2003 p34), presents three typical response curves published by McCormick al. (1985) representing the relationship between the frequency of action potentials and the intensity of the current injected into a dendrite.

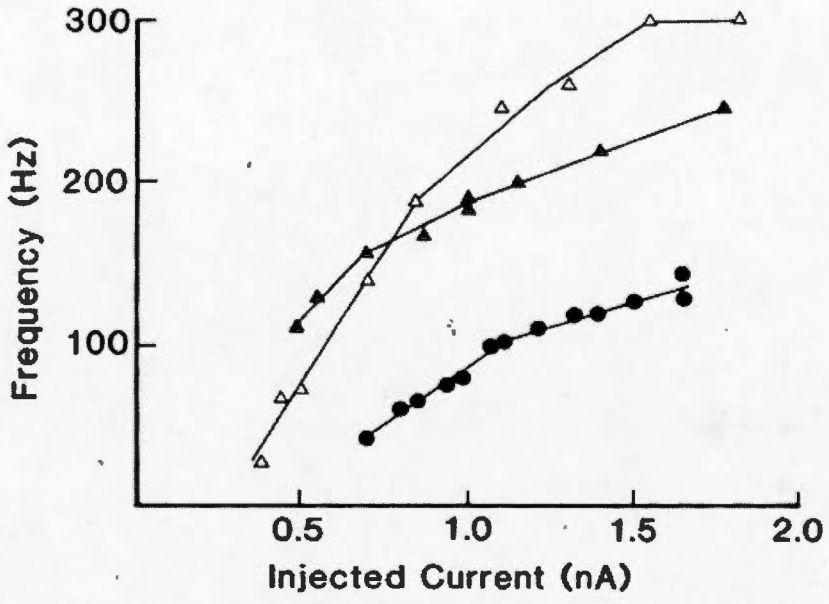


Figure I.1 - Rate neurons' frequency-current curves (see footnote 7)

On the other hand, as Maass points out, some (Perrett, Rolls, and Caan, 1982 and Thorpe and Imbert, 1989) have demonstrated that the computation time of pulse rates was not compatible with the ultra-fast response (20-30 msec) of some complicated cortical networks (ten synaptic connections chain) including neurons at relatively low frequencies (<100 Hz). Maass provides also a long list of experimental evidence indicating that the timing of each pulse taken individually may be important in the encoding of information. The third generation models, called spiking neurons, include these biological considerations by using space-time differential equations in modeling the dendritic integration. These differential equations necessarily introduce the dynamic aspect essential to cognition according to van Gelder (1996). We will return to these issues in the second chapter where we will talk about modeling the brain.

We will keep in mind that connectionism is fundamentally computational in the sense of the Church/Turing thesis although it is radically opposed to the abstract computationalism of the functionalists and representationalists of putnamian and fodorian influence. We will also remember that the continuous and analog appearance of second generation models may provide some answers to the question of symbol grounding and that it is important not to sacrifice this advantage to take into account the individual dynamics effect of each pulse in our third generation models. We will try to justify the merits of these allegations in the second chapter where we will explain more precisely the mathematics and semiotic aspects of biological and artificial neurons.

Connectionism, situated embodiment and biologically plausible dynamics are based on a strongly constructivist conception of brain development and therefore, for their proponents, of thought, intelligence, mind, cognition. Elman al. (1996) redefined nativism according to this conception and state (p361):

> [...] representational constraints (the strongest form of nativism) are certainly plausible on theoretical grounds, but the last two decades of research on vertebrate brain development force us to conclude that innate specification of synaptic connectivity at the cortical level is highly unlikely.

When Elman al. (1996) talk about architectural nativism, they refer to the constraints that heredity imposes directly on brain organization. We would go further and say that these genetic constraints are imposed indirectly on brain development by the body architecture, which itself is defined by genes.

Long before them, Piaget (1936) also observed and analyzed the development of pre-verbal intelligence by focusing on the importance of what is acquired in relation to the innate. Yet it must be admitted, with Elman al., that, if the representations are

(most likely) not innate, the constraints controlling the development of the brain are fully defined[8] by genetic inheritance (by the set of sensors and actuators[9] at its disposal including their initial settings) and by the biological rules of operation, learning and development.

However, the embodied and situated connectionism is not sufficient to completely solve the symbol grounding problem. Floridi (2011b section 6.2) introduces a Zero semantic commitment requirement, or Z condition, stating that:

> The challenge posed by the SGP [symbol grounding problem] is that
> a. No form of innatism is allowed; no semantic resources (some virtus semantica) should be magically presupposed as already preinstalled in the AA [artificial agent]; and
> b. No form of externalism is allowed either; no semantic resources should be uploaded from the 'outside' by some deus ex machina already semantically proficient. [...]
> c. The AA may have its own capacities and resources (e.g. computational, syntactical, procedural, perceptual, etc., exploited through algorithms, sensors, actuators, etc.) to be able to ground its symbols. (Floridi 2011b p137)

Floridi's entire chapter 6 demonstrates that the proposed solutions to the symbol grounding problem, whether representationalist as in Harnad (1990), Mayo (2003), Sun (2000), semi- representationalist as in Davidson (1993), Vogt (2002a, 2002b), Rosenstein and Cohen (1998), or non-representationalist as in Brooks (1990), Billard and Dautenhahn (1999), Varshavskaya (2002), do not pass the test of the Z condition.

---

[8] Waddington (1956) would have said "channeled".

[9] Although it is not quite right, the terms "actuators" and "effectors" are often used interchangeably. Effectors affect the environment (e.g. an arm) while actuators are the elements (e.g. muscle, motors) enabling the effector to execute the action. Effectors may have multiple degrees of freedom while actuators generally refer to a single degree of freedom. We will use actuators, in parallel with sensors, because we are interested in decomposing action, and perception, to the cellular level.

1.1.5 - The computationalism - connectionism cleavage

In summary, we have attempted to show that connectionism, as well as computationalism, draws its source from the axiom of computationality, thus from the Church-Turing thesis. We consider that connectionism does not reject functionalism, nor representationalism, nor symbolism, nor cognitivism, but it differs mainly from computationalism by the object of its modeling which is at the physical level, the concrete and mechanical level, of the brain, rather than at the level of the mind, the thought, the intelligence, the cognition, which is much more abstract and phenomenological. This computationalism-connectionism cleavage, which is the contemporary version of the Leibnitzian gap, is at the heart of the difficulties in unifying the cognitive sciences and must be bridged to enable them to achieve their ultimate goal and produce artificial intelligence.

1.1.6 - Argument summary

Table I.1 - Is (strong) artificial intelligence possible?

P1[10]. All computable functions can be executed by a Turing machine (TM) (Turing's thesis)
P2. Mental functions are computable functions (computationnalism)
C1. Hence, mental functions can be executed by a Turing machine (Turing's test, Putnam's functionalism, Fodor's representationalism)

[10] Pn = Proposition n. Cn = Conclusion n. On = Objection n

C2. Physical symbol systems have the necessary and sufficient means for general intelligent action (Newell and Simon's Physical Symbol Systems hypothesis)

O1. The physical symbols manipulated by the system are not understood by the system (Searle's Chinese room argument [CRA])

O1 rephrased. The physical symbols manipulated by the system are not grounded in reality (Harnad's symbol grounding problem [SGP])

P3. Hybrid systems can use embodied connectionist sub-systems situated in a given reality to ground symbols in that reality (Harnad's solution to SGP)

P3 is explicitly or implicitly supported by numerous connectionist approaches such as Brooks', Pfeiffer's, Steels', etc..
P3 is compatible with Smolensky's thesis; cognitive systems must be explained at a sub-conceptual level by sub-symbolic systems.

O2. Situated embodied connectionist sub-systems cannot develop their own semantic without external contributors such as innateness or programming (Floridi's Z condition - « zero semantic commitment condition ») (P3 is necessary but not sufficient)

C3. Hence, to have general intelligence, a system must :

1. use physical symbols grounded in reality and
2. have no external semantic contribution, neither innate, nor programmed.

1.2 - Fundamental hypotheses

We cannot talk seriously about cognition and knowledge, computing and information processing, without having an idea of what information is. But where does information come from? We believe that there are some basic assumptions implicitly hidden behind the computational axiom.

1.2.1 - Physicalism

If we reject out of hand the Cartesian substance dualism, we can, actually we must, posit a highly materialistic ontological assumption.

H3 - Ontological hypothesis
Everything is matter/energy.

To include energy and wave phenomena, we prefer to speak of physicalism rather than materialism; a physicalism that focuses on the physical properties and conceives objects as instantiations of a set of physical properties.

Each element of matter/energy (each atom) has its own properties. There are intensive properties, as the type of atom and the energy levels, and extensive properties which depend on the number of bonded atoms and on the intensive properties of each of these atoms. Modern quantum physics deals with particles that make up the atoms and their specific properties, but to talk about cognition, the molecular level is well below the levels commonly used and appears to us as being sufficient, but also necessary, to understand the concepts of object, signal and information.

1.2.2 - Information

Without matter/energy, there is nothing, no space, no time, nothing. But as soon as there is matter/energy, information appears. Space-time comes from a differentiation between elements of matter/energy and any difference is a piece of information. This is what Floridi, in his Semantic Conceptions of Information (Floridi 2011a), refers to as the « diaphoric[11] definition of data ». According to Floridi (2011b p85-86),

> the diaphoric definition of data can be applied at three levels:
> 1 Data as diaphora *de re*, that is, as lack of uniformity in the real world put there. [...] They are pure data or proto-epistemic data, that is, data before they are epistemically interpreted [...].
> 2 Data as diaphora *de signo*, that is lack of uniformity between (the perception of) at least two signals such as [...] a variable electrical signal in a telephone conversation, or the dot and the line in the Morse alphabet.
> 3 Data as diaphora *de dicto*, that is lack of uniformity between two symbols for example the letters A and B in the Latin alphabet.

Any difference in any of a thing's (*res*) physical properties is a *de re* (about this thing) data. Note that by *thing* we mean any object, from the simplest atom to the universe in all its complexity, and this object is physical if it has physical properties. A *de signo* (about the signal) data is a difference in the signal's (*signum*) physical properties, thus a *de re* data about this thing called the signal, and finally a *de dicto* (about the word) data is a difference in the word's (*dictum*) physical properties, thus a de re data about this thing called the word. So, why three categories? What is so special about signals and words?

---

[11] From diaphora (*διαφορα* in greek) meaning difference.

Usually, we are not interested in the *de signo* data of signals because they carry more data, *de re* data, about other things. To take Floridi's example, « the variable electrical signal in a telephone conversation » is not interesting as a variable electrical signal, but because it hides at another level (because it contains or bears) a conversation. And this conversation is not interesting for the *de dicto* data of the words that make it up, but for the *de re* data of the things that those words represent. The word is out, "represent"; *de signo* and *de dicto* data are representations of other *de re* data about other things whether or not directly visible in the immediate spatiotemporal reality. In fact, the spoken or written word is a sound or visual reification of mental object representations; this mental object reified thereby becoming a noticeable physical signal by others. In this thesis, we will only accidentally mention *de dicto* data (words), since we are interested in the preverbal sensorimotor intelligence. We will focus mainly on signals and on the two levels of *de signo* data (the properties of the signal) and *de re* data (the meaning of the sign carried by this signal).

Floridi (2011a) considers that « Data [...] can have a semantics **independently** of any informee. » It is not intended to define where this meaning comes from, only that it exists independently of the informee. To support this assertion, he recalls that:

> [b]efore the discovery of the Rosetta Stone, Egyptian hieroglyphics were already regarded as information, even if their semantics was beyond the comprehension of any interpreter. The discovery of an interface between Greek and Egyptian did not affect the semantics of the hieroglyphics but only its accessibility.

However, he is reluctant to support, « *the stronger, realist thesis, supported for example by Dretske [1981], according to which data could also have their own semantics independently of an intelligent producer/informer.* » Inspired by Barwise and Seligman (1997) and Dretske (1981) he defines environmental information in

relation to an observer, as follows: « Two systems *a* and *b* are coupled in such a way that *a*'s being (of type, or in state) *F* is correlated to *b* being (of type, or in state) *G*, thus carrying for the information agent [(the observer)] that *b* is *G*. »

Therefore, he concedes no semantics, especially no semantic processing (interpretation), to environmental information making it a non-semantic semiotic system based strictly on *de signo* data. This environmental information is a potential information as it becomes meaningful only for the observer able to interpret it. However, because it does not suffer any interpretation, this potential information is *de facto* essentially true.

> Plants (e.g., a sunflower), animals (e.g., an amoeba) and mechanisms (e.g., a photocell) are certainly capable of making practical use of environmental information even in the absence of any (semantic processing of) meaningful data.

The light, generated by the sun, affects the plants and photocells by environmental coupling. This coupling is not merely correlative but essentially causal. An observer may confuse the two, but physically only the causal coupling exists, whilst the correlation is pure semantics.

Isn't this causality sufficient to make sense of the received (or perceived) information? The practical side of the resulting usage is certainly a viewer's interpretation, but it is not without giving some meaning to the entire causal chain leading from the data transmitted to some useful action. We will talk about this utilitarian semantic when we will discuss H8 and H9 towards the end of this chapter.

Leaving the observer out of the equation (who needs a homunculus?), it is possible to reformulate the definition of environmental information in terms of causality: *two*

systems a and b are ***causally*** *linked if, first, they are of* ***compatible*** *types and second, that a is (in the state) F* ***as a physical reaction to*** *b being (in the state) G.*

This causal and utilitarian semantics of potential information is the basis of our second fundamental assumption:

H4 - Semiotic hypothesis

As soon as there is matter/energy, there is potential information, therefore signs.

This hypothesis does not, in anyway, contradict the previous one (H3) as it makes information a property, in fact all the properties, of matter/energy. One could speak of supervenience of information on matter/energy. So, this is not substance dualism; there is only one substance which we call matter/energy. This is not property dualism, but rather communication of properties by semiotic transmission.

So, there are two distinct parts to information: the causal part, or *de signo* information, at the signal level and the semantic part, or meaning, of the sign carried by this signal. The first is physically transmitted from one system *a* to a second system *b* as explained by Dretske, while the second depends on interpretation requiring a move to a higher level of abstraction. We will use some basic notions of C.S. Peirce's semiotics to mark the difference, but, for now, we are concerned by the causal level of *de signo* information and, as mentioned above, we will return later on semantics.

Peirce's theory of signs (1897, 1903), spoke of icons, indices and symbols. We will use the same words to explain the difference we introduced between causality and semantics. Without rejecting, without even questioning, the definition given by Peirce to the terms 'icon' (« firstness »), 'index' (« secondness ») and 'symbol' (« thirdness

»), we will change the order to put emphasis on a perspective of spatiotemporal proximity in the reference to reality.

We will place the index first because it comes from an immediate physical reality. Take, for example, the smell of the cheetah that repels the gazelle. The smell, present in the immediate spatiotemporal reality, causally triggers the action without reference to the well-hidden cheetah in the tall grass. The *de signo* data (the smell) acts directly. Piaget (1936) used the example of floorboards creaky as the mother approached the cradle to take the child and feed him. The link is still causal but indirect; this is obviously not the creaky floorboards that nourishes the child, but the mother is causing both effects sufficiently correlated to seem causally related... especially for a baby. The index is predictive rather than representative; the smell triggers the flight because every time the smell was perceived, the herd fled. The creaking announces feeding because the latter is almost always preceded by a creaking. The smell and the creaking are indices inasmuch as they are fully present in the immediate reality. In the same vein, one could speak of Pavlov's dog salivating at the sound of the bell or Bickerton's (1990) vervet monkeys which respond differently to three specific sounds produced by conspecifics when seeing an eagle, a snake or a leopard. The vision of a given predator creates a specific flight movement; it is this movement that modulates breathing in an equally specific sound and hearing the sound triggers a similar flight movement; the flight movement is not connected to the predator, but only the sound; the latter is connected to the predator, but only causally, without representation of the predator in the listener's mind.

The icon is second. It usually connects two physical objects by their resemblance. By resemblance we mean the common subset (intersection) of both sets of properties specific to each of the two objects. The icon acts as an index relative to its referent; a physical object in the immediate reality recalls a second similar object which appears in a shared mental reality (imaginary); this is the beginning of representation. These

relationships must be understood in the context of neural networks. Similar properties activate the same sensors. Each object has a set of properties which stimulate a set of sensors which, in turn, activate a more complex neuron network. If two objects have common subsets of properties, there will be activation of a common sub-network of sensors and neurons integrating, in the immediate mental reality, the two similar objects. The icon, present in the immediate physical reality, activates its own specific neural network including the common subset which, in turn, excites the rest of the network specific to an object which, though absent from the immediate physical reality, takes place in the mental reality.

The symbol, which comes third, inherits from both the index and the icon. In itself, the symbol is a physical object like the object it represents. However, these two objects refer to mental realities which have nothing in common, no resemblance, except that, in physical reality, they are frequently associated like the name of an object and that object. The presence of one or the other in the immediate reality causes the appearance of the other in mental reality as in the icon case. The relationship may also be strictly mental if, for example, an icon recalls the face of a person, like in cartoons, then the name of the person may also appear in the mental reality and may, in turn, trigger other mental objects associated with that name.

From this we can conclude that:

1. the meaning of the sign is inseparable from the signal, but becomes significant only via the semantic interpretation of a conscious agent, and
2. the signal is sufficient for the propagation of a neuronal causal chain; its interpretation, its meaning, is not necessary for sensorimotor behaviors.

Finally, this conception of information is closer to de Saussure's (1913) than to Peirce's. The signal (the sign carrier) is the signifier while the meaning of the sign is the signified although signifier and signified have an intentional character involving a

conscious generator which we do not deem essential for potential information. It might also be worth noting that this dual aspect of information also corresponds to the fundamental principle on which Chalmers (1995) based his theory of consciousness... but our research is limited to the preverbal and preconscious sensorimotor brain.

1.2.3 - Modeling and simulation

The link that we have established in the previous subsection between information and causality brings us back to Turing who was certainly part of the mathematician elite of the first half of the twentieth century, but if we recognize his genius, is it not mainly for establishing a similar link between mathematical functions and physical causality (mechanics)? In purely mathematical terms, his machine was equivalent to Church's recursion, but the Turing machine was definitely causal whereas the Church's recursion was purely abstract. So, behind Turing's computational axiom, there must be an essentially physicalist hypothesis in the most causal sense of the term.

Figure I.2 shows that an algorithm according Turing establishes, by its action table, the causal links (mechanisms) between the outputs and the inputs. It is the input symbol (the signal) and not its meaning, which causes the internal change of state, but also, in a way, the change of state of the environment through the written symbol and the displacement of the tape to the right or left.

By reducing causality to its simplest form (any new state s^+, any written symbol and any resulting action are function of, and only of, the current state and the input symbol), the Turing machine promotes the atomization of causality into its simplest elements (mechanisms). The new state becomes the current state the next symbol to be read depends on the action taken (move to the right or to the left).

Any causal simulation of a physical phenomenon carries its load of potential information.

By physical phenomenon, we mean the observable behavior of a physical system. A physical system is a closed set of physical properties subject to disturbances by the environment and capable of disturbing the environment. To be observable, the system properties must be measurable by an observer. The observer is outside the system and can interact with (disturb) the system.

To simulate a phenomenon, the observer must be in the presence of two systems: one system (A), observed, and a second system (B), manipulated. The observer manipulates (disturbs) system B to reproduce the phenomena observed in system A. Both systems can be identical, in which case we will call them replicas of each other.

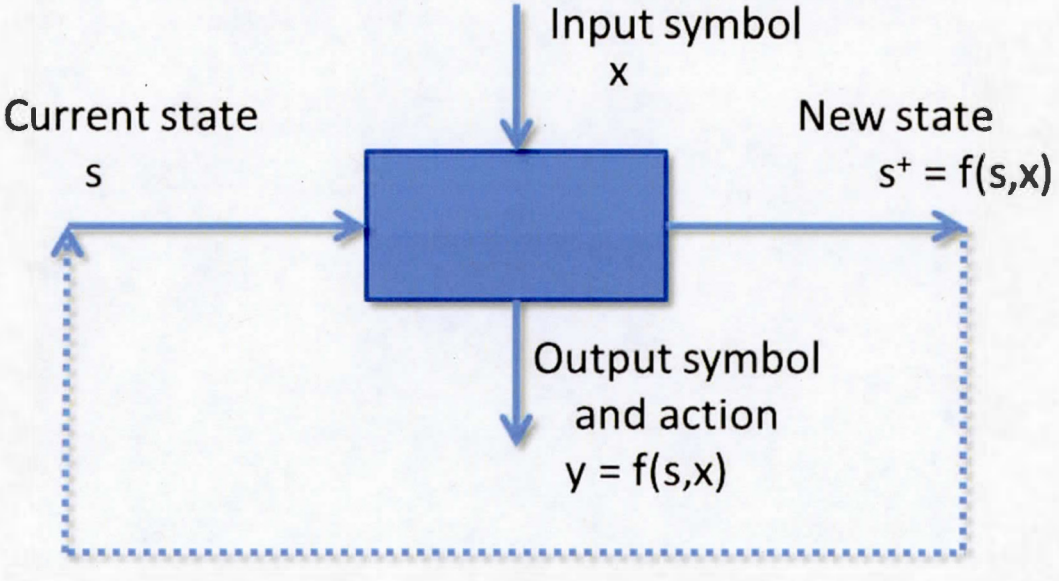


Figure I.2 - Physicalist perspective of the Turing machine

When replicas are placed in the same environment to eliminate (or at least minimize) the unintended disturbances, it is sufficient to identify the independent (or manipulated) variables to verify that a change in an independent variable in B causes, for all measured variables in this system, changes identical to the changes observed for the same variables in A for a similar change of the equivalent manipulated variables. In short, for two identical systems (replicas), identical changes of incoming signals result in identical state changes and identical changes of output signals since, the systems being identical, the causal relationships are identical.

Simulations and models usually reproduce only a subset of the original system's properties. For example, statues reproduce the shape of the original respecting its proportions with a scaling factor. Some scale models, like children's toys for example, simply reproduce the properties of shape and color, while others are used to study the relationship between certain physical properties of the environment and the shape of the object such as the effect of the air speed on a miniature airfoil in a wind tunnel. The model's physical system remains the same as the studied physical system except for scaling factors, but some interesting properties (lift and drag) cannot be observed with the naked eye; so, analog translations (transductions) are required for measurements. Stretching or compressing springs can be used to measure lift and drag. The results of these measurements are necessarily of numerical order: stretching or compression of the order of a few millimeters. This brings us back to the Dretske's principle we previously translated in causal terms: *lift (or drag) and the air speed are* ***causally*** *coupled since any change of speed will cause a proportional change in stretching or compression of the springs which are indicators of the lift (or drag).* These indicators, in centimeters, are numerical, like any measure, and this takes us from the physical world to the mathematical world where the Church-Turing thesis can simulate the relationship between physical properties using computable functions relating numerical measurements of these physical properties.

A measure is a dretskeian transduction (projection) of a physical world into an informational mathematical world.

H5 - Strong informational hypothesis

Any physical phenomenon can be simulated at the information level.

If two different physical systems have, for some properties, the same mathematical (informational) representation, they are, one for the other, an analog simulation. The Turing machine (nowadays we can say the computer) becomes, in a sense, the physical medium of numerical (digital) simulation.

1.2.4 - Argument summary

Table I.2 - But where does information come from?

P4.	Everything is matter/energy (Our physicalist ontological hypothesis)
P5.	Physicalism ⇒ the existence of material objects/properties (∃x, ∃y), of states for these objects/properties (∃s) and of causal relationships (∃f, ∃g).
C4.	These objects/properties and relationships form the potential information which can be actualized by whatever is capable to perceive it.
C5.	As soon as there is matter/energy, there is potential information (PI)
P6.	Potential information is transferred from one physical system to another physical system by causal coupling (inspired by Dretske)

P7. A causal coupling ⇒ that a change of state in a physical system produces (or results from) emission (or absorption) of matter/energy producing (or resulting from) a proportional change of state in another physical system (first law of thermodynamics)

P8. A mechanism $=_{def}$ a causal coupling between two physical systems.

P9. A semiotic system $=_{def}$ a system capable of sending and receiving potential information.

P10. This information transfer is, by definition, grounded in physical reality and independent of any semantic interpretation (escapes objections O1 and O2)

C6. Semiotic systems are sets of mechanisms grounded in physical reality and independent of any semantic interpretation. (from P7, P8 and P9)

P11. Any computable function can be represented by an algorithm of the form $y = f(s,x)$ and $s^+ = g(s,x)$ (Turing's thesis)

P12. Any causal coupling (thus any mechanism) may be represented by an algorithm of the form

$\Delta s = s^+ - s \doteq x - y$ which can be approximated by

$y = f(s,x)$ and $s^+ = g(s,x)$ (from P7 and P8)

C7. All, and only, mechanisms are algorithmizable (from P11 and P12)

P13. Any physical phenomenon can be decomposed into mechanisms.

C8. Any physical phenomenon is algorithmizable.

1.3 - Complementary hypotheses

However, a simulation is not a physical reproduction. If a simulation of a waterfall does not wet anybody, why would anyone think that a simulation of the brain would think?

1.3.1 - Emulation

It is obvious that a flight simulator will not move the plane, nor its content, nor its pilot to a remote destination. But in its most refined versions of virtual reality where the pilot sits in a full replica of a specific cockpit, he (the pilot) perceives and feels all reactions of the apparatus as if he was really in a plane. It is clear that much of the information processing required to guide the aircraft is in the brain of the pilot.

Consider now a similar experience where a drone flying over Afghanistan is controlled from a bunker somewhere in Colorado, USA. We are not talking simulation anymore; the drone is a real airplane actually flying above Afghanistan, potentially attacked by surface-to-air missiles and able to drop bombs on very real target. The pilot, seated in his bunker, immune to all danger, receives all kinds of information about the state of the drone, but does not physically feel anymore the drone's reactions. This is made possible due to Shannon's mathematical theory of communication (1948). So, the pilot, in Colorado, can remotely control the drone, in Kabul, because it receives and transmits information to the drone regardless of the means used for telecommunication.

As Shannon has shown, such telecommunications is subject to very specific time constraints. Let's go now to Houston, USA, where the Mars Rover engineers would love to, like the pilot in Colorado, remotely control their robot in its exploration of

the Red Planet. But communication is not possible because a radio signal takes from 3 to 21 minutes to travel the distance between Earth and Mars depending on their relative positions, not to mention the interference of the sun when it is in between. For these engineers, there is no escape, cognitive functions (at least some of them) must be implemented in the robot; no human brain can be close enough to remotely control the robot. Obviously, they will use all available digital and analogue simulations to study how a human would guide such a robot, analyzing the reactions of humans to different information provided by the simulated robot. These functional relationships between the information received by the human and the action taken can simulate human cognition at an informational level. As rough as this simulation can be, it allows us to extrapolate Shannon's theory on the transport of information (communication) to the level of information transformation (information processing or cognition), and to posit the following hypothesis:

> **H6 - Informational hypothesis of cognition**
> Cognitive phenomena can be emulated because they are strictly informational.

By emulated, we understand that the human intervention between the information received and the action taken can be replaced by its simulation (as rough as it might seem), provided that some causal chain at the signal level can translate the received information consistently into a message decodable by the robot's effectors while respecting the system's dynamics, or if you prefer the time constraints mentioned above.

1.3.2 - Granularity

However, do not be mistaken, the causal chains do not exist at the phenomena or behavior level. At best, one could find mere correlations. To refine the simulation, it

is necessary to detail the description of the human perceptions and actions, but those cannot be described beyond the language concepts associated with the categories of identifiable objects while perception and information differentiation begin at a significantly subconceptual and subobjectal[12] level as Smolensky said (1987) in his hypothesis 8:

> **The subsymbolic hypothesis**: The intuitive processor is a subconceptual connectionist dynamical system that does not admit a precise formal conceptual-level description.

In fact, any information used by engineers to control Mars Rover, like any information transmitted by the drone's remote pilot is already semantically interpreted. Take, for example, three simple instructions to the drone pilot: speed, altitude and orientation of the aircraft. These properties are measured, respectively, by a Pitot tube, a barometer and a compass. Each measure, each indication is transmitted to the driver with a label; so, clearly pre-interpreted.

On the other hand, the drone may also be provided with a camera. In this case, it is quite different for the interpretation of the captured images. In agreement with the Shannon's theory, the drone can telecommunicate all collected information to the pilot with sufficient precision for him to interpret the content of the image and fly visually by identifying significant objects in the landscape. It must be understood that the image, transmitted after digitization of punctual signals, can be approximately (but accurately) reconstructed on a receiving screen which, in turn, emits physical (light) signals capable of causally interacting with the pilot's brain in the same

---

[12] Objectal = related to the object. In French, objectal is a neologism introduced by Lacan because objective had taken a different meaning. With the prefix sub-, we use it to refer to properties of an object which, although inexistent independently of the object, can be perceived before or without perception of the object itself.

manner that the signals picked up by the camera would have interacted with his brain had the pilot been on board the drone. As the signal carrier has been translated several times between the emitting objects[13] in Afghanistan and the pilot's eyes in Colorado, one can conclude that most of the essential information was saved at the punctual level in each pixel point by reproducing the intensity and geometric structure.

In fact, dretskeian physical transduction always occurs at the punctual physical level, at the pixel level for visual signals, which we could generalize as the level of "sensels" for any sensitive element, taking sensitive as active for the receiver capable of sensation and perception and passive (or perceptible) for the signal emitted or modified by an emitter or reflector. These sensels are, somehow, the atoms of information and it is from these sensels that must be built all semantics.

Going back to Houston, we understand that the Mars Rover engineers can (and they do it very well indeed!) establish between semantically interpreted signals, such as speed, direction and power, equally preconceived links bringing back to mind Braitenberg's vehicles (1984), Brooks' subsumption architecture (1986) and Brooks' (1989) and Arkin's (1998) behavioral robotics. They can even program some shape and color recognition software to interpret the pixels transmitted by high definition cameras, but not without instilling in the robot a minimum of preconceived semantics. They certainly produce subsymbolic systems which might approach symbol grounding, but which certainly do not meet Floridi's Z condition. Although the signals are natural and real, the links between these signals are still artificial, externally designed and programmed. The algorithms are implemented by an external observer capable of semantic interpretation. To achieve the Z condition, it is not sufficient to automate the robot's response, it is necessary to automate the external

---

[13] Technically we should say reflectors since it is by selectively reflecting sunlight that different objects differently affect the camera sensors.

observer's work, his algorithm generation. We must therefore automate automation, algorithmize algorithmization; do what von Foerster (1974) called cybernetics of cybernetics or second-order cybernetics

In the Smolensky's context, one must understand that the subsymbolic systems are from a subconceptual level, but not yat the neuronal level; that is to say not yat the brain's strictly physically causal level, and therefore not totally free of semantic interpretation. Harnad (1992) pinpoints essentially the same thing by distinguishing between the Turing Test (TT or T2), the Total Turing Test (TTT or T3) and the Total Total Turing Test (TTTT or T4) which could be identified with Smolensky's conceptual, subconceptual and neural levels. Symbolic systems, working at the conceptual level, can pass the T2 if and only if they are supported by sub- symbolic systems working at the subconceptual level and able to ground symbols in a physical reality. For Harnad, robotics is a necessary complement to symbolic systems for TOTAL performance evaluation... and symbol grounding is no more than an appreciable « bonus ». This overall performance criterion (intellectual performance plus sensorimotor performance) is not part of Turing's performance criterion; it was added by Harnad to account for all human capacities not limited to pen pals' activities. Yet, he notes further that « [It] may be that even successful TT capacity has to draw upon robotic capacity. »

Admitting that some sensorimotor elements are necessary for symbols grounding does not imply that all the sensorimotor capabilities must be indistinguishable. A severely physically handicapped human could well aspire to full pen pal recognition without any doubt about his intellectual capacity; think Stephen Hawking. On the other hand, these sensorimotor elements are not sufficient to explain how relations develop between well-grounded symbols. Without claiming indistinguishability, we believe that some elements of T4 are essential for T2. No robot can aspire to the title of pen pal if its symbols **and the relations between these symbols** have not been

established and grounded in reality by the autonomous development of a sensorimotor brain. For an independent grounding of any preconceived semantics, we must emulate the brain at the neuronal level. This artificial brain is not a neuro-molecular reproduction, but a computational emulation of a biological brain and a rough approximation of the latter from a sensorimotor perspective (T3) while being indistinguishable from a symbolic point of view (T2). This leads us to the following hypothesis:

> **H7 - Epistemic hypothesis**
> Phenomena emerge from underlying mechanisms which must be explainable by other underlying mechanisms as long as such mechanisms have a significant effect on the phenomena to be explained.
>
> (Only mechanisms can be algorithmized; phenomena emerge from underlying mechanisms.)

In other words, intelligence (thought, mind) is not a mechanism nor a machine (set of mechanisms) in itself, but a property of a complex system, the brain. The concepts are possible only by composition of a multitude of sensels of different types and intensity because information is only available in this form. Recall Floridi's (2011a section 1.3) *diaphoric definition of data* as discussed on page 19. The perception of the object necessarily goes through the capture of its properties in a punctual space-time. The composition of these sensels in representations and concepts is the first step of cognition and corresponds to an organism's sensorimotor development. Note that these concepts do not wear labels, no symbolic referents, and can therefore only be activated by the presence of the object in the immediate sensory environment. Symbolic referents will emerge with the advent of language.

The epistemic assumption thus favors the biological hypothesis based on neural mechanisms rather than the psychological hypothesis based on conceptual phenomena. However, this biological option is not without constraints as we will see in the next section.

1.3.3 - Argument summary

Table I.3 - Is emulation possible?

P14. A simulation $=_{def}$ a composition of algorithms representing a physical phenomenon at the informational level.
P15. An information system $=_{def}$ a semiotic system where the physical carrier is of secondary importance in relation to the information provided.
P16. An emulation $=_{def}$ a simulation of an information system
P17. Cognitive systems $=_{def}$ informational systems
C9. The brain is a cognitive system, thus informational.
C10. Cognitive systems can be emulated, i.e. replaced by equivalent information systems instantiated by different physical carriers. (multiple realizations)

1.4 - Biological constraints

It is hardly surprising that we have to consider the biological constraints since all known cognitive systems are biological systems. While these constraints seriously

complicate simulation and emulation efforts, they provide an opportunity to meet the requirement of Floridi's Z condition since they involve autonomous (without external control) and evolutionary (random and selective) development.

1.4.1 - Autopoiesis

Maturana and Varela (1980) introduced the concept of autopoiesis according to which, in biology, the product of the process is the process itself. Their original definition (Maturana and Varela, 1980 pp 78-79) is as follows:

> An autopoietic machine is a machine organized (defined as a unity) as a network of processes of production (transformation and destruction) of components that produces the components which:
> (i) through their interactions and transformations continuously regenerate and realize the network of processes (relations) that produced them; and
> (ii) constitute it (the machine) as a concrete unity in the space in which they (the components) exist by specifying the topological domain of its realization as a network.

Thompson (2007 p100) paraphrases this definition at the cellular level in the following terms:

> [...] a molecular autopoietic system is one in which chemical reactions produce molecules that (i) both participate in and catalyze those reactions and (ii) spatially individuate the system by producing a membrane that houses those reactions.

Figure I.3, slightly modified from Thompson, indicates that a cell delimited by a membrane (bounded system) generates a network of metabolic reactions which, through DNA, RNAs and proteins, produce the components determining the molecular membrane and the cell contents. The membrane is necessarily semi-permeable to let the required elements enter and the unnecessary waste leaves the

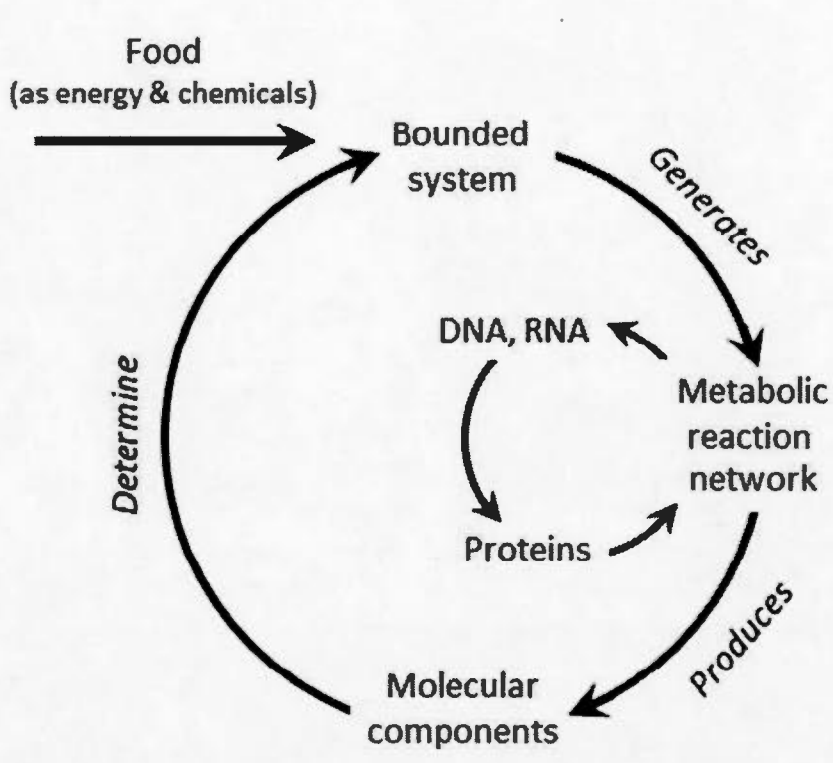


Figure I.3 - Cellular autopoiesis
(adapted from Thompson 2007 Figure 5.1 page 102)

cell. Its primary role is to "selfishly" keep intermediary products useful for its maintenance. The Figure I.3 presents a "modern" cell with a complex network of reactions requiring sophisticated elements such as DNA, RNAs and proteins, but one must understand that this metabolism began with the production of semi-permeable membranes closing on themselves and promoting reactions involved in the production of elements of semi-permeable membranes. This was only the beginning of a long history of Darwinian evolution, hence of natural selection, even before the appearance of genes, but clearly generated by chance and utility.

Autopoiesis sheds some light on the circular causality proposed in the thesis project a few years ago and described by paraphrasing Descartes: "I eat, therefore I live, therefore I eat, therefore..." Indeed, the semi-permeable membrane lets in some

elements (eating) which favor certain metabolic reactions (living) which transform these elements and reduce their concentration, causing the membrane to let in more of these elements (eat).

For Maturana and Varela, autopoiesis is necessary and sufficient to define any biological (living) system as indicated by the following quotes: « the notion of autopoiesis is Necessary and Sufficient to Characterize the organization of living systems » (Maturana Varela 1980 p82) and « Autopoiesis in the physical space is necessary and sufficient to characterize a system as a living system » (Maturana and Varela, 1980 p112). Thompson (2007 p124) would probably have preferred to stick to this position, but he still accepted arguments from Bitbol and Luisi (2005) and Bourgine and Stewart (2004) saying that « all living systems are both autopoietic and cognitive systems, but an autopoietic system is not necessarily a cognitive system ». Thompson had no objection to accept as non-living, because non-cognitive, cases where « the system autocatalytically produces its own boundary but does not actively relate to its environment » (Thompson, 2007 p125) as long as it preserved the identification of biological systems to cognitive systems made by Maturana in his first article on autopoiesis: « Living systems are cognitive systems, and living as a process is a process of cognition. This statement is valid for all organisms, with and without a nervous system. » (Maturana, 1970 p13).

This identification of the cognitive with the living presupposes extended definition of cognition including not only animals' sensorimotor activities, but also microorganisms' taxes and plants' tropisms. While accepting that the taxes and tropisms could be signs of cognition, we can still be reluctant to recognize them as signs of intelligence. Like the first autopoietic systems were not quite biological, we could say that the first biological systems, although cognitive, were not quite intelligent. We can distinguish three levels of cognitive systems: the strictly causal

cognitive systems (SCCS), the sensorimotor cognitive systems (SMCS) and the verboconceptual cognitive systems (VCCS).

The SCCS include microorganisms' taxes and plants' tropisms. We say they are strictly causal because they are nothing more than a causal coupling *à la* Dretske or a *de re* information exchange (which we previously called "potential information" and which is not considered as real information by Floridi but simply as data or differences). These are simple unidirectional causal chains without inhibitory interactions (e.g. bacteria, sunflowers, thermostats). There is no abstraction; this is pure and simple causality (abstraction level 0).

The SMCS correspond to animals' movements. The information is *de signo*. Indexical links are created by composing information. While SCCS have exclusively local sensors (taste, touch), the SMCS have remote sensors (hearing, smell and vision). In a way, we can speak of smell as remote tasting and of vision as remote touching (although vision adds color to the shape and does not include thermal components). This multiplication of modes and the creation of intermodal links produce unidirectional, but interconnected and possibly inhibitory, causal chains. Rats can smell the cheese and taste the cheese; the smell of cheese becomes a hint (an index) of an interesting meal. The same rats can smell the cat and, if the smell has previously been associated to an unfortunate encounter with a cat or simply with a frantic flight with its conspecifics in similar circumstances, they will certainly inhibit any curiosity and take action to escape. There is a first degree of abstraction (level of abstraction 1) where simple, but multiple, causal chains interact for stimulation or inhibition. This level of abstraction is similar to the hidden layer of a multilayer perceptron.

Finally, the VCCS correspond to human conscious movements using *de dicto* information. Links are created between words and sensory perceptions or motor

actions forming a parallel plane of interactions between increasingly complex causal chains. This is the second level of abstraction (level of abstraction 2) which we will not discuss in this thesis as we have already mentioned a few times, but which allows us to better understand the SMCS's upper limit. The word level allows an imaginary reproduction of a situation (independent of the immediate spatiotemporal reality) for the approximate evaluation of possible outcomes and the adjustment (including inhibition) of some less beneficial causal chains. The signals from the remote sensors are not only composed, but they are interpreted to activate a causal semantic level, a level allowing the creation and validation of hypotheses.

AI generally argues that intelligent systems are not necessarily biological and, more specifically, the artificial neural networks suggest that an (intelligent) nervous system is not necessarily biological. This thesis focuses on the question: «Is it possible to develop a non-biological, but still autopoietic, nervous system? »

So, if we focus on autopoiesis, it is not so much at the biological level, although it is, in nature, essential to any other level, but rather because it allows the conception of another level for the autonomous development of the brain. Figure I.4 shows a superposition of two autopoietic loops. At the bottom, we see the biological loop producing and maintaining cells, the neurons, which become the cellular components of a neural network fed by information which stimulates reaction networks in several neurons in order to produce neurotransmitters, ion channels and pumps, which modify these neurons and consequently the network's response to future stimulations.

In the lower loop, the food, in the form of matter/energy, is directly involved in the physical structure of cells, while in the upper loop, information, a different form of matter/energy, activates the cells' metabolism only to change the organization of the neural network. One must see the brain, or more precisely the (single) central and peripheral nervous system, as a bounded system with an interface (not really a

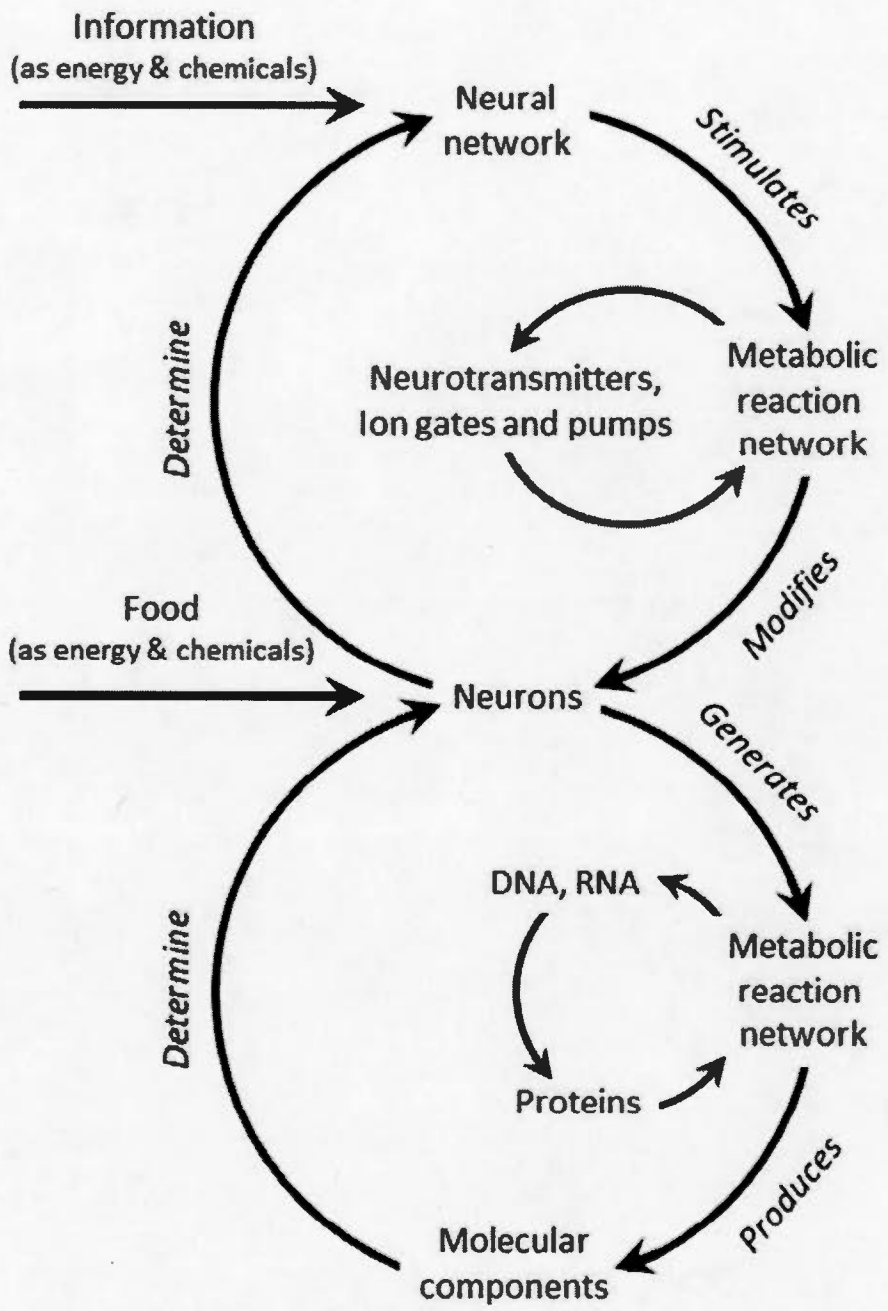


Figure I.4 - Double autopoiesis

membrane) consisting of frontier cells reacting to external physical objects/properties (matter/energy) and transform information into neurological signals (neurotransmitters) and vice versa; therefore, this neural network has its own semipermeable barrier, more specifically semipermeable to information. To be more precise, perhaps we should say transfer (translate, "transduce") information on neurological signals.

Clearly the adaptation of each neuron depends on its genetic content, but the network develops (evolves) on random stimulations (received information) according to the utility of the behavioral reactions.

H8 - Autopoietic hypothesis

Biological mechanisms develop autonomously in an evolutionary manner.

In this double autopoiesis, *information is to cognition what food is to digestion and what oxygen is to breathing.* To better understand the relationship between information, food and matter/energy, let's take the example of sharks which, according to experts, are able to detect blood even in concentrations smaller than one part per million. When a shark detects a drop of blood, it is far from a meal, but if it uses the information appropriately (if you allow this anthropomorphism), it will be guided by the concentration gradient, or countercurrent, to enough food to satisfy its appetite. This example also helps us understand the difference between a physical system and an information system. When the shark smells the blood drop, it receives very little material content, but enough information to find the blood source. When it eats this source, it gets a lot of material content and very little additional information. The first (top loop) is a cognitive information system while the second (bottom loop) is a digestive physical system (hardware). The former can be emulated while the latter can, at best, only be simulated.

1.4.2 - Evolution and development

Like DNA did not appear as soon as the first vesicle was formed, neurons are the result of a long evolution by natural selection. Unicellular organisms, such as bacteria, obviously do not have neural network since everything is within a single cell. Yet, food (matter/energy) entering the cell carry its own information. This information affects some reactions of the metabolic network and, thereby, participates in the management of the organism's behavior. This behavior can be vital or fatal for the organism and at the same time for reaction types. Only useful reactions will "survive" and will become part of the DNA.

H9 - Evolutionary hypothesis

Only the mechanisms, which are nondetrimental to survival, survive (Darwin).

We are interested in the survival mechanism, not the individual's survival or the species' survival, although the three are strongly intertwined. Indeed, *like the reproduction of the individual favors the survival of the species, the reproduction of the mechanism promotes the survival of the individual and the survival of the species favors the survival and the reproduction of the mechanism.*

As organisms become more complex and multicellular, cells specialize and organize themselves (to be taken as much in the sense of organization as in the sense of agglomeration into organs). In this evolution, neurons specialize in information processing and generate neurotransmitters and ion channels sensitive to neurotransmitters. As these components facilitate behaviors which are particularly useful for survival, they proliferate rapidly. Obviously, when we talk about utility, we are talking about nutility taking into account all the negative and positive effects. Neurons are energy intensive cells; the brain uses about 25% of the total energy of the body. To compensate for this exorbitant cost, neurons must produce more than

equivalent beneficial results. It is not as individual elements that they can produce such results; it is not even as an independent organ (the brain or neural network) that they can justify this energy consumption. The usefulness of neurons can only be assessed at the level of the whole organism; it is the effect of the neural network on the organism's behavior that will be subject to the evaluation of natural selection. As shown in Figure I.3, cellular autopoiesis is based on the reproduction of a DNA coded reaction network. Any new cell is similar to the neighboring cells because it contains a complete copy of the "recipe". As we have already mentioned, this "recipe" is the result of a long evolution; this is genetic inheritance.

On the other hand, Figure I.4 shows that the neural network in the top loop is only indirectly affected by DNA. When stimulated by incoming information, the upper loop uses the reaction network in the lower loop to produce neurotransmitters or ion channels and pumps to modify and adjust the neural network. DNA specifies how to make these chemical components, but contains no information about the structure of the neural network. This structure is the result of the ontogenetic development of the brain and not its phylogenetic evolution. This is the cultural and epigenetic inheritance.

Referring to Maturana and Varela's original definition, one could say that the brain is:

An autopoietic machine is a machine organized (defined as a unity) as a network of **(interconnected neurons)** processes of production (transformation and destruction) of components **(connections)** that produces the components **(connections)** which:

(i) through their interactions and transformations continuously regenerate and realize the network of processes **(interconnected neurons)** that produced them; [...] and

(ii) constitute it (the machine) as a concrete unity in the space in which they **(the connections)** exist by specifying the topological domain of its realization as a network.

Having defined a "membrane" (interface) semi-permeable to information and a self-adaptation process, we use this type of autopoiesis to define an organized level of cognition which we call intelligent systems (including SMCS and VCCS).

For Bourgine and Stewart, like for Bitbol and Luisi, minimal autopoiesis cannot be described as living (biological) because it is not cognitive, because it is not actively involved in its interaction with the environment. The stone heated by the sun passively receives its energy. The vesicular membrane, which automatically mends itself in presence of the necessary reagents, is passively reacting to some homeostatic equilibrium that ensures constant supply of these reagents; it will never move to find these reagents if the immediate environment is impoverished. The bacterium, activating its flagellum when the internal glucose concentration decreases, no longer passively suffers environmental changes, but actively responds. The causal coupling is just as determined as for the stone heated by the sun, but the reaction causes an environmental change leading to an adjustment of the homeostatic equilibrium in play. The homeostatic mechanism has not changed, but the behavior modified the interacting forces. In this simple case, the environment change is caused by movement, but it is also possible to change the environment in many ways when the means of action (actuators) and the assessment of conditions (sensors) are multiplied and become more complex.

1.4.3 - Argument summary

Table I.4 - So, (Strong) artificial intelligence is possible, but how is it built?

P18. All known cognitive systems (including the brain) are biological systems
P19. Biological systems are autopoietic systems capable of autonomous and evolutionary development (Maturana and Varela)
C11. All known cognitive systems (including the brain) are autopoietic systems capable of autonomous and evolutionary development (meeting Florida's Z condition)
P20. The cell is the elementary component of biological systems
C12. The cell (the neuron) is the elementary component (the mechanism) of any (known) cognitive systems
P21. Biological systems evolve with natural selection
C13. Only the mechanisms, which are nondetrimental to survival, survive (Darwin)
P22. The advanced cognitive systems are capable of practical (sensorimotor) intelligence and even of general intelligence
QED. Autopoietic semiotic systems are capable of practical (sensorimotor) preverbal and preconscious intelligence (from P9, P15, P17, C9 and P22)

1.5 - Conclusions

1.5.1 - Identified hypotheses

In this chapter, we identified nine hypotheses that we will use as postulates for the development of our model. We rejected the psychological hypothesis (H2a) and have preferred the biological hypothesis (H2b).

H1 - Computational Hypothesis (or computational axiom)
Any algorithm may be performed by a Universal Turing Machine.

H2a - Psychological hypothesis (which we will not accept)
Intelligence is directly algorithmizable.

H2b - Biological Hypothesis
The brain is algorithmizable.

H3 - Ontological hypothesis
Everything is matter/energy.

H4 - Semiotic hypothesis
As soon as there is matter/energy, there is potential information, therefore signs.

H5 - Strong informational hypothesis
All physical systems can be simulated at the information level.

H6 - Informational hypothesis of cognition
Cognitive phenomena can be emulated because they are strictly informational.

H7 - Epistemic hypothesis

Phenomena emerge from underlying mechanisms which must be explainable by other underlying mechanisms as long as such mechanisms have a significant effect on the phenomena to be explained.
(Only mechanisms can be algorithmized; phenomena emerge from underlying mechanisms.)

H8 - Autopoietic hypothesis

Biological mechanisms develop autonomously in an evolutionary manner. (Maturana Varela).

H9 - Evolutionary hypothesis

Only the mechanisms, which are nondetrimental to survival, survive (Darwin).

The analysis of these hypotheses allowed us to emphasize the two main objections seemingly blocking all attempts in the development of Artificial Intelligence. The first, described by Searle (1980) in the Chinese room thought experiment, was identified by Harnad (1990) as the symbol grounding problem. The second, more recently identified by Floridi (2011) as the *Zero semantic commitment condition* tells us that the symbol grounding cannot be solved externally neither by inneism nor by programming, and must result from evolution and/or autonomous development.

1.5.2 - Cognitive thesis

We present a cognitive thesis: « only autopoietic semiotic systems have the necessary and sufficient means for general intelligent action » which is a precision of Newell and Simon's thesis replacing physical symbols systems by autopoietic semiotic

systems. The evolution of semiotic systems eliminates the symbol grounding problem[14]. The autopoiesis involves self-generation of algorithms (bootstrapping) suggesting algorithmization of algorithmization[15]. Some would rather call it automation of automation, cybernetics of cybernetics, or second-order cybernetics (von Foerster 1979). There might be other ways to realize symbol grounding and zero semantic commitment, but until they are discovered, we must consider semiotics and autopoiesis necessary to produce cognitive systems.

1.5.3 - Neuro-computational thesis

Our neuro- computational theory, to be described in the next chapter, claims that it is possible to produce such autopoietic semiotic systems if the neurons are dynamic (to perceive temporal changes of information), analog (to proportionally represent the basic signal) and asynchronous (to allow digital communication of the state without external intervention).

---

[14] According to Searle and Harnad, no symbol grounding implies no cognition. The contrapositive of this proposition tells us that cognition implies symbol grounding. If we add to this that symbol grounding implies causal coupling which requires causality level semiotics, we can conclude that cognition implies semiotics which is therefore necessary for cognition.

[15] Floridi's Zero Semantic Commitment condition tells us that cognition implies no innateness and no programming which implies some form of autogeneration which we interpret as autopoiesis bringing us to conclude that cognition implies autopoiesis which is therefore necessary for cognition.

CHAPTER II

Modelling the brain

In the preceding chapter, our hypotheses led us to the conclusion that only autopoietic semiotic systems have the necessary and sufficient means for practical (sensorimotor, preverbal and preconscious) intelligence which is a necessary developmental step (Piaget 1936) towards general intelligent action. We will now review the evolution of neuron models to identify which mechanisms, if any, make the brain semiotic and autopoietic.

2.1 - The neuron doctrine

The scientific study of the brain started in the late 1700 when Luigi Galvani (1791) discovered that muscles and nerve cells produced electricity. In the late 1800, Camillo Golgi developed a method of staining neurons with silver salts that revealed their entire structure under the microscope which was used by Santiago Ramon y Cajal to elaborate his neuron doctrine. Bullock al. (2005) wrote:

> [...] it was Cajal who envisioned the neuron as an individual functional unit, polarized such that signals are received through its rootlike dendrites and transmitted through its long axonal process [(generally referred to as dynamic polarization)]. He posited that although an axon terminates adjacent to a dendrite of the next neuron [...], the cleft between them would act as a synaptic switch regulating information flow through neural circuits. The synaptic cleft went unseen until a half-century later, when in 1954 the electron microscope provided convincing evidence that essentially refuted the earlier "reticular" view of a nerve fiber web.

Sherrington (1906) was also an ardent champion of cellular connectionism.

With the advent of electron microscopy in the 1950s, Palade and Palay (1954), Palay and Palade (1955) and De Robertis and Bennett (1955) demonstrated the existence of synapses specialized in the chemical and electrical signaling between neurons.

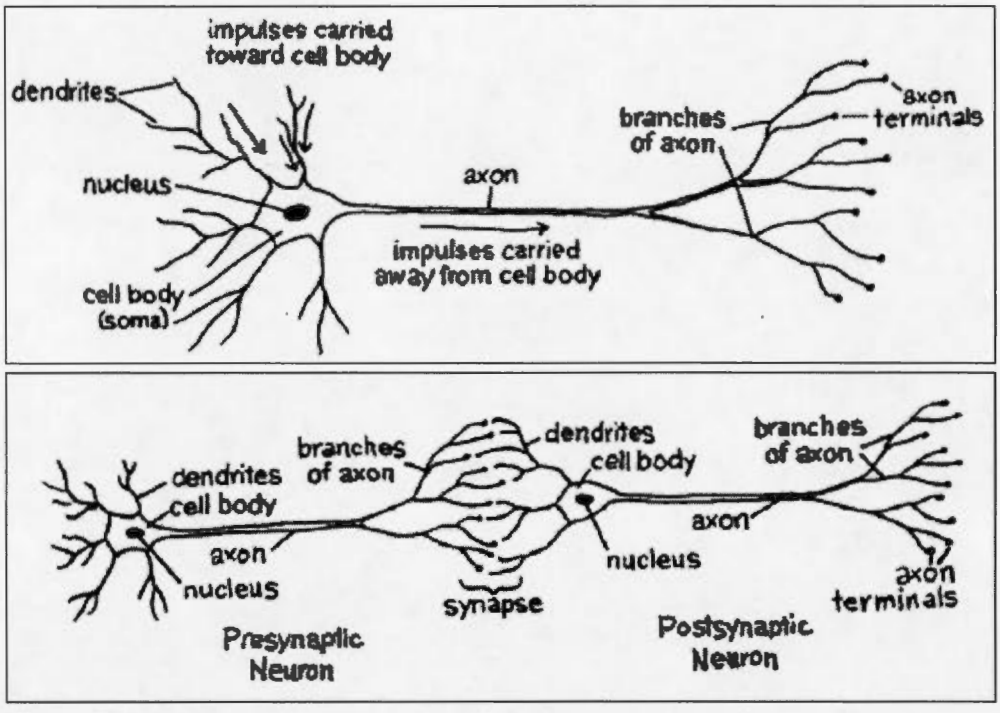


Figure II.1 - Nerve cells (neurons) and main components (from https://science.education.nih.gov/supplements/nih2/addiction/guide/lesson2-1.htm)

Kandel al. (5th ed. 2013 ch. 2) supports Ramon y Cajal's neuronal doctrine specifying that « nerves cells [neurons] are the signaling units of the nervous system » leaving a support role to glial cells. They also add that « signaling is organized in the same way in all nerve cells » (ibid. p29) which generate « four different signals in sequence, each at different sites within the cell: an input signal, a trigger signal, a conducting signal and an output signal » (ibid. p29). This organization holds for all types of neurons including unipolar, bipolar or multipolar cells (classification already

recognized by Ramon y Cajal). It also holds independently of the role of the neuron: sensory neuron, motor neuron or interneuron (ibid. p30, figure 2-9, reproduced below as Figure II.2). This does not deny the recent developments in « single-channel recording, live cell imaging, and molecular biology » reported by Bullock al. (2005), but it clearly differentiates the signaling role of the neurons from the biological support role of the glial cells; a differentiation we already alluded to when discussing double autopoiesis in section 1.4.1.

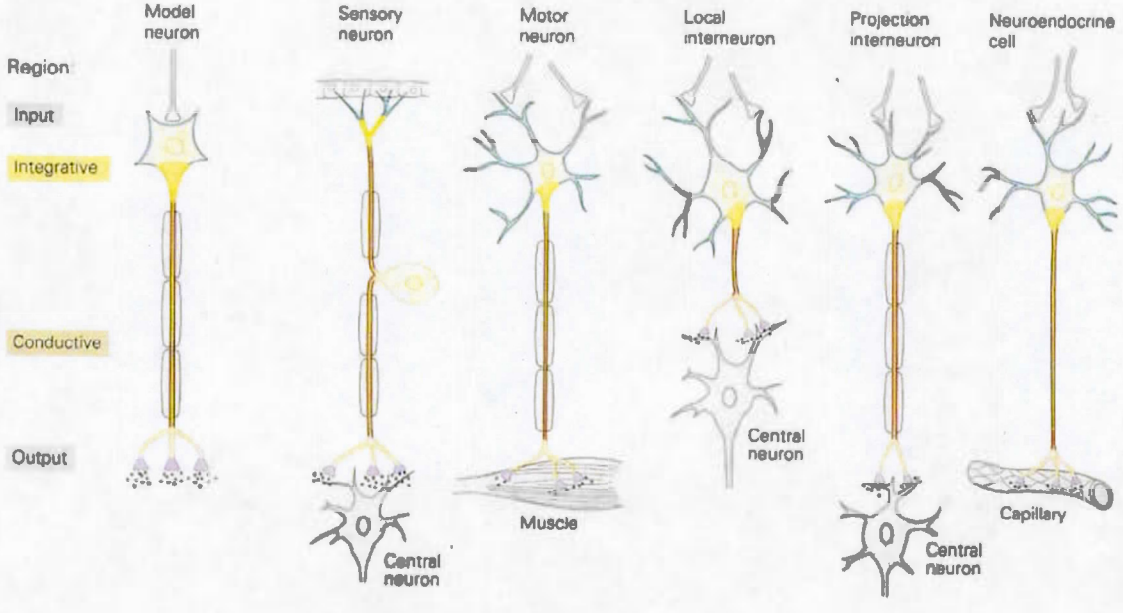


Figure II.2 - Different types of neurons, one signaling system.
(From Kandel al., 5th ed. 2013, p30, figure 2-9)

The presence of a (potentially information carrying) signal is sufficient to declare the neuron semiotic. The "meaning" is not yet obvious, but we will come back to semantics later on.

2.2 - The spiking neuron

Lapicque (1907)[16], inspired by Nernst (1899) and Nernst and Barratt (1904), was the first to use an equivalent electrical circuit to analyze the threshold behavior of semi-permeable membranes like organic tissues. This equivalent circuit is the first element of a model of a biological neuron. It is not a graphical representation of the form of a neuron like those produced by Ramon y Cajal using Golgi's staining technique. Forms and colors are not part of this representation like we have been accustomed to by arts such as sculpture, painting and photography. However, the dynamics of the threshold behavior is fully represented like the dynamics of movements can be represented in animated movies and, to some extent, in music. When we talk about modelling neurons and, later on, the brain, we are referring to this kind of representation of dynamics which implies solution (computation) of differential equations or numerical approximation (also computation) thereof.

Hodgkin and Huxley (1952) provided « a quantitative description of membrane current and its application to conduction and excitation in nerves » focusing on « the flow of electric current through the surface membrane of a giant nerve fibre ». By « *giant nerve fibre* », we should understand the giant axon of the squid which does not include the entire neuron. Hodgkin and Huxley, like Lapique, were interested in modelling the dynamic behavior of the neuron, more specifically the *action potential* running down the axon. They were, in fact, modelling a mechanism which will be later identified as voltage-gated ion channels. Lapique had shown that the membrane became permeable when the voltage exceeded a given threshold; a first hint at the triggering portion of Kandel's canonical model. Hodgkin and Huxley extended the observation to sodium, potassium and chloride ions showing that the conductance of

---

[16] For an english version see Brunel and van Rossum (2007).

sodium and potassium were functions of time and membrane potential. Their model could simulate the triggering of an action potential and its propagation along the axon up to the terminal.

Meanwhile, McCulloch and Pitts (1943), interested by the computational possibilities of the neurons, had introduced a rudimentary model neglecting the dynamic behavior, but combining a linear weighted summation model of the dendritic tree with a threshold triggering of the action potential. Although they were addressing most of the behavior of the neuron (dendritic summation, threshold triggering, weighted communication), their interest for binary logic (a mental function) brought them to oversimplify some known properties (dynamics) of the neuron. The boundary between computationalism and connectionism was already blurred: the model was clearly neurally inspired, while the objective was mentally directed.

McCulloch and Pitts' simplification was not limited to neglecting the transient dynamics of action potentials to carry the signal along the axon (which is probably acceptable in most cases except for very long nerves where there is a significant delay between the triggering of the action potential and the emission of neurotransmitters at the axon terminal), but also neglected the phenomenon originally reported by Adrian (1926ab) and Adrian and Zotterman (1926ab) about rate (or frequency) coding stating that the rate (or frequency) of action potentials increases when the intensity of the stimulus increases (as shown on Figure I.1 page 13).

Stein (1965) proposed an algorithm describing the operation of a leaky-integrate-and-fire neuron (LIF) with linear accumulation of input impulses until a threshold is reached which triggers firing and resets accumulation (or depolarization) to zero, while for subthreshold levels the accumulation decays exponentially between impulses. Clearly in this case the intent is to represent the entire neuron from the dendritic summation to the emission of action potentials (AP). Like McCulloch and

Pitts, Stein reduces Hodgkin and Huxley's detailed calculations of membrane currents to a simple impulse to represent APs; we have already accepted this simplification. On the other hand, the reduction of the dendritic tree to a single point where pulsed (current) inputs are linearly summed up until the threshold is reached, at which moment the accumulation vanishes to restart from zero, is probably an oversimplification. An oversimplification because 1) the entire dendritic tree cannot be instantly repolarized nor 2) can the opening of a sodium channel have the same effect when the neuron is highly depolarized as when it is fully polarized.

Biological neurons are essentially dynamic. At the same time, they must also implement some kind of logic. Compromising one characteristic to accommodate the other leads to incompleteness. The dynamics have to encompass the entire dendritic tree with all its synapses as well as the axon with its communication power. Clearly, the logic transforming synaptic inputs into axonal action potentials is not classical, not binary; values have to be continuous, analog, graded and the logic becomes fuzzy with decisions which are neither conjunctions nor disjunctions, where no input can be sufficient nor necessary, except in very specific cases (e.g. for sensors, a specific input is both sufficient and necessary).

2.2.1 - Synaptic plasticity

Hebb (1949) introduced the hypothesis that would define neural networks and explain the adaptation of neurons in the brain during the learning process.

> When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased.

The mathematical formulation of Hebb's rule took many forms and the most popular and successful was proposed by Rosenblatt (1957, 1962). Applying Hebb's rule to McCulloch and Pitts neurons, he developed the perceptron, a linear classifier, which seemed promising until Minsky and Pappert (1969) showed that it was impossible for these networks to learn an XOR function.

Rumelhart, McClelland and the PDP[17] Research Group (1986) brought the perceptron back to life by adding hidden "units" between the input "units" and the output "units". This multilayer perceptron, as they called it, could handle the XOR function and other non-linear functions thanks to the additional degrees of freedom in the hidden layers. A clever mathematical procedure, called back-propagation (Rumelhart, Hinton and Williams 1986), made it possible to adapt supervised learning (Hebb's rule) to the multilayer network. Like McCulloch and Pitts and Rosenblatt before them, the PDP group put the emphasis on the function and neglected the biological dynamics of the neurons. According to our computationalist/connectionist cleavage, their methodology was decidedly connectionist, but their objective was clearly computationalist; even the vocabulary shifted from "neurons" to "units" to avoid having to justify any biological plausibility.

The discovery of long-term potentiation (LTP) by Bliss and Lømo (1973) provided the first experimental evidence for synaptic plasticity. High-frequency tetanic stimulation, driving postsynaptic neurons to fire, leads to LTP, an increase in the synaptic response to single stimulus. Subsequently Hebb's postulate was extended to encompass long-term depression (LTD) as a necessary converse of LTP (Stent, 1973; Sejnowski, 1977). Low-frequency stimulation, not driving postsynaptic neurons to fire, leads to LTD, a decrease in the synaptic response to single stimulus. LTP and

---

[17] Parallel Distributed Processing

LTD of inhibitory synapses (referred to as LTPi and LTDi) have also been observed as reported in Maffei (2011).

With new technologies and new procedures (such as dual whole-cell voltage recordings), it became possible to perform new experimental studies on the precise (sub millisecond) relative timing of spikes emitted on both sides of a monosynaptic connection between two paired neurons (Markram al., 1997ab). According to Sjöström and Gerstner (2010),

> Spike Timing Dependent Plasticity (STDP) is a temporally asymmetric form of Hebbian learning induced by tight temporal correlations between the spikes of pre- and postsynaptic neurons. As with other forms of synaptic plasticity, it is widely believed that it underlies learning and information storage in the brain, as well as the development and refinement of neuronal circuits during brain development (e.g. Bi and Poo, 2001; Sjöström al., 2008).

This procedure is difficult to generalize to multiple pre- and postsynaptic spikes as indicated by the multiple attempts including: Kempter al. (1999), Song al. (2000), Izhikevich and Desai (2003), Abbott and Nelson (2000) and Wittenberg and Wang (2006). Clopath (2010) proposes an interesting solution based on virtual traces of the cellular potential at the synapse which are suitable for a phenomenological model, but do not provide any explanation or understanding of the underlying causes. We will come back with more details on this subject when time comes to define how our model deals with synaptic plasticity.

STDP emphasizes the importance of the temporal correlation, temporal coincidence, of pre- and postsynaptic activity for learning to take place. It is like something happens (locally) at the synapse while the channels are opened and that something depends on the global state of the postsynaptic neuron.

2.2.2 - Metaplasticity

However, it is well known that repeated hebbian (associative) strengthening results in runaway synapses (Trappenberg, 2002, section 4.4; O'Reilly and Munakata, 2000, Ch. 4). Oja (1982) proposed to normalize the strength of synapses on a given neuron such that their sum remained constant. This approach eliminates the possibility that all synapses saturate at a prescribed maximum, but tends to isolate the most active synapse in a kind of principal component or eigen-value analysis. Bienenstock, Cooper and Munro (1982), hereafter referred to as BCM, introduced yet another approach where the threshold (θ_M) between LTD and LTP (antihebbian and hebbian learning) evolves with the recent activity of the postsynaptic neuron. The BCM method ensures the relative selectivity of each synapse while avoiding saturation of any of them except for the very unlikely case of a continuously activated synapse in a continuously spiking neuron. These methods (Oja's and BCM's) having been designed from the observer's point of view are not easily encapsulated in self-contained neuron model. Encapsulation is an essential property for biological plausibility, especially when looking for autopoiesis which depends on the existence of a well-defined boundary (membrane) determining operational closure.

Abraham and Bear (1996), Abraham (2008) and Abraham and Philpot (2009) have shown that biological metaplasticity is a reality in the brain even if the underlying mechanisms have yet to be clarified.

2.2.3 - Summary

From the preceding brief review of (computational) neuroscience, we can draw three principles which are necessary conditions for the proper functioning of the brain.

Principle 1. The neuron doctrine - « each neuron is a discrete cell [...] and [...] neurons are the signaling units of the nervous system » (Kandel al., 2013, p23).

Principle 2. Dynamic polarization - « electrical signals within a nerve cell flow only in one direction: from the receiving sites of the neuron, usually the dendrites and cell body, to the trigger region of the axon. From there the action potential is propagated along the entire length of the axon to its terminals. In most neurons studied to date electrical signals in fact travel in one direction. » (Ibid.).

Principle 3. Connectional specificity - « nerve cells do not connect randomly with one another in the formation of networks. » (Ibid.).

We can also derive three postulates identifying neuronal behaviors generally accepted as the basis for neural network development.

Postulate 1. Spiking neurons are « substantially more realistic » (Maass 1997 p1661) than previous models (McCulloch and Pitts neurons or rate neurons) even though they are still « simplified models that focus on just a few aspects of biological neurons » (Ibid. p1661).

Postulate 2. Spike-Timing-Dependent Plasticity (STDP) - The bidirectional change in synaptic efficacy (strengthening LTP or weakening LTD) is conditional on the activity of the specific synapse (i.e. local channels are open because the presynaptic neuron has fired) and proportional to the global activity of the postsynaptic neuron.

Postulate 3. Metaplasticity - The crossover from LTD to LTP as a function of postsynaptic activity varies according to the (recent) history of this postsynaptic activity (BCM 1982). Other changes in the properties of the neuron (e.g. "size") can affect its "learning" behavior.

Our objective will be to integrate these principles and postulates in a coherent model reproducing the semiotic and autopoietic behavior of a cognitive (intelligent) system. We are looking only for the emergence of some kind of sensorimotor intelligence which could be a stepping stone toward higher levels such as the concrete operational stage and, ultimately, the formal operational stage (Piaget).

2.3 - LIF (Leaky-Integrate and Fire neuron)

Having accepted the spiking neuron as the « most realistic » model (postulate 1) does not mean that we consider it to be complete. In fact, it is mainly a good representation of the trigger zone. The propagation of the AP from there to the axon's terminals is trivially perfect and instantaneous, but we already deemed this simplification to be acceptable. On the other hand, the representation of the dendritic tree is minimalist. It includes some consideration of the temporal integration of the input current to the capacitance, but the summation of spatially parallel stimulations (multiple synapses simultaneously excited) is strictly linear. The E/IPSPs[18] (e.g. Maass 1997) or the E/IPSCs[19] (e.g. Gerstner and Kistler 2002, Eliasmith and Anderson 2003, Izhikevich 2003) are summed linearly without explaining how they are produced. In other words, the model of the synapses is not well defined. Furthermore, the instantaneous reset of the entire dendritic tree to resting potential is also difficult to explain.

---

[18] E/IPSPs = Excitatory or Inhibitory PostSynaptic Potentials
[19] E/IPSPs = Excitatory or Inhibitory PostSynaptic Currents

This second problem (reset of the dendritic summation to zero) is still fairly common for all kinds of spiking neuron networks (e.g. Izhikevich 2003, 2004) especially that the consequences are very limited when the network is connected randomly. Precisely the kind of network that Markram attacked so vehemently in his letter (Adee 2009) against IBM's claims on DARPA's SyNAPSE project « These are point neurons (missing 99.999% of the brain; ... » Of course, Markram is a purist and, from an information processing perspective, the LIF model is doing better than 0.001% even if it is representing only the axon and even if it does it with « no detailed ion channels. » Still, he has a point: LIF-type models fall short of a complete neuron simulation.

Gerstner and Kistler (2002), in their figure 4.1, show the model (neuron) in two parts: a spiking LIF (soma) and an integrating low-pass filter (synapse). This approach addresses both the spatial and temporal integration of incoming current pulses into a postsynaptic current which decays with time according to the time constant of the low-pass filter. However, it does not solve the second problem (linear integration) since each incoming pulse produces an equal amount of postsynaptic current. Rospars and Lánský (1993), following Kohn (1989), proposed a stochastic model of a two-compartment neuron to eliminate the total reset of the dendritic tree.

Stein (1965) talks about « unit depolarization » summed up linearly to the threshold (« 4. If the depolarization reaches a threshold of r units, the neuron fires. » p.175).

Eliasmith and Anderson (2003 p84) represented the LIF as shown in Figure II.3. The input $J(t)$ is a current (even though it is sometimes referred to as a voltage/potential). This begs the question: where does this current come from? It makes sense in the context of physiology experiments where the AP is triggered by injection of a (steady) current via an intracellular electrode, but does it hold when the neuron is stimulated through a synapse by a presynaptic AP? Clearly, the modeling of the

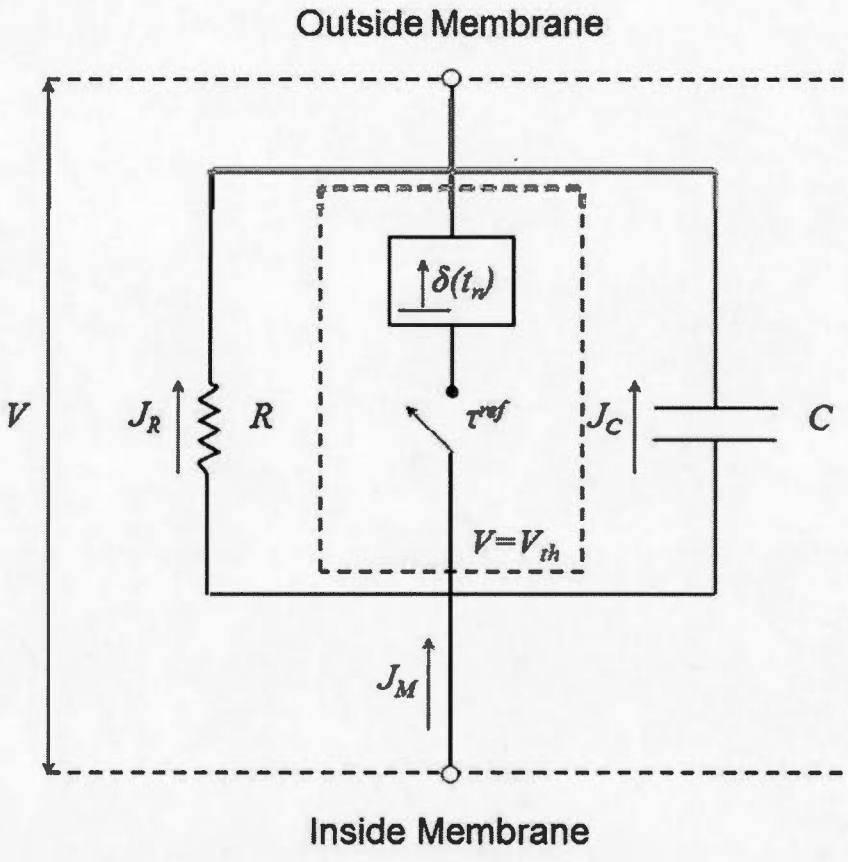


Figure II.3 - LIF according to Eliasmith and Anderson
(From Eliasmith and Anderson, 2003, p84, Figure 4.2)

synapses is not compatible with the model of the trigger zone. In the caption of that figure, Eliasmith and Anderson describe « [t]he active, super-threshold behavior » of the LIF as:

> [w]hen the membrane potential is equal to the voltage threshold, V_{th}, at a time t_n, the short-circuit switch is closed, resulting in a 'spike', $\delta(t_n)$. The switch remains closed, resetting the circuit and holding $V=0$, until it is opened after the refractory period, t^{ref}.

This is the short version; somewhat like describing Dretske's proverbial doorbell as: when the door button is pushed, the bell goes "ding dong". For the doorbell, a more

precise explanation would say: when the button is pushed, an electrical current flows through the copper wire inducing, via the solenoid, a magnetic field capable of projecting the hammer against a metallic plate that goes "ding"; when the solenoid is deenergized, the hammer falls back on another metallic plate with a "dong". Similarly, for the LIF, one could say: when the membrane potential reaches the voltage threshold, Na channels open letting in floods of positive ions until the threshold of K channels is reached opening neighboring channels and flushing out K^+ ions while Na channels further down the axon are already opening since their threshold has been reached, and propagating a causal chain which, by succession of Na channels and K channels, reaches the axon terminal where Ca^{2+} channels finally trigger the ejection of a quantum of neurotransmitters in the synaptic gap. It is this final ejection of neurotransmitters that is represented by the impulse in the inner block of Eliasmith and Anderson's diagram. It is also this impulse, or, as we have just explained, the quantum of neurotransmitters, that is responsible for the current *J(t)* potentially triggering the postsynaptic neuron.

So, the explanation of the impulse being caused by the triggering current is a short version, but it is an acceptable simplification as we have agreed to above. However, at the time, we mentioned that the linear summation of impulses to generate the triggering current in the postsynaptic neuron was an oversimplification. Maybe that part of the model was inherited from Stein's depolarization units, perhaps inspired by McCulloch and Pitts's summation and/or by Adrian's electrical experiments, but somehow it does not fit the expected role of neurotransmitters.

When the neurotransmitters cross the synaptic gap, they can act on chemically-gated channels causing a rush of ions in or out the nerve cell (Na^+ in for excitatory presynaptic neurons, K^+ out or Cl^- in for inhibitory presynaptic neurons. Technically it could be a mixture or hybrids of those, but we are interested only in the net effect.) Clearly the current is dependent on the potential difference between the two sides of

the membrane and the conductivity of this membrane. Each chemically-gated channel activated by a neurotransmitter increases the conductivity.

If there were voltage-gated Na channels in the neighborhood, they would quickly be triggered open by the rising membrane voltage, quickly followed by any K channels and the ensuing AP. However, this is generally not the case. We will come back to specific special cases later, but, for now, we will posit the simplifying hypotheses that there are no (or very few) voltage-gated channels in the dendritic tree like there are no (or very few) chemically-gated channels in the axon.

Simplifying hypothesis 1 (SH1):
There are no voltage-gated channels in the dendritic tree.
Simplifying hypothesis 2 (SH2):
There are no chemically-gated channels in the axon.

As a corollary to these SHs we could say that currents flow passively in the dendritic tree and actively in the axon. To be more exact, it also flows passively in the myelinated axons in between Ranvier nodes.

Whether we are talking ligand-gated channels in the dendritic tree or voltage-gated channels in the axon, the conductance-based model is « the simplest possible biophysical representation of an excitable cell, such as a neuron, in which its protein molecule ion channels are represented by conductances and its lipid bilayer by a capacitor » (Skinner 2006). We can see the similarities between this conductance-based model and the representation of the LIF (Figure II.3) by Eliasmith and Anderson (2003).

The same conductance-based model is also the basic module for compartmental neuronal modeling developed by Wilfrid Rall (1957, 1959, 1960, 2009) as the

neuroscientific application of the cable theory[20] originally developed by Lord Kelvin, in the 1850s, to model the signal decay in submarine telegraphic cables.

According to our SHs, in the dendritic tree we are mainly interested by chemically-gated channels as shown in Figure II.4. The opening of one channel under the influence of neurotransmitters affects directly and significantly the voltage in that compartment which goes back to equilibrium partially due to a small leak through the membrane, but mainly by electronic diffusion to neighboring compartments and so on through the entire tree. Detailed simulations based on the differential equations from the cable theory is possible with public domain software packages like GENESIS (GEneral NEuron SImulation System) developed at Caltech (see http://www.genesis-sim.org visited 2014.02.19) and NEURON developed at Duke and Yale Universities (see http://www.neuron.duke.edu visited 2014.02.19).

This compartmentalist approach is based on the assumption that the spatial distribution of synapses in the dendritic tree affects the impact of the incoming spikes

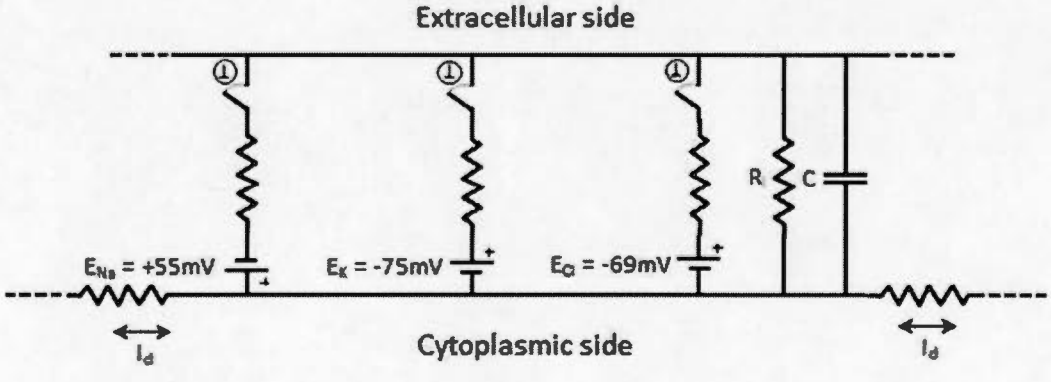


Figure II.4 - Conductance-based model

[20] Summarized in Niebur (2008).

on the voltage at the APs' triggering point giving the dendrites a computing ability, a power of decision on triggering or not an AP. Most other approaches neglect this spatial influence and consider that the temporal aspect supersedes any morphologic interactions such that the whole tree can be reduced to a single isoelectric compartment identified with the triggering point. We will look at a combined approach featuring a single compartment separated from the triggering point by some resistive connection. While rejecting the compartmentalists' hypothesis that the morphology has a significant impact on the fully developed neurons, we will postulate[21] that, being highly involved in the learning process, any morphology would find its way to a common equilibrium behavior for a set of given inputs.

Simplifying hypothesis 3 (SH3):
Dendritic morphology is trumped by temporal correlations in defining the impact of synaptic inputs on triggering of APs.

2.4 - DoubleLIF

2.4.1- The model

We are therefore proposing a neuron model based on a LIF (representing the axon) connected to a single compartment (representing the dendritic tree) via a single longitudinal resistance (as per SH3). Figure II.5 shows a complete artificial neuron with a surrounding line ($V_{ref} = 0\ mv$) representing the external side of the cell. On the left (A), we see a series of interconnections representing one or more synapses between m presynaptic neurons and the depicted neuron k each including a series of N_m chemically-gated excitatory ionic (Na^+) channels activated by impulse trains δ^+_{ijk} (i = 1, N and j = 1, m). Similarly, on the right (B), we see a series of interconnections

[21] This hypothesis, by itself, could be the subject of an interesting research project in computational neuroscience.

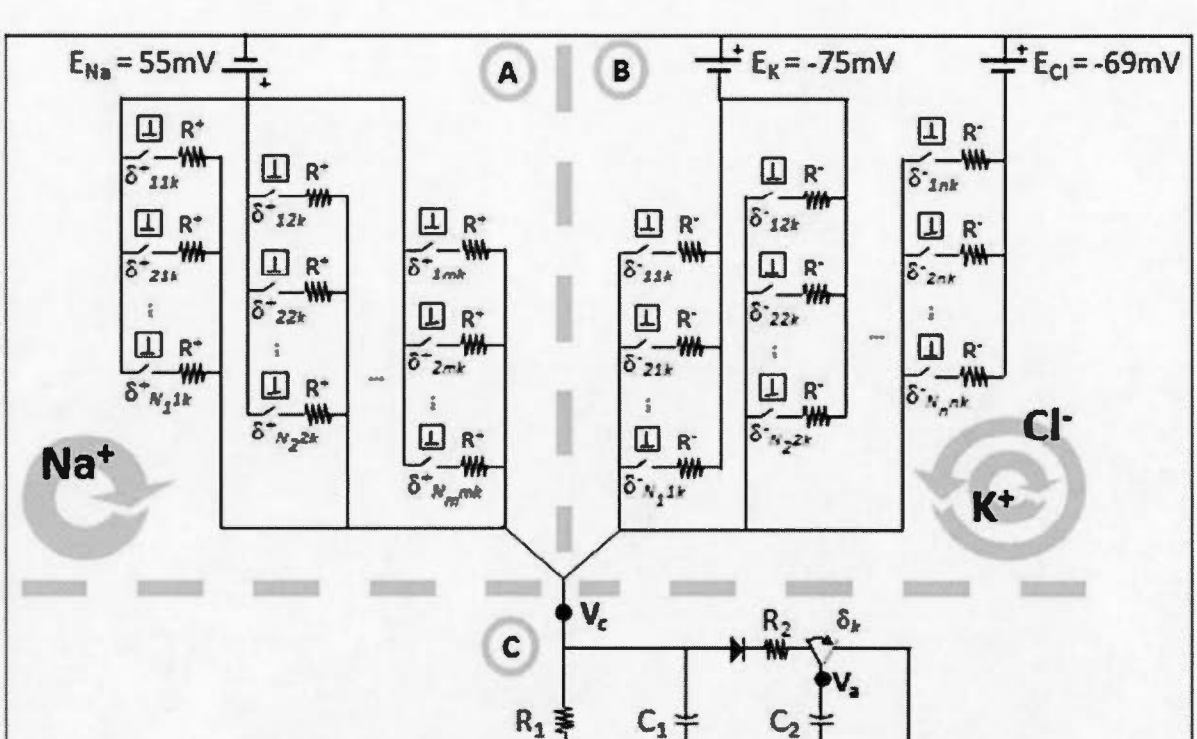


Figure II.5 - DoubleLIF model

Electric diagram of the DoubleLIF model. Section A represents the excitatory portion of the dendritic tree where incoming Na^+ ions tend to depolarize the cell. Section B represents the inhibitory portion of the tree where K^+ ions flowing out (or Cl^- ions flowing in) work against depolarization. Section C represents the soma and axon portions including the ionic pumps continuously repolarizing the cell by rebuilding concentration gradients for all ions between the inside and the outside, as well as an action potential triggering mechanism.

between n presynaptic neurons and the depicted neuron k each including a series of N_n chemically-gated inhibitory ionic (K^+ or Cl^-) channels activated by impulse trains δ^-_{ijk} (i = 1, N and j = 1, n). Of course, the segregation by type and the array-like arrangement are strictly for clarity purposes and the organization of real synapses is not that simple. To avoid excessive complexity, we will assume that ionic equilibrium at the "Y" intersection (connection of the A, B and C sections) is reached in the sub-millisecond time frame.

To begin with, we will look only at the dendritic tree, assuming that R_2 is infinite. In the most quiet state of the dendritic tree, all the switches ($\delta^{+/-}_{ijk}$) are open (i.e. the ionic channels are closed) and the cellular potential (V_c) rests at E_{rest} (-69 mV). When excitatory inputs start activating channels on the left side of our dendritic tree, V_c rises slowly proportionally to the number of channels activated and the frequency of their activation. V_c rises because C_1 accumulates charges while R_1 leaks the excess outside of the cell.

From an ionic perspective, this "R_1 leak" should bring the concentrations inside the neurons to equilibrate with the concentrations in the surrounding solution and all the E_{ion} would, according to Nernst law, become 0 mV thereby eliminating any potential reaction by the neuron. However, the ionic pumps restore the relative internal concentration of the different ions against the gradient imposed by the constant external concentrations. In other words, they keep the batteries (E_{ion}) fully charged at all times. So, the resting potential (-69 mV) is a homeostatic equilibrium resulting from the action of the ionic pumps and being disturbed by the opening of ionic channels activated by neurotransmitters.

Similarly, when inhibitory inputs start activating channels on the right side of the tree, V_c slowly decreases proportionally to the number of channels activated and the frequency of their activation. However, the ions in play are not the same that caused the rise of V_c on the left side and, although they work in conjunction with the ionic pumps to polarize the membrane, they also increase the workload of these ionic pumps for different types of ions.

The result of all this gating and pumping is that V_c can assume any value between -69 mV and +30 mV (and even exceed these values during temporary transient excursions) depending on the net instantaneous relative permeability of the overall tree including the effect of the ionic pumps.

Until now, assuming that R_2 was infinite, we have neglected a significant part of section C; the neuron is not firing. However, if we reduce R_2 such that a small current can trickle through it, C_2 will accumulate charges and V_a will rise (as long as V_c is larger than V_a) until a threshold (~-40 mV), dictated by the voltage-regulated ionic channel (represented by switch δ_k), is reached draining C_2 and resetting V_a to E_{rest} (-69 mV). It should be noted that this reset has only a marginal effect on V_c since R_2 is much larger than R_1 and only C_2 is short-circuited. It is, however, sufficient to trigger a self-propagating action potential along the axon as demonstrated by Hodgkin and Huxley (1952).

By comparison, a standard LIF neuron (see Figure II.6) has only one accumulator (C_0) coupled with one resistor (R_0) which define the current (I) required to bring the membrane potential (V_m) to the threshold and trigger an AP. In this case, V_m goes to

E_{rest} (-69 mV) and recharging C_0 starts from scratch with no memory of previous activation. In some applications, the current (I) is limited to the number of incoming impulses from that time to the next reset. At best (e.g. Gerstner and Kistler 2002,

Eliasmith and Anderson 2003), the remaining effect of previous impulses, dampened through first-order filters, is added to new incoming impulses similarly dampened.

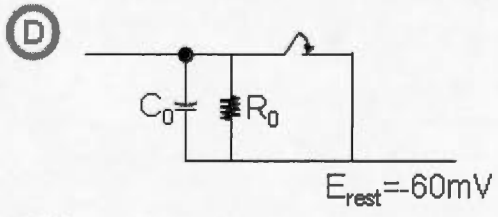


Figure II.6 - LIF model

2.4.2 - Biological plau

In the previous section, we have described the neuron as an electric circuit with some hints to the actual neurobiological equivalent. It should be clear that the electric

circuit is just a model (an approximate representation) of the ionic currents actually taking place in a real cell. Most importantly, the relative concentrations of the different ions are not explicit in the model and, as previously noted, the essential action of ionic pumps is not explicitly represented, but somehow implicit in the permanently charged batteries in the different branches of the circuit. Still, we consider that this DoubleLIF model includes the main features of a complete spiking neuron:

1. a dendritic tree,
2. excitatory and inhibitory synapses,
3. spatiotemporal integration (accumulation) of incoming signals,
4. very short term memory (milliseconds) of the resulting state (V_c),
5. translation of the state into an action potential through V_a, and
6. excitation or inhibition of other neurons as a result of these action potentials.

Before tackling the differential equations of Figure II.5, we will complete the neurological description of the process represented by the diagram.

The representation of the dendritic tree is evident from the diagram and we have multiplied the representations of ionic channels to carry the net impression of the volumetric and numeric importance of dendrites relative to the soma and the axon. As mentioned previously, we have segregated the excitatory and inhibitory synapses mainly because the ionic currents are different through the different channels involved. This means that from an electrical perspective the results are similar (although reversed), but the concentrations of different ions are affected. All the ionic channels in the dendritic tree are chemically-gated channels (SH3) and will therefore open only when activated by a neurotransmitter forced through the synaptic gap by an action potential on the axon of the presynaptic neuron. The switches (δ^{+}_{ijk}) on the left have been associated to channels which are excitatory with an influx of Na^{+} tending

to depolarize the dendritic tree (+55 mV). Similarly we have associated the switches (δ^-_{ijk}) on the right to inhibitory channels with an influx of Cl^- or an outflux of K^+ tending to polarize the tree (respectively -69 and -75 mV). As we don't have sufficient information to differentiate between Cl^- and K^+ channels, we will have to settle for a single type of inhibitory channels for which we will use K^+ properties.

The dynamic integration of these ionic fluxes results in a cellular potential (V_c) somewhere between these two attractors created by concentration gradients maintained by the continuous action of ionic pumps (certainly sodium-potassium pumps and probably chloride pumps). This cellular potential (V_c) represents the neuron's (very) short term memory which is then translated in the soma for transmission via the axon. The electronic diffusion resulting from the ionic fluxes reaches the first voltage-gated ionic channels in the axonic hillock triggering the propagation of an action potential along the axon as described by Hodgkin and Huxley (1952).

2.4.3 - Dynamics

The dynamics of the DoubleLIF include a mixture of current square pulses (inputs), continuous current leaks and discharge impulses. Essentially, we have a capacitance (C_1) accumulating positive (I_{EPSP}) and negative (I_{IPSP}) charges, continuously leaking to ground (I_{leak}) and charging the oscillator's capacitance (C_2) with a continuous current (I_{fire}) until instantaneous discharge to ground when C_2's potential (V_a) reaches the threshold (θ_S).

When a switch (δ^+_{ijk}) closes in the upper left quadrant (A) of fig. 1, the current through the resistance R^+ is

$$i_{EPSP}(t) = \frac{1}{R^+}\left(E_{EPSP} - V_c(t)\right) \qquad (1)$$

and similarly the current in the upper right quadrant (B) is

$$i_{IPSP}(t) = \frac{1}{R^-}(E_{IPSP} - V_c(t)) \qquad (2)$$

where $E_{IPSP} = E_K$ (assuming we neglect the Cl^- branch). Note that i_{IPSP} is always negative, while i_{EPSP} is always positive, since $E_{IPSP} < V_c < E_{EPSP}$.

These currents are not continuous; they last only for a short period of time (T_ε) resulting in a square pulse equivalent to an impulse of charge $q = i_{E/IPSP} \, x \, T_\varepsilon$. These impulses are quanta of charges transmitted to the postsynaptic neuron each time an action potential is triggered in a presynaptic neuron; more specifically, each time a neurotransmitter opens an ionic channel in the dendritic tree of a post synaptic neuron.

On the other hand, the losses through R_1 are continuous and equal to

$$I_{leak}(t) = \frac{1}{R_1}(V_c(t) - E_{rest}) \qquad (3)$$

and the firing current through R_2, also continuous, is

$$I_{fire}(t) = Max\left(0, \frac{1}{R_2}(V_c(t) - V_a(t))\right). \qquad (4)$$

The *Max* function is used since, as shown by the diode in section C of Figure II.5, the ions cannot flow back assuming they were involved in an irreversible synthesis of neuropeptides[22]. $V_a(t)$ is temporarily frozen until $V_c(t)$, which can decrease as low as E_{rest}, comes back to exceed $V_a(t)$.

---

[22] The effect of the diode is negligible. V_a is frozen at a given value when V_c becomes smaller and stays there until V_c becomes larger than V_a again. Without the diode, V_a would follow V_c down to 0 and back up afterwards (albeit with a small lag).

Whenever V_a reaches θ_S, C_2 is momentarily short-circuited to ground and instantaneously drained of all charges accumulated until then via I_{fire}. The resulting impulse is denoted as $\delta_k(f(t))$ where $f(t) = Min(0, V_a(t) - \theta_S)$ which is 0 whenever $V_a(t) > \theta_S$.

Applying Kirchhoff's law, we obtain the accumulation in C_1

$$C_1 \frac{dV_c(t)}{dt} = \sum_{j=1}^{m}\sum_{i=1}^{N_j} \delta_{ijk}^{+}\left(Min\left(0,\left(V_{a_j}(t)-\theta_S\right)\right)\right) i_{EPSP}(t) T_\varepsilon$$

$$+\sum_{j=1}^{n}\sum_{i=1}^{N_j} \delta_{ijk}^{-}\left(Min\left(0,\left(V_{a_j}(t)-\theta_S\right)\right)\right) i_{IPSP}(t) T_\varepsilon$$

$$-I_{leak}(t) - I_{fire}(t)\left(1-\delta_k\left(Min\left(0,\left(V_{a_k}(t)-\theta_S\right)\right)\right)\right) \qquad (5)$$

and in C_2

$$C_2 \frac{dV_{a_k}(t)}{dt} = I_{fire}(t) - \left(I_{fire}(t) + C_2\theta_S\right)\delta_k\left(Min\left(0,\left(V_{a_k}(t)-\theta_S\right)\right)\right). \qquad (6)$$

It should be noted that, in the last term of equation 5 and in equation 6, the impulse refers to the axonic potential (V_a) in the neuron k represented by the equations, while, in the first two terms of equation 5, impulses refer to the axonic potential (V_a) in presynaptic neurons j.

Since there is only one impulse per presynaptic neuron, it must be assume that all N_i channels between two synaptically connected neurons can be lumped together for calculation purposes and $\sum_{j=1}^{m}\sum_{i=1}^{N_j} \delta_{ijk}^{+}(f(t))$ can be replaced by $\sum_{j=1}^{m}\left(N_j.\delta_{jk}^{+}(f(t))\right)$ where N_j is the total number of ionic channels forming each of m excitatory synaptic connections between neurons j and k. The same applies for the n inhibitory connections between neurons j and k.

On one hand, DoubleLIF is clearly a spiking neuron where each input spike is treated individually producing a non-linear effect on the cellular potential (V_c). On the other hand, DoubleLIF is also a rate neuron since this cellular potential (V_c) determines the frequency of the output spikes. A constant V_c produces a constant output firing rate, but V_c is rarely constant and continuously changes with time due to multiple inputs coming in at different frequencies. V_c is therefore a direct representation of an instantaneous output frequency which, in other models, can only be approximated by some running average over a time window with some inherent time delay.

2.4.4 - Plasticity

The model described in the previous section assumes that information is passed from a presynaptic neuron to a postsynaptic neuron through synapses made of multiple channels totaling N_j units[23] of connection. In other words, the strength (or intensity) of the connection is proportional to N_j. The information transferred is the state of the presynaptic neuron represented by the cellular potential (V_c) resulting from a spatio-temporal integration (running average equivalent) of all the signals connected to its own dendritic tree via similar synapses. The spiking frequency of this presynaptic neuron is directly proportional to its V_c which is the driving force for the axonic oscillating accumulator (see appendix A).

However, Hebb's « firing together, wiring together » tells us that the connections' strength is not constant, but changes with time. In other words, the equations should refer to $N_j(t)$. As we have seen in section 2.2.1, the changes in connection strength ($dN_j(t)/dt$) can be positive (long term potentiation - LTP) or negative (long term

---

[23] A unit of connection is one ionic channel with a conductivity of $1/R_0$. Channels are added one at a time, but their efficiency matures with time. Therefore N_j does not have to be an integer and can take real values.

depression - LTD) and the neuron, fully encapsulated by cellular definition, has very limited information to make the difference.

In its simplest expression, the LTP/LTD paradigm can be summarized as: high frequency stimulation (HFS) produces LTP and low frequency stimulation (LFS) produces LTD. In the DoubleLIF model, this translates to: HFS produces high V_c which favors LTP and LFS produces low V_c which favors LTD.

2.4.4.1 - DoubleLIF and BCM theory

The Bienenstock-Cooper-Munro (BCM) theory (1982), developed for rate neurons, states that the strength of the synapse is increased (LTP) when the postsynaptic activity is high and decreased (LTD) when the postsynaptic activity is low. This promotes selectivity by favoring cooperating connections which are in phase with the postsynaptic neuron and hindering connections which are out of phase. Frequency-based rules are well-suited for rate neurons. To separate between high and low postsynaptic activity or frequency, BCM includes a threshold frequency (θ_M) which is not fixed but depends on the history of postsynaptic activity. In other words, BCM slowly adjusts the threshold frequency to match the mean firing rate of the postsynaptic neuron thereby balancing the effects of LTD and LTP. DoubleLIF's inheritance from rate neurons makes it particularly well suited for the implementation of BCM theory in spiking neurons.

In its original form (Bienenstock, Cooper and Munro 1982), BCM is given as:

$$c = \Sigma_j m_j d_j \quad (7)$$

$$dm_j/dt = \varphi(c)d_j - \varepsilon m_j \quad (8)$$

where $\varphi(c) < 0$ for $c < \theta_M$ and $\varphi(c) > 0$ for $c > \theta_M$ (9)

with a note stating that « [t]he term, $-\varepsilon m_j$, produces a uniform decay of all junctions [which], in most cases, does not affect the behavior of the system if ε is small enough.»

It should be noted that BCM explicitly states (Eq. 7) a direct correlation between output frequency (c) and input frequency (d) hinting at a natural causal relationship between the two variables. According to Blais and Cooper (2008 - with symbols adapted for equations 7-9 here above):

> The BCM theory of synaptic plasticity … is based on … three postulates.
> 1. The change in synaptic weights [dm_j/dt] is proportional to presynaptic activity ([d_j]).
> 2. The change in synaptic weights is proportional to a non-monotonic function (denoted by φ) of the postsynaptic activity ([c]):
>   1. for low [c], the synaptic weight decreases ([$dm_j/dt < 0$])
>   2. for larger [c], it increases ([$dm_j/dt > 0$])
>
>   The cross over point between [$dm_j/dt < 0$] and [$dm_j/dt > 0$] is called the modification threshold, and is denoted by θ_M.
> 3. The modification threshold (θ_M) is itself a super-linear function of the history of postsynaptic activity [c].

While BCM correlates directly the output frequency (c) to the input frequencies (d_j) by the strength of the synapse (m_j), DoubleLIF uses the state variable V_c provided by the added accumulator as an intermediate step between input and output activities. Equation (5) shows that V_c results from the integration over time of I_{EPSP}, I_{IPSP}, I_{leak}, and I_{fire}. We will assume that there is no inhibitory stimulation and neglect I_{IPSP} for the purpose of this discussion. Combining (4) and (6) and holding V_c constant, it can be shown (see appendix A) that the spiking frequency is directly proportional to V_c. On the other hand, combining (1) and (5), it can also be shown that V_c is directly proportional to the temporal integration of the I_{EPSP} term which represents the spatial integration (Σ_j) of the strength (m_j) of the input impulses at each moment of time (i.e. no running average).

To translate (8) from the rate neuron formalism of BCM to the spiking neuron formalism of DoubleLIF, we propose to replace it by

$$dN_j(t)/spike = 0 \qquad \text{for } V_c \leq \theta_S \qquad (10)$$

$$= k_j(V_c(t) - \theta_S)(V_c(t) - \theta_V)N_j(t) \qquad \text{for } V_c > \theta_S \qquad (11)$$

where θ_V is the potential equivalent of θ_M and

θ_S is the spiking threshold of the neuron in the rate neuron paradigm.

This means that, for $V_c \leq \theta_S$ (i.e. when the postsynaptic neuron is not spiking), there is no potentiation nor depression of the synapse for any spiking frequency of the presynaptic neuron, which is an expected behavior in the rate neuron paradigm where potentiation and depression are dependent on postsynaptic activity. However, in a true spiking neuron paradigm, θ_S should be replaced by 0 (resting potential) since it has been demonstrated experimentally that depression can be induced at subthreshold level of stimulation (e.g. LTD protocol in Enard al. 2009).

In fact, the BCM model (see Equation 7) does not consider any subthreshold level of excitation; even the smallest input activity will generate an output frequency. In the absence of any other stimulation, VLFS (Very Low Frequency Stimulation) will always induce LTD. The decay term ($-\varepsilon m_j$) in the plasticity equation helps, among other things, to compensate for this deficiency. As discussed later, we will neglect this decay term for DoubleLIF.

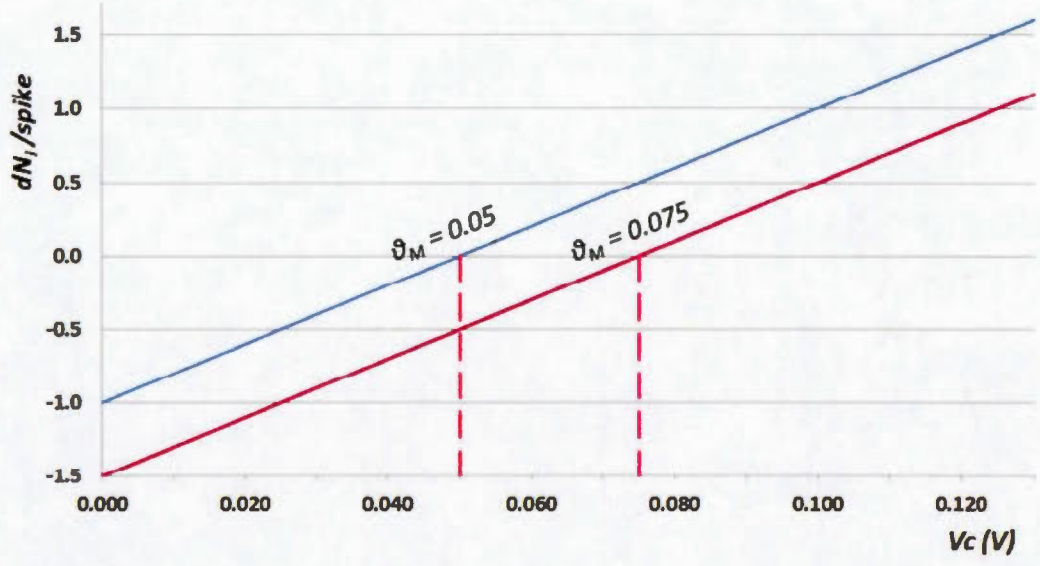


Figure II.7 - DoubleLIF - Linear dN/spike

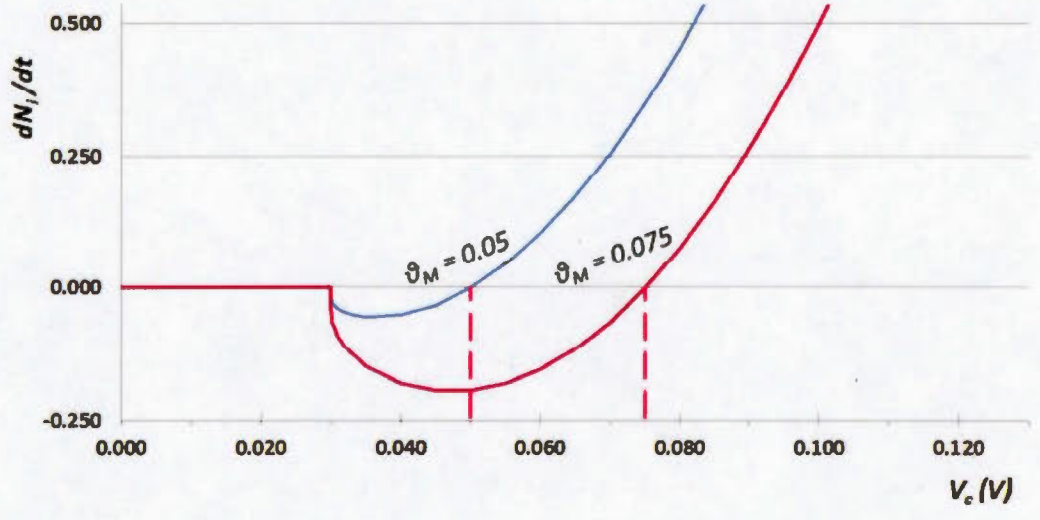


Figure II.8 - DoubleLIF - Non-linear dN/dt

Multiplying $dN_j(t)/spike$ by $\delta_{ijk}(t)$ (incoming impulses) in the temporal integration is equivalent to $dN_j(t)/dt$ based on average frequency.

2.4.4.2 - DoubleLIF and STDP

However, nowadays in the spiking neuron paradigm, the prevalent plasticity theory is Spike-Timing-Dependent Plasticity (STDP) which tells us that, in controlled pairing experiments, the important variable in the determination of $\Delta N_j(t)$ is the time difference between the input spike and the output spike. In its basic form, STDP is a satisfactory model for specific testing protocols involving a pair of pre- and postsynaptic spikes at fairly low frequency (<5 Hz) as depicted in Figure II.9 (Sjöström and Gerstner 2010).

With the symbols in the figure, the data points can be correlated using the following equations:

$\Delta w_{ij}/w_{ij} = A_+ exp(-\Delta t/\tau_+)$ for $\Delta t > 0$ *and*

$\Delta w_{ij}/w_{ij} = -A_- exp(-\Delta t/\tau_-)$ for $\Delta t < 0$.

where $\Delta t = t_j^f - t_i^f$

t_j^f being the firing time of the presynaptic neuron and

t_i^f the firing time of the postsynaptic neuron

Parameters can be estimated for the curves shown in Figure II.9: $A_+ = 0.82$, $\tau_+ = 19$ ms, $A_- = 0.28$, $\tau_- = 27$ ms (as shown by the dotted lines superimposed over the original curves). However, these parameters apply strictly to data generated according to the protocol followed by Bi and Poo in their experiments; any departure from their protocol is likely to produce (slightly) different parameters. Izhikevich and Desai (2003), using data from Froemke and Dan (2002), arrived at $A_+ = 1.03$, $\tau_+ = 14$ ms, $A_- = 0.51$, $\tau_- = 34$ ms.

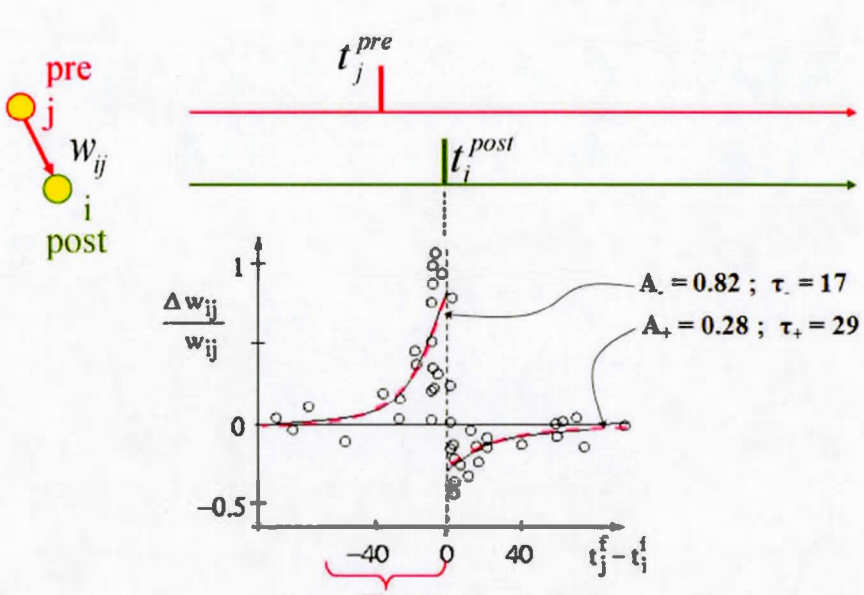


Figure II.9 - Spike-Timing Dependent Plasticity (schematic)
Spike-Timing Dependent Plasticity (schematic): The STDP function shows the change of synaptic connections as a function of the relative timing of pre- and postsynaptic spikes after 60 spike pairings. Schematically redrawn after Bi and Poo (1998)
(Copied with legend from Sjöström and Gerstner 2010)

So, when Sjöström and Gerstner (2010) sums up these synaptic changes linearly, it is valid only for multiple repetitions (e.g. 60) of the same protocol at low frequency (<5 Hz) such that the pre- and postsynaptic neurons have time to return to steady-state equilibrium in between repetitions.

Above that frequency, unwanted spikes start to appear in the window of interest (+/- 100 ms) around the pre- (or post-) synaptic spike and a strategy has to be established to deal with all the spikes in that window. A typical problematic example is a protocol involving a presynaptic spike followed by a postsynaptic spike 10 ms later and repeated at 50 Hz (every 20 ms). There are 10 pairs of spikes in any 200 ms window and it is impossible to differentiate between pre-post and post-pre pairing

since all postsynaptic spikes are equidistant from the preceding and following presynaptic spikes and vice-versa. *In vivo,* neurons are stimulated by tens (maybe even hundreds) of presynaptic neurons firing at frequencies of up to hundreds of Hz (this means inputs at 30 kHz for 100 presynaptic neurons at an average 300 Hz frequency or as much as 500 kHz for 1000 neurons at a 500 Hz maximum frequency); a situation which cannot (yet) be duplicated in a Petri dish for controlled observation.

When it comes to online implementation of STDP, moving from Petri dish simulation to *in vivo* simulation, at least two additional variables must be taken into consideration: first, the postsynaptic voltage at the time of the spike and, second, the spiking frequency (Sjöström al., 2001; Izhikevitch and Desai, 2003; Lisman and Spruston, 2005; Clopath al., 2009; Clopath and Gerstner, 2010; Lisman and Spruston 2010; Shouval al. 2010).

It is important to note that, in the STDP protocol, the delay rule supersedes any causal relationship between the pre- and the postsynaptic spikes and forces the two neurons to spike at the same frequency. This frequency is somehow embedded in the parameter set resulting from a given experiment such that any naturally produced (homosynaptic or heterosynaptic) postsynaptic spike comes in at the wrong frequency (for the parameter set) and corrections must be added to the model. The triplet strategy (Clopath al., 2009; Sjöström and Gerstner, 2010) indirectly brings in information about the postsynaptic neuron's spiking frequency via the interspike interval of the two postsynaptic spikes taken into consideration. The voltage-based skeleton of the model used to generate postsynaptic spikes from presynaptic stimulations provides all required information about postsynaptic cellular potential to adjust for a specific STDP protocol. Since, in such protocols, the spiking frequency of the presynaptic neuron, and consequently that of the postsynaptic one, are fixed, and, due to the regularity of the repetitions, the cellular potential of the postsynaptic neuron is approximately the same each time a backpropagating AP is triggered, a set

of constant parameters can be successfully generated to simulate the specific experiment. However, these parameters are not valid when input frequency is significantly increased like for *in vivo* modeling.

As shown on Figure II.7, for DoubleLIF, the synaptic change per presynaptic spike (*dN/spike*) is linearly proportional (slope to be determined) to the postsynaptic cellular potential (V_c) which is equivalent to assuming that the probability of a postsynaptic spike is distributed evenly over the window of interest and, over time, yields an average contribution which can be integrated with the voltage effect. This is consistent with Sjöström and Gerstner's (2010) paragraph on Voltage dependence:

> [...] the voltage of the postsynaptic neuron just before generation of action potentials influences the direction of change of the synapse, even if the spike timing is held fixed (Sjöström al., 2001), suggesting that postsynaptic voltage is more fundamental than spike timing. Indeed, a model of synaptic plasticity that postulates pairing between presynaptic spike arrival and postsynaptic voltage contains STDP models as a special case (Brader al., 2007, Clopath al., 2008).

If the input spikes were perfectly synchronized with output spikes (same frequency as in STDP protocols), the synaptic change per unit time (*dN/dt*) would become quadratic (see Figure II.8) since the output frequency (φ in spikes/second) is proportional to V_c and $dN/dt = \varphi * dN/spike$. However, *in vivo,* all inputs are not synchronized with the output and we will see later how the covariance of their frequencies affects the selectivity of the connections.

Figure II.8 shows a point (θ_M) on the horizontal axis (V_c) where the synaptic change (*dN/dt*) goes from negative to positive or, in other words, where the plasticity goes from LTD to LTP as in the BCM model and as in special cases of STDP according to Izhikevitch and Desai (2003) and Clopath and Gerstner (2010).

In summary, while STDP is a more precise model of the laboratory experiments where pairs of spikes are observed at low frequency, DoubleLIF provides a better averaging when input frequencies are very high and the postsynaptic spikes are not artificially induced on specific time delays but result precisely from these input frequencies and connections' strength. We should not lose sight of our objective: we are interested in biological plausibility, but only inasmuch as it is necessary for cognition.

2.4.5 - Metaplasticity

As mentioned earlier (section 2.2.2), it is well known that repeated hebbian (associative) strengthening results in runaway synapses and constraints must be added to models to ensure stability.

2.4.5.1 - Constraints

In DoubleLIF, V_c is naturally bounded by the potential reversal of Na^+ (E_{Na} = 55 mv) and of K^+ (E_K = -75 mv). When all gated channels are closed, it finds equilibrium at E_{rest} = -69 mv. V_c would never exceed 55 mv even if the voltage-gated Na channels got stuck open. The same applies for -75 mv in the case of K channels. Since Lapique (1907), it is known that a minimum depolarization (V_c = -40 mv) is required to trigger a spike. So, the output frequency, directly related to V_c (see appendix A), is also bounded at 0 Hz for θ_S (-40 mv) and some maximum frequency for 55 mv depending on the time constant ($\tau_2 = R_2C_2$).

On the other hand, V_c depends on input frequency, the strength of the connections, the capacitance (C_1) of the dendritic tree (in other words, its volume) and the size of the leak due to the conductance ($1/R_1$) of the membrane at rest. In fact, V_c never reaches 55 mv because the leak through R_1 must equilibrate with the incoming current through ligand-gated channels and this current would be zero for Na^+ ions at 55 mv.

We will assume that V_c never exceeds 30 mv which, rounding E_{rest} to -70 mv, gives us a span of 100 mv for the analog signal and could as well be interpreted as 100% of span. This allows us to realize that the important feature of the neuron is the homeostatic force field created by the controllable (gated) selective ionic channels. This force field has a different equilibrium than those generated by the two forces taken separately: ionic concentrations and potentials would normally equalize on both sides of the membrane. Thanks to the action of ionic pumps, they settle at a point which allows reaction to changes in the environment.

Cooper al. (1979) introduced the notion of a « modification threshold » as a constant marking the transition from LTD to LTP, but the system was not robust and all inputs could disappear if the output frequency fell below the selected constant θ_M. The BCM model (Bienenstock al. 1982) replaced the constant by $\theta_M(t)$, a function of time, implying that plasticity evolves with time, which can be referred to as « Metaplasticity: the plasticity of synaptic plasticity » (Abraham and Bear, 1996; see also Abraham and Philpot 2009, and Abraham 2008 for a comprehensive review).

Until BCM, the only parameters changing with time in neuron models was the connection weights; with BCM, θ_M also becomes activity dependent and changes with time.

2.4.5.2 - Stability

While, in BCM, $\theta_M(t)$ is a global property of the postsynaptic neuron representing the running average of the output frequency, in DoubleLIF, $\theta_M(t)$ becomes a local property of each synapse representing the running average of the postsynaptic neuron's cellular potential V_c (hence indirectly the output frequency) when the presynaptic neuron is firing. This can be interpreted as an approximation of the covariance of the pre- and postsynaptic neurons' potentials and indirectly of their instantaneous frequency.

Applying a HFS to a presynaptic neuron already capable, on its own, of eliciting a spike in the postsynaptic neuron can produce LTP of the interconnecting synapse. As the synapse strengthens, the postsynaptic potential keeps rising and θ_M follows such that the probability of LTP and the ΔN diminish up to the point of vanishing when the potential reaches V_{max}. The same applies to LTD when the presynaptic neuron is subjected to LFS and θ_M tends toward zero. So, this variable θ_M ensures that the weights do not grow or diminish forever without having to san arbitrary constraint on their value. The implicit assumption is that, in nature, all stimulations have a finite maximum intensity which will correspond to a finite maximum connection strength.

2.4.5.3 - Selectivity

The previous description applies to homosynaptic stimulation. Generally, there are many synapses competing to connect to one postsynaptic neuron. If a group of neurons jointly produce the equivalent of HFS, they will cooperate in maintaining V_c above θ_M and this heterosynaptic stimulation will favor strengthening of all their synapses albeit at different rates depending on their relative frequencies. If some presynaptic neurons spike at a relatively low frequency, they will benefit only marginally of the strengthening boost and, if they happen to spike at higher frequencies when the group is quiet, they will auto destroy their own connection. So, one or a group of neurons take control of a common postsynaptic neuron and favor the connection of associated (with highly correlated instantaneous frequencies) neurons while they let non-correlated ones slowly eliminate their connection. For each synapse, θ_M finds an equilibrium where the running weighted sums of LTP and LTD cancel one another.

2.4.5.4 - Neuronal development and bootstrapping

Metaplasticity is not limited to BCM's changing θ_M as can be seen in Abraham's review (2008). In DoubleLIF, we could consider having other parameters changing with time including R_1, C_1, R_2 and C_2. For now, there are clear advantages at keeping

R_2 and C_2 constant. That gives us a standard neuron with a given frequency to voltage response. We are not suggesting that all the neurons of a human brain have such a standard response, but rather that the diversity found in a human brain is (maybe) not absolutely required for a simpler cognitive system. On the other hand, we think that the adaptation of R_1 and C_1 is useful, maybe even essential, to understand the development and bootstrapping of neurons.

To implement these cellular modifications, we propose two differential equations:

$$dC_1/spike = k_C\, C_1(t) \quad \text{for } V_c > V_{max} \quad (12)$$
$$= 0 \quad \text{for } V_c \leq V_{max} \quad (13)$$
$$dR_1/spike = -k_R\, R_1(t) \quad \text{for } V_c > V_{max} \quad (14)$$
$$= 0 \quad \text{for } V_c \leq V_{max} \quad (15)$$

where k_C and k_R are positive gains for the increase of C_1 and the decrease of R_1 respectively.

Figure II.10 shows the development of a nascent sensor stimulated by a constant stimulus. The top portion displays V_c and V_a during the first 50 ms of the first 8 seconds of stimulation. It can be seen that, at the beginning, V_c exhibits a bang-bang behavior as the very small capacitance (1 pF) fills instantaneously under stimulation to empty immediately into the second capacitance in the following millisecond. After some 5 seconds, the first capacitance has grown sufficiently to exceed the loading rate of the second capacitance even though the leaking current has increased as the conductivity of the membrane has increased due to the reduction of its resistance from an initial 3 MΩ to slightly more than 2 MΩ. These changes can be seen in the very first seconds of the bottom portion. The middle portion shows the firing rate of the neuron starting at 125 Hz for 5 seconds and climbing to a steady state equilibrium of 215 Hz. The oscillator's parameters, R_2 and C_2 have been set to limit the rate to 250 Hz when V_c reaches maximum depolarization (30 mV or 100 mV above resting potential) which is equivalent to setting a 4 ms refractory period. The bottom portion

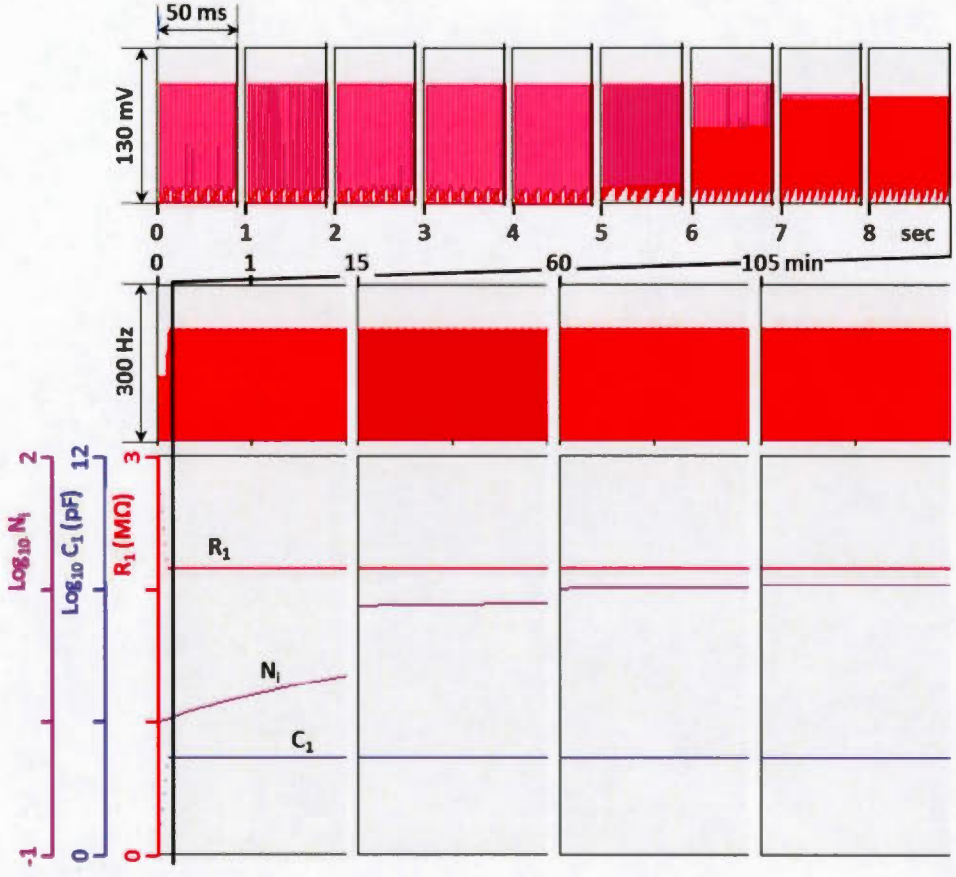


Figure II.10 - Neuronal (meta)plasticity

shows that the connection strength (N_i) continues to grow asymptotically even after the 8 seconds required to stabilize R_1 and C_1. The evolution over nearly 2 hours can be seen in two-minute snapshots after 15, 60 and 105 minutes.

2.4.6 - Special cases

Considering the great diversity of neurons in human brains, it is very easy to find exceptions which could not be directly modeled by DoubleLIF as presented until now.

2.4.6.1 - Purkinje cells

Purkinje cells are a very peculiar type of neurons which react more like a network of neurons than a single one. Although it is not our intention to model specific types of neurons individually, if we had to, it would be necessary to make an exception to the basic "one cell, one neuron" rule and to simulate a Purkinje cell using as many DoubleLIF neurons as required to represent the internal stimulation paths.

2.4.6.2 - Axo-axonic synapses

Some other neurons present direct connections of their axon to the axon of another neuron. This implies the presence of ligands-gated channels on the second axon which is in contradiction with our second simplifying hypothesis (SH2). We assume that the addition of an extra DoubleLIF neuron could produce equivalent results.

2.4.6.3 - Inhibitory neurons

Since our main objective is not to simulate specific neurons, the workarounds presented in the previous two cases are perfectly acceptable. We have mentioned inhibitory neurons before and indicated that they were essential to the operation of the brain. DoubleLIF includes the simulation of such inhibitory neurons, but we have not discussed their plasticity rules. Some articles (Hass al. 2006; Lamsa al. 2010) show that their synaptic plasticity is very similar to that of excitatory neurons. Maffei (2011) provides a few leads on the synaptic plasticity and metaplasticity of inhibitory neurons, but remains inconclusive on the exact phenomena. We will have to explore these possibilities in the numerical implementation of the model.

2.5- Conclusion

In this section, we have shown that the DoubleLIF model is a complete representation of the neuron from the synapses in the dendritic tree to the axonic terminal. Each neuron can have excitatory and inhibitory stimulations, but produces only one or the

other. Its response is proportional to the intensity of the stimulation with non-linearity depending on the existing level of polarization.

We have suggested a plausible neurobiological explanation (at the cellular level, not the molecular level) clearly indicating where the electrical model combines ionic currents of Na^+, K^+ and Cl^- in a single net current hiding the important role of the ionic pumps in maintaining a homeostatic equilibrium responsible for the neuronal activity.

We have also shown that the newly-added accumulator brings into play a new state variable, the cellular potential, which is well-suited to develop a translation of the Bienenstock-Cooper-Munro (BCM) theory from rate neurons to spiking neurons on the basis of a voltage-dependent synaptic plasticity (without any need for externally computed average firing rates). The metaplasticity introduced by BCM can also be extended to the adaptation of other parameters of the same accumulator and interpreted in terms of cellular development.

As shown in Table II.1, the main innovation of DoubleLIF is its ability to include metaplasticity in a fully encapsulated biologically plausible state of the art spiking neuron. The concept of metaplasticity is not new since it was already an integral part of the BCM model (1982) and an external add-on of multiple models since Oja (1979). The BCM model was decidedly a rate neuron model thereby lacking proper individual treatment of input spikes in the dendritic tree. Third generation spiking models, like the LIF and other point neurons, were not encapsulating the dendritic tree although some (Gerstner and Kistler 2002, Eliasmith and Anderson 2003) were properly processing the inputs as square current pulses into a charging capacitance (but without relating this capacitance to any biological equivalent). So, the real question is not: "How much better than the LIF is DoubleLIF reproducing Hebbian learning, or even LTP/LTD learning or STDP learning?", but rather: "How are

Table II.1 - Summary table - State of the art

	Lapicque	Hodgkin and Huxley	McCulloch and Pitts	Stein (LIF or other point neurons)	Rosenblatt (Perceptron)	Bienenstock, Cooper and Munro	PDP Research Group (MLP)	Multicompartments	Izhikevich	Clopath	DoubleLIF	Biological plausibility	Cognitive necessity
Threshold	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Action potential		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Spiking dynamics		✓		✓		✓?		✓	✓	✓	✓	✓	
Detailed spiking dynamics		✓						✓	✓			✓	✘
Networking			✓	✓	✓	✓	✓	✓?	✓	✓	✓	✓	✓
Learning (Hebb)				✓		✓		✓	✓	✓	✓	✓	✓
Backpropagation							✓					✘?	
LTP/LTD						✓		✓	✓	✓	✓	✓	✓
STDP				✓				✓?	✓	✓	?	✓	
Encapsulation				✘	✘	✘	✘	✓?	✘	✘	✓	✓	
Metaplasticity				✘	✘	✓	✘	✘	✘	✘	✓	✓	✓

DoubleLIF learning capacities (including bootstrapping) developing under continuous stimulation (something that cannot even be tested with the LIF and other point neurons)?" If point neurons are too simple, we could consider

multicompartment models if they provided some insight into their development instead of overparameterizing a static view of their current structure. Izhikevich's model (2003), like multicompartments, are very good at mimicking numerous types of spiking patterns, but there is no discussion of the impact of these patterns on the neurons' cognitive abilities. Clopath's model (2009), fundamentally voltage-based like DoubleLIF, uses STDP learning based on results from experimental protocols with frequency-dependent parameters not related to any biological component. Stretching our imagination, we might see some connection between MLP's error backpropagation and the biological process triggered by backpropagating action potentials, but it remains difficult to grant biological plausibility to the mathematical formulation of the backpropagation algorithm.

The table is filled with our best understanding of the different neurons' properties, but it clearly implies judgement calls biased by our specific research objectives and based on Ockham's razor and Einstein's caveat. At some point, one has to define in the details what is necessary and what is sufficient: point neurons (e.g. LIF) are too simple, multicomponent neurons are too complex, could two-point neurons (DoubleLIF) do the job?

We contend that the level of activity-driven metaplasticity implemented in DoubleLIF is necessary and sufficient to support our hypothesis of information-fed second level of autopoiesis for the development of neuron networks. We are probably still guilty of some gross oversimplification, but we hopefully have added some refinement in the conception, if not the explanation, of the complex dynamics of neuronal communication and its self-organization. So defined, DoubleLIF has the properties, semiosis and autopoiesis, identified in the previous chapter as essential to the emulation of cognitive systems. Some might say that, along the way, we have lost some credibility about biological plausibility considering the lack of supporting evidence for the dynamics of neuronal growth, but, although extremely young, a

field, called « dynamic morphometrics » (Chen and Haas 2011), is being developed with the help of emerging technologies (e.g. single-cell electroporation, two-photon microscopy). The field is based on the *synaptothrophic hypothesis* elaborated by Vaughn (1989) and stating that:

> [...] the formation of synaptic junctions may take place as an ordered progression of epigenetically modulated events wherein each level of cellular affinity becomes subordinate to the one that follows. The ultimate determination of whether a synapse is maintained, modified or dissolved would be made by the changing molecular fabric of its junctional membranes. ... Key elements of this hypothesis are 1) epigenetic factors that facilitate generally appropriate interactions between neurites; 2) independent expression of surface specializations that contain sufficient information for establishing threshold recognition between interacting neurites; 3) exchange of molecular information that biases the course of subsequent junctional differentiation and ultimately results in 4) the stabilization of synaptic junctions into functional connectivity patterns.

In this definition, Vaughn does not explicitly mention autopoiesis, but this epigenetic activity-driven self-organization is functionally very close to what we are trying to achieve, at the most simplistic level, with activity-driven development of DoubleLIF's parameters (C_1 and R_1).

The next step of the research project will investigate the behavior of DoubleLIF neurons in pre-wired networks to verify if they could autonomously develop strictly on the basis of external stimulation. Ultimately, we expect to show that "free-wiring" networks of DoubleLIF neurons can develop and organize themselves when externally stimulated.

CHAPTER III

Numerical simulation and experimentation

In the previous section, we have described the differential equations representing the dynamics governing the operation and the development of neurons. These equations cannot be solved analytically, but we can provide a numerical approximation. Considering that neurons typically fire at frequencies in the order of hundreds of Hertz, we will select one millisecond (1 ms) as the integration time step which will allow us to process spiking rates as high as 500 Hz without losing any information. This frequency is an acceptable compromise allowing a strict adherence to the spiking paradigm's digital aspects (the $\delta^{+/-}$ impulses and the transformation of dX/dt into $\Delta x/spike$) while fully representing the continuous behavior of the electrical analog model components.

We will now describe the prototype developed to instantiate the DoubleLIF model. This prototype is only a proof of concept since the environment and the body are included in the simulation and thereby at the same information level as the brain which is meant to be an emulation precisely because it exists only at that information level. If we could have a physical version of this simple body in a similar physical environment, we should be able to use exactly the same informational model of the brain. However, physical bodies and environment are usually much more complex than what we could simulate here and the simplistic nature of that simulation is decidedly an advantage when it comes to explain the most basic principles. True robotics is left for future developments.

3.1 - The numerical simulation

The simulation (written in Java) is composed of two threads. A main thread handling the simulation of the brain, the body and the environment and starting a second thread for the graphical user interface (GUI).

3.1.1 - The GUI

In the graphical user interface (GUI) of Figure III.1, we can see:

1. A schematic top-view of a simple environment (top left rectangle) including, in this case, an agent (quasi-triangular shape) and two stimulation sources (*A* and *B*).
2. A schematic of the agent's simplistic brain, in this case three neurons (top right rectangle).
3. A set of buttons to control the simulation (top center).
4. A graph of a selected neuron's cellular and axonic potentials (V_c and V_a) showing 1000 milliseconds of history continuously rewritten from the left (second row).
5. A graph of the same selected neuron's spiking rate showing 900 seconds of history also continuously rewritten from the left (third row); followed on the right by the average spiking rate during the last second.
6. A graph of the same selected neuron's main characteristic variable parameters R_I, C_I, and strength of all connected synapses (fourth row) followed on the right by a menu of different scenarios.

3.1.1.1 – The environment window

The environment window provides an overview of the agent's behavior and means to activate/deactivate stimulation sources. These sources can be seen as emitting a stimulant (light, odor, etc.) and affect specific sensors. Sensors react to one and only one type of stimulant, but can be affected by multiple sources simultaneously. The

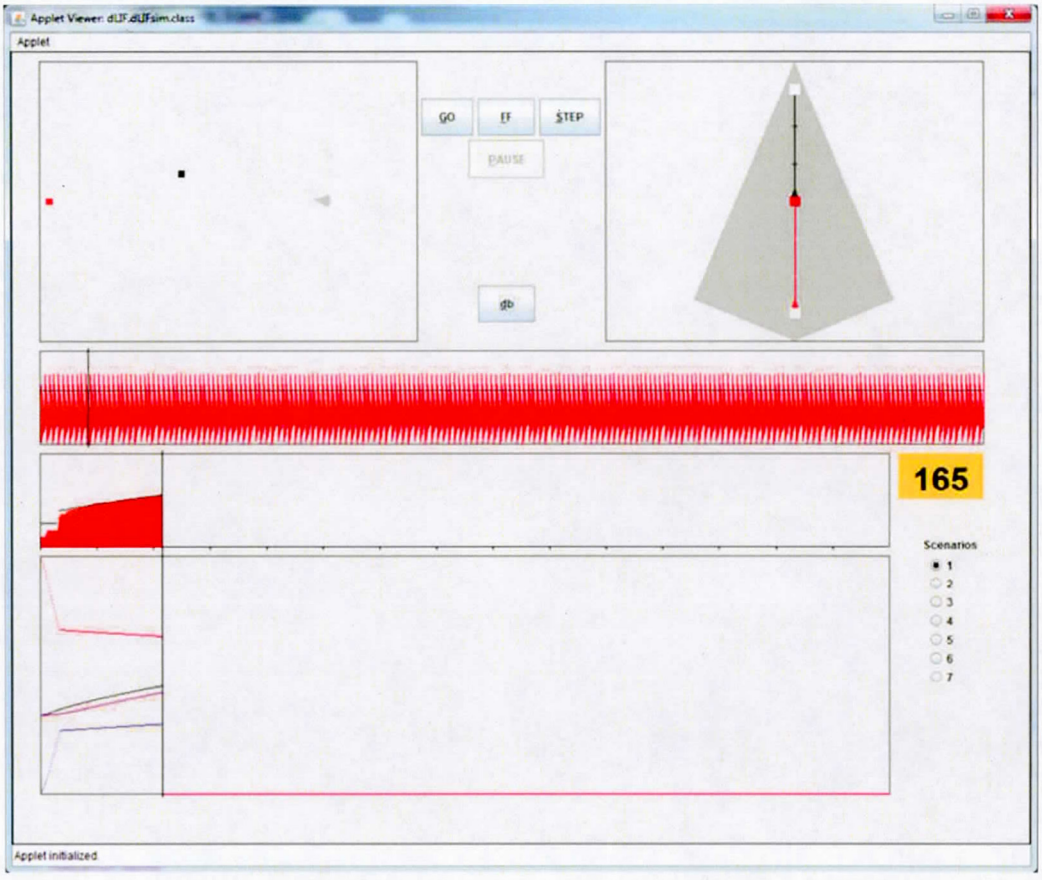


Figure III.1 - The graphical user interface (GUI)

reaction intensity depends on the source intensity, the sensor's position relative to sources and the sensor's development. Sources can be switched on/off by clicking on them.

3.1.1.2 – The brain scan window

The brain scan window provides an image of the agent's body with a gross approximation of the location of sensors and actuators, and a schematic of the interneurons and their interconnections from sensors to actuators (the neuron networks). It allows to select one neuron (and one synapse) for observation, i.e. the

state and parameters of this neuron (and synapse, if one selected) will be shown in the other windows.

3.1.1.3 – The control buttons

There are 5 control buttons:

1. The GO button starts the simulation in real-time.
2. The FF button accelerates the simulation to maximum speed and displays the speed factor when accelerated.
3. The STEP button runs for 1 second and stops 50 ms after the beginning of the following second.
4. The PAUSE button stops the simulation allowing for analysis or changes.
5. The db button is not relevant for demonstrations; it controls debugging messages during software development.

3.1.1.4 – The millisecond window

The millisecond window shows the state (V_c and V_a) for the selected neuron and the modification threshold (θ_M) for the selected synapse. The window uses 1000 pixels to display a full second on a millisecond resolution. The vertical scale, in mV, goes from E_{rest} (-70 mV) to a maximum of +60 mV for a full range of 130 mV.

3.1.1.5 – The spiking rate window

The next window displays the trend of the observed neuron's one-second spiking rate (i.e. the number of output spikes in the preceding second in spikes/second or Hz). The last value is digitally displayed on the right of the window. The window is updated every second and contains 900 seconds or 15 minutes. It is continuously refreshed from the left. The vertical scale goes from 0 to 300 Hz.

This window also shows the selected synapse's modification threshold (θ_M) averaged over 1 second and translated in frequency terms.

3.1.1.6 – The dynamic parameters window

The last window displays, in synchrony with the rate window, the trends of the observed neuron's development parameters:

- C_I starts at 1 pF in the lower left corner of the window and can only go up as the capacitance (the volume) of the selected neuron increases. The vertical scale is logarithmic, ranging from 1 pF (10^{-12} F) to 1 F.
- R_I starts at 3 MΩ in the upper left corner of the window and can only go down as the leak through the membrane increases with the surface increase. For this variable, the scale is linear, ranging from 0 to 3 MΩ.
- N_i, the strengths of all synapses connected to the observed neuron (if one of these is the selected synapse, it is shown in black), start at 1, one third up the scale on the left side of the window, and move up or down depending on the relative stimulation of the pre- and postsynaptic neurons. This variable has no units and the scale is logarithmic ranging from 10^{-1} to 10^{2}.

The scales are provided for analysis purposes; they are not shown on the GUI since, during simulations, we are much more interested by the trends than by absolute values.

3.1.1.7 – The scenario menu

On the right of the dynamic parameters window, there is a list of scenarios which set the environment, the agent and the agent's brain for different experiments. We will discuss these experiments in detail after a closer look at the simulator.

3.1.2 - The simulator

Having started a second thread for the graphical user interface (GUI), the main thread defines the environment, the body, and the brain, and execute a "forever" loop providing a numerical approximation of the differential equations describing the

dynamic model of the neurons combined into the brain, embedded into the body (embodied), and situated in the environment.

```
Start thread for GUI
Initialize agent's body
Initialize environment
  Set position and status of stimulation sources
  Set body's position
Initialize brain according to selected scenario
Adjust stimulation sources' and body's position and status according
            to selected scenario
Start forever loop
  If not running wait for GUI input
    While waiting, check for changes
      If new scenario, reinitialize simulation
      If switch(es) flipped, update environment
      If neuron (and/or synapse) selection changed,
            update brain scan
    If no GO, keep waiting
    On GO from GUI
      Move sources and agent in environment
      If body has moved, update sensor's world position
      Compute effect of sources on sensors according to new
            relative positions
      Stimulate sensors
      Update neurons
      Every 25 ms (40 images/second), update GUI
      If not fastforward, wait to complete reporting period (25 ms)
        Otherwise display speed factor
      If stepping, snot running 50 ms after full second
Repeat forever loop
```

3.1.2.1 - The environment

The environment is a two-dimensional space, providing world reference coordinates, where agents can be stimulated by sources of different types. The effect of a source on an agent depends on the source's type, its intensity and the distance between the agent and the source. If the agent does not have sensors for that source type, the source is not affecting the agent in any way. If the source is of low intensity or very far from the agent, the effect might be below the agent's reaction threshold. Because the effect is inversely proportional to the distance, it is necessary to use a third quasi-dimension to avoid dividing by zero when the agent is located right above or under the source.

3.1.2.2 - The body

The agent's body is positioned in the environment relative to world coordinates. The agent's sensors are located on the body allowing calculation of the sensors' world coordinates when the body moves around and then calculation of sources to sensors distances when agent and/or sources move around in the environment. The agent also has actuators which, when activated, affects the agent's body's world position.

3.1.2.3 - The brain

The brain is (literally) a neuron network connected to sensors (physical properties transducers) as input layer and to actuators (neuronal to action transducers) as output layer, with, in between, a handful of interneurons (very far from the 100 billion in a human brain such that we might have a chance of understanding what is going on).

As previously mentioned in subsection 2.4.5.4 (*Neuronal development and bootstrapping*), all neurons, including sensors and actuators, will have the same values for R_2 and C_2 such that their spiking frequencies will range between 0 and 250 Hz. While this fixes the frequency to cellular voltage response, it does not fix the frequency to current response which depends mainly on the leaky resistance (R_1) or

the inverse of membrane's conductivity. So, our neurons can still assume any possible curves on Figure I.1 (page 13). According to APPENDIX A, DoubleLIF, the oscillator's time constant ($\tau=R_2C_2$) should be 0.0112 to spike at 250 Hz when the cellular potential is 125 mV. If we sC_2 at 1.0 pF, that yields 11.2 GΩ for R_2 (11,214,693,008 Ω to be precise).

We will also assume that each unit of input (each ion channel) has a conductivity of 10^{-6} S; in other words, the resistance to incoming currents is 1 MΩ when the channel is open (relative to infinite when it is close). We should remember that adding resistors in parallel reduces the overall resistance and, as mentioned previously, units do not get full capacity from the beginning such that the number of units (N_i) can take any positive real value.

3.2 - The experimentation

In this section, we will describe a few experimental scenarios designed to understand the basic principles of the model. Since the dynamics are of the essence in understanding the observations, these scenarios are available on the web at www.DoubleLIF.uqam.ca[24].

3.2.1 – Scenario #1: Causality

As shown on Figure III.2, the first scenario presents an extremely simple brain composed of a single straight chain of three neurons: one sensor, one interneuron and one actuator. The scenario is called "Causality" to emphasize the causal effect

---

[24] At the time of publication (end 2014), it is possible to run the simulation from an internet browser at high level of security by adding the site (http://www.DoubleLIF.uqam.ca) to the trusted site list in the JAVA control panel. (For procedural information, see https://www.java.com/en/download/exception_sitelist.jsp). Considering the limited distribution, the application uses a self-signed certificate which might not be tolerated by the next release of JAVA.

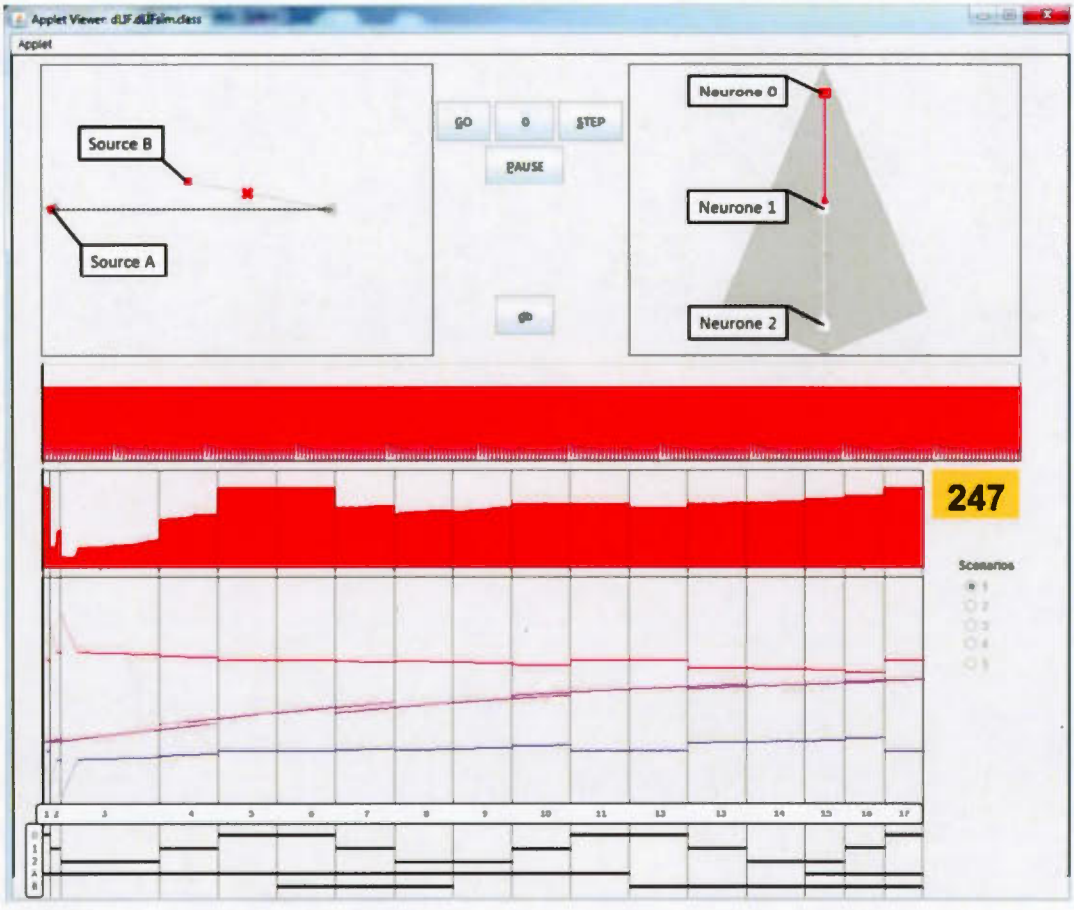


Figure III.2 - Simple neuronal chain

relationship in all neuronal chains. Having established this fact in the simplest possible neuronal arrangement, it should be understood that it applies to all neuronal arrangements however complex they might get.

Each of them have specific properties and we will discuss them starting with the sensor.

3.2.1.1 - The sensor (Neuron 0)

Sensors, like any other neuron, start as highly sensitive cells (the tiniest stimulation generates a full, albeit weak, response) and develop under environmental stimulation

into mature neurons responding only to very specific stimuli with self-adapted scaling for analog transduction. For example, if a light sensor is located at a distance d from a light source A, the sensor receives a stimulus equal to the intensity of A divided by the square of the distance d^2, but, if the sensor has never received such a strong signal ever before, the stimulus is reduced to the maximum previously sensed (0 at the very beginning) plus a development factor (let's say 0.001). The sensor responds at its maximum frequency (in our case: 250 Hz) as long as the received stimulus is equal to or larger than any previously experienced stimulus. If the intensity of the source diminishes, or the distance between the source and the sensor increases, the sensor's response becomes proportional to the ratio of sensed stimulus to maximum previously experienced stimulus; it is therefore automatically limited by the physical constraint on the intensity of such signals in the environment.

3.2.1.2 - The interneuron (Neuron 1)

Interneurons react essentially the same way, except that their input, coming necessarily from another neuron, sensor or interneuron, is pulsed and cannot build up enough potential in C_1, in a single pulse, to energize the oscillator. When the presynaptic neuron is spiking at 250 Hz, as it takes 4 spikes to trigger a postsynaptic spike, the interneuron starts spiking at 62.5 Hz [25] and it takes that much longer to reach internal equilibrium and build connection strength.

3.2.1.3 - The actuator (Neuron 2)

Actuators develop exactly like interneurons. Their reaching full frequency should not mean that they reach the full strength of a mature muscle. The biological development of muscles is not part of neuron development, but it greatly affects brain

---

[25] Without the diode between C_1 and C_2, there would be no spike at all until the internal equilibrium is reached to generate such a spike, but the overall behavior of the neuron would not be significantly different.

development since the dynamics of the response is directly involved. In control systems in general, and in robots in particular, actuators have a given strength, like sensors have a given sensibility, and the interrelation is provided taking these fixed values into account. In a developing brain, it is important to realize that the development of the relationship follows the development of both physical interfaces.

3.2.1.4 - The (causal) sensorimotor chain

With a straight, and short, chain like this one, it is easy to see the causal effect relationships. The physical signal in the environment affects the sensor which reacts and produces neurotransmitters causing channels in the dendritic tree of an interneuron to open letting in electrically charged ions which change the cellular potential of the interneuron triggering a spike along its axon to eject more neurotransmitters towards the actuator which, in turn, reacts and produces an effect in the environment. This is possible only when the agent has the specific type of sensor required to react to that specific signal from the environment. The response is only possible according to the degrees of freedom of the actuator; for this scenario, the agent can only move forward in one direction.

The scenario was organized such that the agent moves towards the light source (*A*), but that should not be interpreted as intentionality, there is nothing more than pure causality. The reaction is similar to that of a bacterium activating its cilium when sensing low food content in its environment. As a result of this activation, it finds itself in a different environment which might, or might not, be richer in nutrients. The move was not triggered by a probability of getting more food, but simply by the fact that there wasn't any food around. This is not different from the action of a thermostat; a bimetallic thermostat on a shelf will click on and off with changes in the ambient temperature even if it does not, in any way, affect the temperature. All sensorimotor activities, however complex, must be explained by such causal chains at the physical level.

However, if we use a different source (e.g. *B*), we see (Figure III.2 in the environment window) that the agent is moving exactly the same way as before not reorienting itself towards the new source (it does not have enough degrees of freedom to follow the path marked with a red X) and continues its way beyond the source to stop, as previously, against the wall.

In the dynamic parameters window, we can see that the parameters are developing continuously. They clearly tend asymptotically towards a stable equilibrium and they preserve their values when the stimulation is stopped. As shown in the extended window at the bottom, the time axis has been divided in 17 sections where the selected neuron (0, 1 or 2) alternate under different stimulating conditions (*A*, *B*, or *A*+*B*). The first three periods show the rapid early growth of the three neurons in sequence with delays and lags in stimulation down the chain. Then, longer periods let see the continuous development of the parameters under sustained stimulation.

Periods 5 and 6, 11 and 12, and 17 show the development of the sensor (neuron 0). In 5, the sensor, stimulated by *A* only, is at maximum intensity. Adding source *B*, in 6, increases the intensity to a new maximum, but not the response frequency which is already maxed out. However, having switched *B* back off, we see, in 11, that the response to *A* only is now less than it was in 5. Switching *B* on and *A* off, the response to *B* only is yet somewhat lower since B is further away (see 12). Finally, when *A* and *B* are both back on, the response goes back to maximum frequency close to 250 Hz as shown by the spiking rate indicator at 247.

The interneuron (neuron 1) and the actuator (neuron 2) undergo similar development, but they haven't reached full maturity in the fifteen minutes covered by the display. After a while, they will also show definite differences in their responses to only *A*, only *B*, and both *A* and *B* being on. The spiking frequency becomes an analog representation of an external physical property. We are not using "representation"

lightly; the neuronal state is causally coupled to the external property. It is a better "representation" than a picture of the source since it follows dynamically any, and all, changes of the property at the sensing point. This "representation" is transmitted down the causal chain producing what Peirce called a semiosis, a chain of signs. However, a dimensionless (punctual) "representation" (previously referred to as sensel) is very limited if not associated with (many) more sensels.

3.2.2 – Scenario #2: Bilaterality

Clearly, a single sensel and a single degree of freedom are not sufficient to talk about cognition. As shown on Figure III.3, in this second scenario, we double the simple causal chain of the previous scenario and we introduce the concept of bilaterality.

3.2.2.1 - The sensors and interneurons

In a bilateral arrangement, the relative physical positions of sensors become of paramount importance. Adding a second sensor (of the same type) immediately provides a different perspective on the environment. The two sensors cannot be exactly at the same position in space and they will generally be excited differently by a single source anywhere in the environment (except in the bisecting plane between the two sensors). The fact that the two sensors are rigidly (or semi-rigidly) interconnected necessarily introduces some correlation between their respective outputs.

Physical reality also constrains the sensitivity angle of all sensors. In scenario 1, the sensor was not constrained and could be stimulated by any source located anywhere 360° around it. With two sensors, one on each side of the body, it is normal to consider that the sources will be effective only when the body is not between the source and the sensor. Considering the quasi-triangular shape of the body, this imply

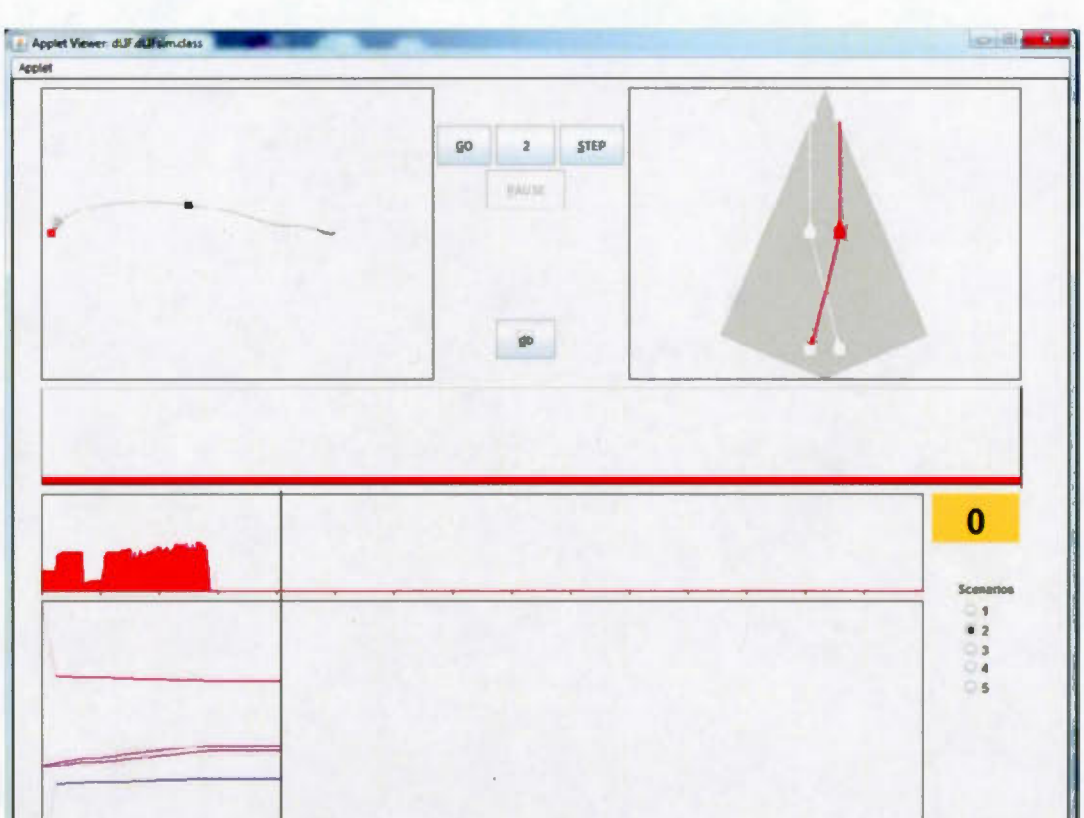


Figure III.3 - Bilaterality

a sensitivity angle of 180° with a small overlap in front where sources simultaneously affect both sensors and a blind spot in the back where neither sensor is affected.

All this means differences; Floridi would say "information". The source is emitting *de re* information and the sensors are affected by causal coupling as discussed in scenario 1. The information is not yet interpreted, but it could be processed. For now, we will not do any processing and the signals will simply be fed, unmodified, to actuators via dedicated interneurons. However, the opportunity exists to extract this potential information by proper interconnection of interneurons and it is precisely the rules of interconnections that we are looking for.

3.2.2.2 - The actuators

Adding a second actuator to our agent provides a second degree of freedom as long as the two actuators are stimulated differently by any common source in the environment. It means that the agent is not limited to moving forward on a straight line, it can cover the entire 2D environment assuming that when an actuator is more excited than the other, that side of the body will advanced faster (one could say: the wheel on that side will turn faster). Of course, the positioning of the actuators is as critical as the positioning of the sensors. This is where genetics plays an essential role in cognition. If you do not have light sensors, you cannot see. If you do not have legs or wheels or whatever mechanisms, you cannot move. If you cannot store energy, you cannot generate actions. Genetics sets the landscape (Waddington 1956) and cognition is "*canalized*" by the available set of sensors and actuators. On that basis, we try to identify how the interconnections between sensors and actuators could, solely on stimulation (and constraints) from the body and the environment, develop a network identifiable as a cognitive architecture.

Since we need a starting point for our observations, we selected to cross the median plane when connecting the output of the interneurons to the actuators. We could have chosen a different starting point, but our objective is to verify that the simple observable sensorimotor behaviors resulting from a given neuronal connection are predictable according to the model. In this scenario, we can predict that:

1. under constant stimulation, the connections will strengthen,
2. the agent will be attracted by the source, and
3. whenever all sources are in the agent's blind spots, the agent will be (fatally) immobilized.

Experimenting with various starting positions and orientations, we observed that the agent effectively turned towards the source at first. If the agent was positioned facing the source such that both sensors were stimulated, it would then move forward slowly

curving until, passing under the source, this one ends up in the blind spot. Of course, whenever the agent was positioned such that the source was already in its blind spot (e.g. facing in the opposite direction), it would not move at all. When the source was in the sensitive angle of one sensor but not of the other, the agent would turn pass the direction of the source until this one ends up in the blind spot of the first exposed sensor, to catch up the strengthening of the second chain until the actuators receive approximately equal stimulations when both sensors are affected by the source moving then the agent forward curving (and wobbling) until, passing under the source, this one ends up in the blind spot.

If two sources are used instead of one, the agent continues to move from one source to the other according to the relative attraction (intensity/distance) of the sources until both sources are simultaneously in the blind spot. Multiplying the number of sources, it is possible to keep the agent going forever.

After some time with a given configuration of sources, the agent will repeatedly follow a pattern around the scene. Turning sources on and off will modify the pattern; the agent (re)learning a new pattern after each modification. Figure III.3 shows the path followed by an agent stimulated only by the leftmost source (*A*) until it reached that source and stops in its blind spot after close to 3 minutes. In Figure III.4, we see that it starts again a minute or so later when the second source (*B*) is turned on. The agent turns then towards source *B* and, subject to the competing attractions of both sources, describes a series of elliptical convolutions until it settles in a circular pattern in between the sources, close to *B*. We can see that the interneuron's dynamic parameters (continuously observed from beginning to end) keep changing throughout this experiment.

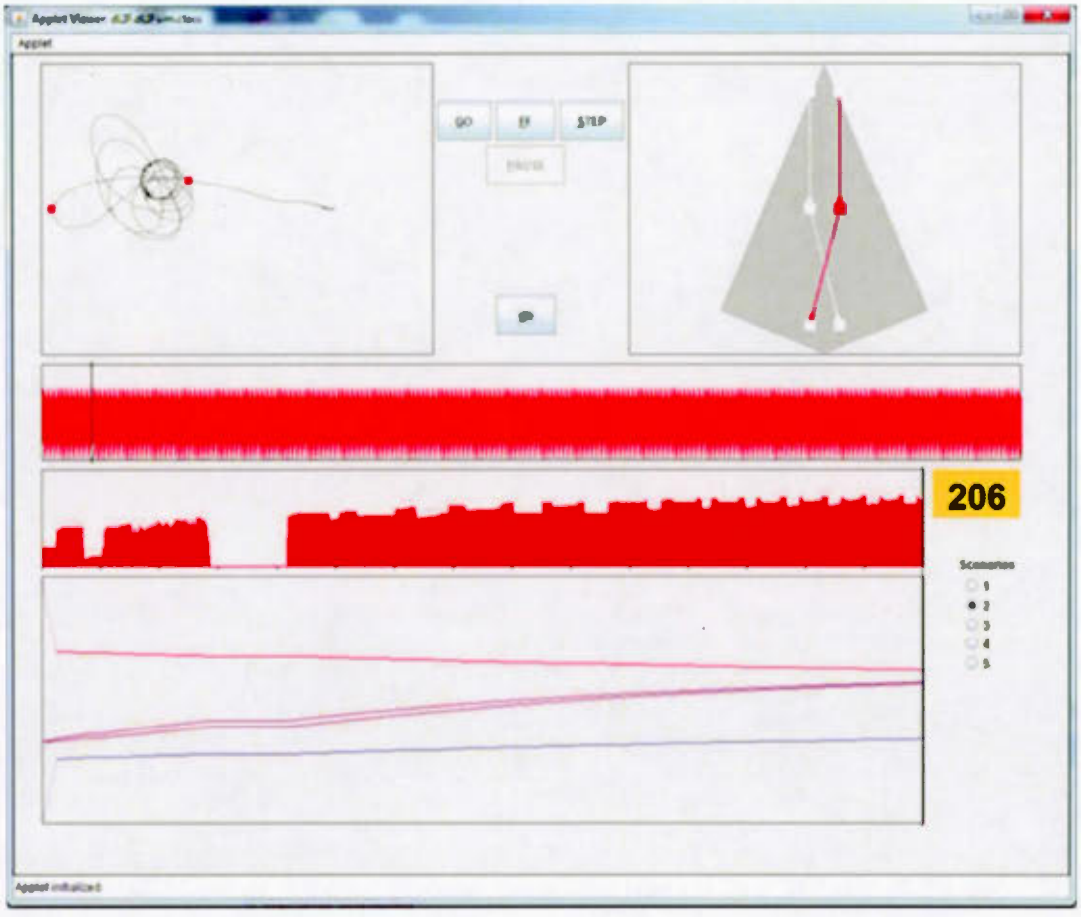


Figure III.4 - Two sources (attractors)

3.2.3 – Scenario #3: Random walk

In the two previous scenarios, we have seen that any sensor stimulated by an appropriate source will necessarily respond proportionally to the intensity of the stimulus. Similarly, any interneuron and any actuator will necessarily respond proportionally to the combined intensity and frequency of all stimuli from interneurons or sensors synaptically connected to their dendritic tree. On the other hand, this implies that whenever the causal coupling between the source and the sensor is broken, the activation of the sensor disappears as well as any synaptically

transmitted causal stimulation. No stimulation without response; no response without stimulation.

In our previous examples, the agent often suffered fatal lack of stimulation due to the extremely limited number of sensors and the fixed configuration of sources. In the real world, stimulations continuously appear and disappear from the agent's immediate environment. Furthermore, the agent is also equipped with a multitude of internal sensors and actuators continuously activated by its own metabolism.

In the present scenario, we can see that, if a source is randomly moving in the environment, the agent will keep following it around. There might be periods of time where the source will be stuck in the agent's blind spot, but it will eventually move in such a way that the stimulation will resume. If we multiply such sources, the agent becomes incessantly stimulated. If the movement of the sources is not totally random, there might be some patterns for the agent to learn.

The essentially dynamic aspects of this scenario cannot be presented in a snapshot of the GUI; it can only be fully appreciated in a real (or accelerated) time display of the behavior. However, Figure III.5 shows a few neuronal spiking patterns which are typical of such lively stimulations when the agent, stimulated by two sources, follows a randomly moving source while the other source remained at its starting position. The agent's path, in black, allows us to imagine the random walk of one source (starting where source *A* normally stood in previous scenarios) moving up, left and diagonally down close to the second source (in position B in previous scenarios) at the time of the screenshot. Clearly, this experiment could go on forever with its dull moments when the two sources are in the agent's blind spot, but always revived when the moving source randomly gets out of that blind spot.

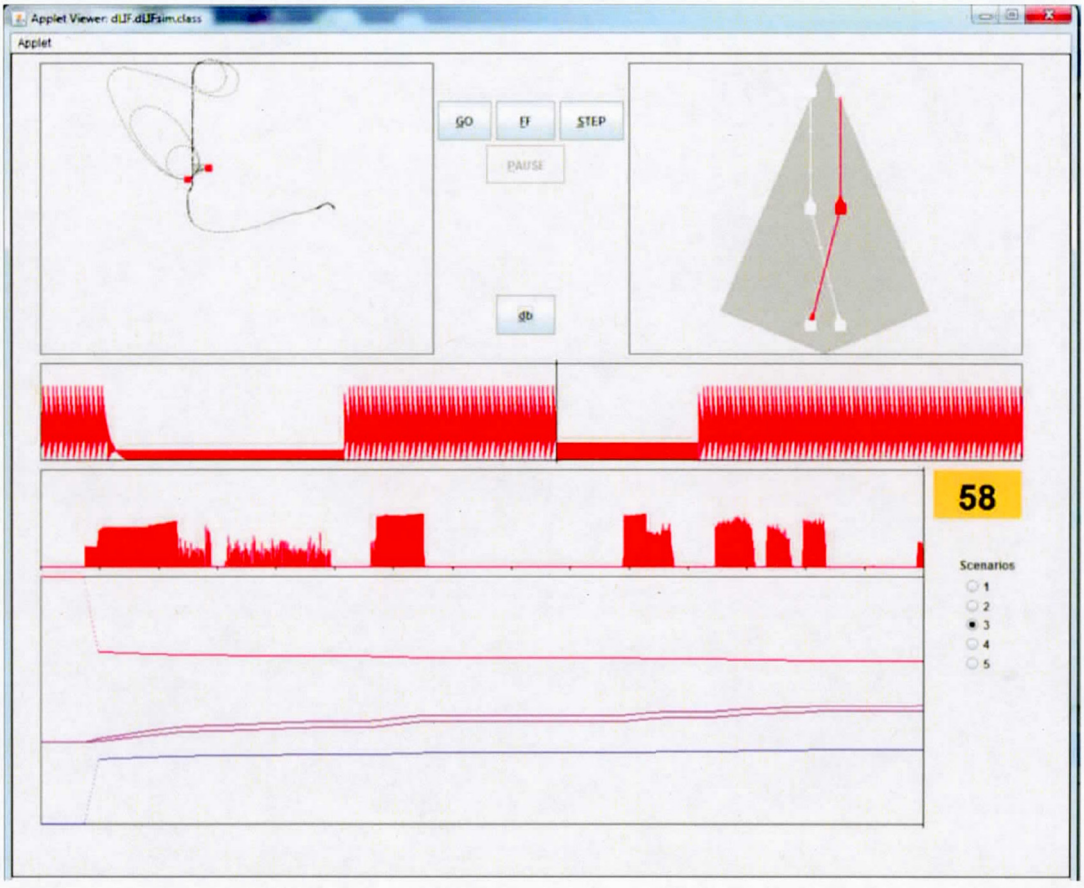


Figure III.5 - Random walk

All this is reminiscent of Braitenberg's *Vehicles* (1984) except that electrical wires have been replaced by chains of artificial neurons.

3.2.4 – Scenario #4: Inhibition

Thus far, we have dealt only with excitatory neurons. We will now introduce inhibitory neurons which send different neurotransmitters through the synaptic gap thereby activating K^+ or Cl^- channels instead of Na^+ channels in the postsynaptic neurons. The opening of these channels produces IPSC's instead of EPSC's (Inhibitory instead of Excitatory PostSynaptic Currents) driving the postsynaptic

potential down instead of up. At first glance, the effect seems simple enough: an inhibitory signal can neutralize the effect of an equivalent excitatory signal after proper weighing of both signals by their relative synaptic strength. But how do they connect together?

As a postsynaptic neuron, an inhibitory neuron develops exactly the same way as an excitatory neuron; its dendritic tree is not different and reacts identically to stimulations. As a presynaptic neuron, things are different: it sends a negative message. While homosynaptic stimulation from an excitatory neuron provides positive feedback (i.e. stimulation raises postsynaptic voltage which favors LTP which increases stimulation which again favors LTP and so on), homosynaptic stimulation from an inhibitory neuron produces negative feedback (i.e. stimulation reduces postsynaptic voltage which favors LTD which decreases stimulation and therefore dampens the response instead of amplifying it. On the other hand, inhibitory connections thrive from competition with heterosynaptic excitatory connections. The excitatory connections raise the postsynaptic voltage which favors LTP for all connections, inhibitory as well as excitatory, which increases stimulation with mitigated results considering the canceling competitive effects. So, inhibitory connections can only strengthen when they fire into an already positively stimulated postsynaptic neuron. They follow Hebb's law in that they wire with other neurons firing at unison, but not to activate them further, rather to stop them from firing.

However, experimenting with this assumption, we found out that:

1. the inhibitory connections can never fully catch up with competing excitatory connections because these excitatory connections strengthen as a result of their own action whether the inhibitory connections are interfering or not while, on the other hand, the inhibitory connections cannot strengthen themselves in the absence of excitatory stimulations,

2. on the contrary, iterative homosynaptic or "cooperative" inhibitory stimulations, without any excitatory counterpart, can only drop the postsynaptic potential and weaken any active incoming signal,
3. by repeatedly stimulating an already polarized postsynaptic neuron, inhibitory connections tend to eliminate themselves, and
4. although it was possible to get the desired behavior by tweaking the external stimuli, the equilibrium was unstable and, under specific changes in the stimulation, the neuron would suddenly favor the rapid strengthening of the excitatory synapse rendering the inhibitory one totally ineffective.

Having exhausted all imaginable alternatives with the original set of equations (not to mention the time allowed for the project), it became necessary to look at any other mechanism potentially offering a stable solution to the inhibition problem. Our observations of the model led us to believe that the inhibitory stimulation had to act directly on the excitatory synapses converging on the same postsynaptic neuron and not only on the cellular potential. However, such an action would have been a violation of our encapsulation principle. The inhibitory synapses do not know, when activated, if there are excitatory synapses connected to that same postsynaptic neuron, even less which synapses these could be. We therefore postulated the existence of an unknown mechanism involving the presence, in the postsynaptic neuron, of a messenger (inhibitor) generated by, and proportional to, any inhibitory stimulation and neutralizing proportionally any future excitatory stimulation before exponentially decaying out of the system. It was also assumed that the neutralized excitatory stimulation could not participate in any form of synaptic potentiation or depression.

$$dM_{inh}(t)/spike = -k_{inh} * I_{IPSP} \quad \text{for inhibitory spikes and}$$
$$= -k_{decay} * M_{inh}(t) \quad \text{for excitatory spikes.}$$

and I_{EPSP} in equation 1 is reduced by an amount equal to the residual M_{inh} before affecting the synaptic strength. The numerical approximation method used to process

synaptic plasticity on a per spike basis might be, at least partly, responsible for the problem, but detailed investigation of this method is not feasible at this stage in the project considering time and complexity.

Figure III.6 shows an agent with two sensors of different types responding to two sources (*A* and *B*) of corresponding types. The sensor on the right side of the body (neuron 0) is stimulated by source *A*, but not by source *B*, while the sensor on the left side of the body (neuron 1) is stimulated by source *B*, but not by source *A*. The signal from the left sensor is sent to an inhibitory interneuron (neuron 2) which produces an inhibitory signal whenever the left sensor is active which means whenever source *B* is on. If we send this inhibitory signal to another interneuron (neuron 3) jointly with the excitatory signal from the right sensor, we could expect some kind of competition between the two signals. When source *A* is on and source *B* is off, neuron 3 receives an excitatory signal from neuron 0 and develop normally as we have seen in previous scenarios since there would not be any (inhibitory) signal coming from neuron 2. When source *A* is off and source *B* is on, neuron 3 receives an inhibitory signal from neuron 2 and nothing happens since neuron 3 is already fully depolarized and the inhibitory signal can only try to depolarize it further. When both sources *A* and *B* are on, neuron 3 receives both an excitatory signal from neuron 0 and an inhibitory signal from neuron 2. With the inhibitory messenger, the inhibitory signal always wins and neuron 3 is inactivated or at least below spiking threshold.

Table III.1 summarizes the resulting responses. As shown, the response of neuron 3 corresponds to an A-not-B gate ($A \wedge \neg B$).

Figure III.6, also shows how this response develops over time under external stimulation by flipping sources *A* and *B* on and off in pseudo-random sequences. With time, it becomes as necessary that *B* be off for neuron 3 to be on as it is necessary that *A* be on. In other words, the absence of *B* is significant for our agent.

Table III.1 - Neuronal A-not-B gate

Source A	ON	OFF	ON	OFF	
Source B	OFF	ON	ON	OFF	
Neuron 0	1	0	1	0	A
Neuron 1	0	1	1	0	B
Neuron 2	0	-1	-1	0	-B
Neuron 3	1	0	0	0	$A \wedge \neg B$

The bottom portion of Figure III.6 is composed of four spiking trains showing how each neuron (0, 1, 2 and 3) responds to different combinations of sources' activation.

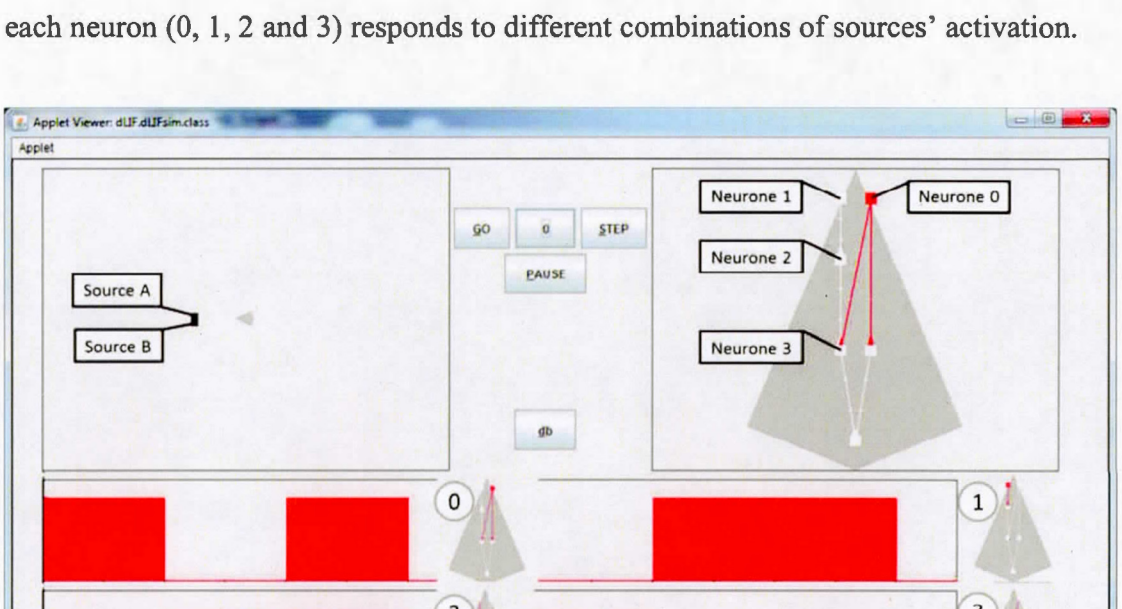


Figure III.6 - Inhibition

To complete the network, we needed an actuator (not identified, but the only neuron in the output layer, i.e. without an output synaptic link). In order to avoid the additional complexity of having to follow a moving agent while flipping switches, we fed strictly inhibitory signals to this actuator such that it was never excited. The other unidentified neuron, on the right of neuron 3, is an inhibitory neuron transforming the excitatory signal from neuron 0 into an inhibitory signal before sending it to the actuator.

At time of submitting this PhD dissertation, this inhibition mechanism seems to be the best solution to our stability problem with the A-not-B gate. However, some doubt persists that it might well be an indirect fix to a quirk in our numerical integration method where we process stimulation and synaptic plasticity on a per spike basis within integration periods.

3.2.5 – Scenario #5: Neuronal logic

Having established a stable mechanism for the combination of inhibitory and excitatory signals, we can now entertain more complex networks.

Figure III.7 shows a network of 12 neurons, some excitatory, some inhibitory, which performs the basics of neuronal logic.

As shown in Table III.2, we expect each neuron to perform a specific logical operation.

Neuron 0 is a sensor responding to source *A*. Whenever neuron 0 is physically linked to source *A*, it produces a signal representing, in the network, the presence (the existence) of *A* in the environment. As long as *A* has not been activated in the environment, neuron 0 remains dormant in the network. When *A* is turned on for the first time, neuron 0 is excited and begins to develop. Any repetition of *A*'s activation,

will strengthen neuron 0 and improve its capacity to represent different levels of intensity of stimulation from source *A*. This part of the logic is clearly inductive since the representation of *A* is based solely on the repetitive stimulations by *A*. That representation of *A* is the ontological establishment of *A* in the representational structure. Before the first stimulation, there is no *A* in the structure; with repetitions, the existence of *A* becomes more and more ascertained and the neuron becomes more and more dedicated to representing *A*. The same applies to interneurons and actuators in their representation resulting from composition of representations fed to them.

Neuron 1 is a sensor responding to source *B*. Everything we said about neuron 0, with respect to source *A*, applies to neuron 1, with respect to source *B*.

Neuron 2 is an inhibitory neuron transforming an excitatory signal from neuron 0 into an inhibitory signal. In itself, *-A* is not different from *A* from a representational point of view, but, when combined with other signals, the effect is equivalent to $\neg A$, the complement of *A*, as we have seen in scenario 4.

Neuron 3 is to neuron 1 what neuron 2 is to neuron 0.

Neuron 4 and neuron 6 replicate scenario 4 respectively producing $(B \wedge \neg A)$ and $(A \wedge \neg B)$.

Neuron 5 combines excitatory signals from neuron 0 and neuron 1. The development of this neuron is also based on induction and its output signal becomes proportional to the exposure of neuron 0 to source *A* relative to the exposure of neuron 1 to source *B*.

Neuron 7, combining the results of neurons 4 and 6, produces an exclusive-or-gate behavior resulting from $((A \wedge \neg B) \vee (B \wedge \neg A))$.

Table III.2 - Neuronal logic

Source A	ON	OFF	ON	OFF	
Source B	OFF	ON	ON	OFF	
Neuron 0	1	0	1	0	A
Neuron 1	0	1	1	0	B
Neuron 2	-1	0	-1	0	-A
Neuron 3	0	-1	-1	0	-B
Neuron 4	0	1	0	0	$B \wedge \neg A$
Neuron 5	1	1	1	0	$A \vee B$
Neuron 6	1	0	0	0	$A \wedge \neg B$
Neuron 7	1	1	0	0	$A \veebar B$
Neuron 8	-1	-1	0	0	$-(A \veebar B)$
Neuron 9	0	0	1	0	$A \wedge B$

Neuron 8 is an inhibitory neuron producing the complement of neuron 7.

Neuron 9, combining the results of neurons 5 and 8, produces an and-gate behavior resulting from $((A \vee B) \wedge \neg(A \veebar B))$.

Figure III.7 shows the responses of the 10 neurons to the different combinations of sources' activation. Like in scenario 4, an actuator and two inhibitory neurons (all three unidentified) were added to the network to keep the agent immobile.

Figure III.7 - Neuronal logic

This agent, with its 12-neuron brain, has a complete internal representation of all the potential states of its admittedly over simplistic environment. The state of the world[26] cannot change without the agent's neuronal state changing and the agent's neuronal

---

[26] We should remember that we are referring to an oversimplified "world". In a more complex setup, an agent's brain receives as much stimulations from its own body than from the rest of the world. So, in this sentence, the state of the world includes the state of the agent's body and, ultimately, the state of the agent's brain.

state does not change if the state of the world has not. In other words, the agent has its own representational system, its own cognitive system. This system is not symbolic. We call the sources *A* and *B*, but the agent does not use, nor need, any labels; it simply reacts to their stimulation. The system is semiotic since a signal is transmitted (causally) from the source to, and through, the network.

As mentioned briefly when inhibition was introduced in scenario 4, once the sources have been detected by the agent, their absence becomes as significant as their presence. This phenomenon of stimulation *in absentia* has been studied at length by Deacon in *Incomplete Nature* (2012).

3.2.6 – Future scenarios

Having established that:

1. neuronal chains can transport signals as effectively as Braitenberg's copper wires (1984), and
2. neurons can compose signals according to a well-defined logic in the same way that logic gates process information in von Neumann machines,

it is now possible to design networks to improve the agent's behavior by multiplying the number of sensors, the number and complexity of actuators, and the number and intricacies of interneurons.

A matrix of sensors, like the retina, provides redundancy and distribution of information. Each sensor provides unique, but correlated, information such that spatial differentiation of such information is in itself information. For example, speed and acceleration are first and second differentiations of position. What information can we generate by differentiating acceleration? What kind of network do we need to detect objects from sensel information?

More complex actuators require more sophisticated logic. One can easily imagine a quad-input two-wheeler where the wheels could turn forward or backward. It would be interesting to evaluate a network where the forward and backward signals would be interlocked somewhat like antagonist muscles, flexors and extensors, in the body.

It is expected that designing different kinds of such modules would allow the investigation of potential rules in the development of networks.

3.3 - Discussion

These results must be analyzed and understood in the transdisciplinary context of the research. The main question guiding the project has always been: "Is strong AI still possible?" With all that has been written about computers, cognition and neurology, is it still realistic to think that a non-biological machine could, one day, pretend to be a true cognitive system, an understanding machine?

3.3.1 – Emulation vs simulation

The scenarios described in the preceding pages of this chapter depict the results of a computer simulation based on a simplified physical environment where a simplistic agent reacts to environmental stimulations through simulated physical interconnections between its sensors and its actuators. We fully concede that the environment simulation is oversimplified and extremely limited, but this might be an advantage when the time comes to interpret what is going on. Thanks to the limited number of stimulation sources, we need only an equally limited number of sensors. These sensors are also a simulation of physical sensors translating a physical property (light, sound, odor, taste, etc.) into a neuronal signal. These neuronal signals are processed (composed) to produce a physical action from the translation, by actuators, of resulting neuronal signals. At this level, we are talking about a computer

simulation of a behavioral phenomenon where a physical stimulation S calls for a physical response R.

However, if we examine these simulations more closely, they are not identical in nature. The physical environment can never be replaced by its simulation. The same applies to the sensors and to the relationships between the environment and the sensors. Ditto for actuators which are, so to speak, inverted sensors. When we look at interneurons, the story is different; how the signals generated by the sensors reach the actuators is totally irrelevant, as long as, for the same combination of stimulations, the same actions are produced. Something like replacing Dretske's doorbell by a modern electronic version; no more wires between the button and the bell, a radio wave carries the signal. Somebody pushes the button and the bell still goes "ding dong". Same stimulation, same action; even though the physical carrier is completely different. In the case of the doorbell, we could call it a physical emulation. If you have more than one door, hence more than one button (sensor), some coding is required to distinguish sensors and produce appropriately different actions ("ding", "dong", "ding dong" or even the Westminster chimes...) and this becomes a numerical emulation. Note that many years ago, typewriters were using strictly mechanical keyboards; then came IBM Selectrics: an electric reproduction (emulation) of those mechanical keyboards; and nowadays, you probably have a totally numerical (not to say virtual) emulator on your tablet.

Requirement 1 – For strong AI to be possible, the brain, a physical causal system, must be emulated.

The emulation of a causal system can only be done at the level of its simplest component. In the case of the brain, a biological system, this means the cellular level, the neurons. The objective of "artificial" intelligence is to replace this biological component by a non-biological one without losing its cognitive function that is any of

the properties essential to cognitive processes. The main difficulty becomes to define which properties are essentially cognitive and which ones are strictly biological. How can we preserve the former while discarding (replacing) the latter? This is where we introduced the concept of equilibrium between cognitive necessity and biological plausibility.

3.3.2 – Cognitive necessity

Eliasmith (2013 p20) points out: « [There] is [something] identified nearly universally as a hallmark of cognitive systems: the ability to manipulate structured representations. » All cognitive functions require the existence of a representational structure; you may call it conceptual structure, semantic structure, informational system, symbolic system, or, as we did in this document, semiotic system. A mirror, for example, is a simple physical representational system (not the same as a mental representation structure, but we will try to get there). A mirror could be emulated (replaced) by a camera and a video screen (with some processing to flip the image horizontally). This provides some additional insights on what we mean by emulation of a physical system at the information level. The image provided by a closed-loop video circuit is as good (provides the same information) as a mirror; granularity might not be perfect, but an AI system as close to some cognitive system as an HD TV is to a mirror, would be a major achievement. Not that an HD TV is comparable to any kind of cognitive systems. For one, cognitive systems are multimodal; we would have to add at least sound, but the main difference is not there. Cognitive systems do not strictly reproduce the inputs in kind, assuming that HD TV can reproduce visual and auditory signals in kind (i.e. the signals produced by the TV are similar to the signals captured by the camera), but process the input signals and transform them in an action which is as much a representation of these inputs than the image in the mirror is a representation of the object in front of it.

Requirement 2 – For a cognitive system, the emulation is not a simple one-to-one reproduction of each input into an output, but a rather complex composition of many inputs into each single output. The dynamic sum of these outputs produces a behavior.

In scenarios 1 and 2, we can see the basic elements of such emulations. The simple 3-neuron chains act like wires with transducers at both ends, in a causally behavioral fashion. In Braitenberg's vehicles, such connections from input to output are indeed hard-wired with fixed transducing factors at both ends such that the output is proportional to the input. This arrangement is equivalent to the closed-loop video circuit; it guarantees analog dynamic representation of the input signals at the output and partly meets requirement 1. However, the brain is not that rigid; on the contrary it is known to be highly flexible, highly plastic. In other words, the hard-wired fixed-coefficient transducers are not biologically plausible.

3.3.3 – Biological plausibility

Scenarios 1 and 2 also show that, the 3-neuron chains are not as rigidly wired as the Braitenberg's vehicles. First of all, the sensors have self-tuning transducing factors. The fixed span (0 to 250 Hz) of the sensors output is slowly brought to represent the maximum intensity ever sensed. At the beginning, even an infinitesimal stimulation produces a full output (250 Hz). This full output correspond to a highly depolarized cell (~30 mv) which is assumed to trigger some metabolic reactions modifying the cell's parameters (capacitance and conductivity) until, after a while, the ionic leak is in equilibrium with the input spikes. We are not looking for the specific (biological) reactions since, anyway, they will be replaced by an equivalent algorithm, but we try to identify metabolic processes which are not blatantly biologically impossible. In this case, a specific state of the cell is associated with a specific change in the cell's characteristics. Encapsulation is certainly a strong constraint on biological

plausibility; cells do not know anything about what is happening in other cells, they barely react to disturbances imposed on them by their immediate environment.

Requirement 3 – Encapsulation is a necessary condition for biological plausibility.

A similar bootstrapping process applies for interneurons and actuators. In a robotic implementation, the physical transducers necessarily have a fixed span which must be corrected to emulate the bootstrapping of biological neural cells. For sensors, the fixed span is first applied to some analog to numerical transducer and the numerical signal is then used as the value of the external stimulation to bootstrap an internal transduction factor. The process is somewhat reversed for actuators such that only a fraction of the output signal is sent out, delaying the usage of full strength for a fairly long period of time. These artificial filters can only be defined empirically depending of the actuators involved and the type of behaviors being emulated. This is an attempt to compensate for muscle development, but it does not help with the additional problem of growing bones. Body development is not a cognition problem, but it is clearly a significant variable in cognition development.

This bootstrapping plasticity is not usually included in neuronal learning processes because it is very difficult to observe in vitro as well as in vivo. Most experiments involve neurons at a given stage of their development with relatively constant characteristics. STDP experiments, for example, observe marginal development of synapses assuming that other parts of the neurons do not change (ceteris paribus), but is that really the case?

The study of synaptic plasticity displaces the focus from one neuron to the interface between two (pre- and postsynaptic) neurons. Encapsulation has to be redefined. At the cellular level, it is easy to refer to the cell's membrane as the encapsulation boundary. At the synaptic level, we find ourselves at the interface between two

boundaries. Encapsulation becomes the operational closure of the synapse. To produce the appropriate effect on the postsynaptic neuron and on its self-strengthening, all the synapse needs to know is: 1) is there a pulse in the presynaptic axon? and 2) what is the cellular potential of the postsynaptic neuron? From a biological point of view, the mechanisms are much more complex than that including multiple reactions to produce neurotransmitters, to opens ligand-gated channels, to produce additional channels for LTP (or somehow eliminate channels for LTD), to enhance (or curtail) neurotransmitter production processes to maintain equilibrium in the future. From a cognition point of view, it is sufficient to know the effect of a pulse on the state of the postsynaptic neuron and on the strength of the synapse itself; we can assume that, when functioning properly, the biological reactions will reach expected homeostatic equilibria. Still, encapsulation, even redefined in this way, puts strict constraints on what can (plausibly) be achieved by neurons and parts thereof.

It should be noted that synaptic plasticity (strengthening of the connections) can push the cellular potential beyond its prescribed limit and trigger the bootstrapping process modifying the cell's properties thereby affecting future expression of synaptic plasticity. This effect of plasticity on plasticity is referred to as metaplasticity.

As we have seen until now, there is no decision to make, not even the possibility to make decisions, about what a sensor is a representation of. A sensor responds to a given type of stimulation and represents this stimulation as sensed at a specific point in space-time. One cannot decide what a point on a mirror (or a pixel on a TV screen) will reflect, one can only put different objects in front of the mirror and the structure of the representation is determined by the structure of the object. In a similar way, the output of an interneuron is the sum of the representations of its inputs. These representations are built by associations of repetitive stimulations. The representational structure develops from the regularities in the stimulations. At that

level, we are not talking about symbols yet, barely about signs combining into concepts, without labels to assign to these concepts.

Requirement 4 – The representational structure develops autopoietically from repetitive stimulations and associations of repetitive stimulations.

We consider this development autopoietic because there is no way to force a sensor to react differently to a stimulation than its natural way of reacting. And a group of sensors, as a group, will always have correlated responses to stimulations and these correlations will necessarily bring associations of signals into neurons which could be identified as representations (or even concepts) of the object responsible for the repetitive common stimulation.

3.3.4 – Neuronal logic

Scenario 1 tries to illustrate the causality between the response and the stimulation, but also the causality behind the development of individual neurons.

Scenario 2 introduces, with bilaterality, a new type of physical categorization. In scenario 1, physical signals were segregated by the type of sensors only. In 2, the bodily arrangement of sensors creates new ways of making differences, of discovering information. The sources are sensed differently by the two sensors on both sides of the body. If a source is in front of the agent, both sensors are stimulated. At some point, when the source moves far enough to one side of the body, only one of the two sensors is stimulated. There is a region where none of the sensors are stimulated: the blind spot. The information is divided in 4 categories: left, right, in front and behind which could also be labelled "A", "B", "A and B", and "Nothing". Excitatory neurons can easily mark the difference between "A" and "B" since different neuronal signals are active in these cases. "A" and "B" can be associated to

represent "A and B", but the resulting signal is in fact a representation of "A or B" with differences in intensity when only "A" or "B" are present. There is no way to isolate the cases where the source is affecting both sensors simultaneously. Scenario 3 shows that it is possible to generate somewhat realistic behaviors using this incomplete logic and maybe even create the illusion that the agent is intentionally following a randomly moving source or deciding to go to the closest source when more than one is available.

There is still one tool which we have not used and which is well recognized by neurologists: the excitatory neuron. However, its application is not straightforward; it certainly does not give "A and B" directly. At best, it gives us a different version of "A" and "B"; a negative version which could be called "A-" and "B-" since, when connected with an excitatory signal to a common postsynaptic neuron, the signals compete instead of being additive such that "A" and "B-" tend to annihilate each other. The result is that an "A and B-" combination will produce a signal when "B" is not there and no signal when "B" is there; in other words a typical "A and notB" combination as demonstrated in scenario 4. Although "B-" is far from being equivalent to "notB", "A and B-" is equivalent to "A and notB".

Requirement 5 – Excitatory neurons are not sufficient to take full advantage of the available information. Inhibitory neurons are required to complete the isolation of overlapping categories.

Scenario 5 shows that, with "A", "B", "A-" and "B-", it is possible to generate neuronal combinations including "A and notB", "B and notA", "A or B", "A and B", and "A xor B". These are sufficient to develop propositional logic. The presence of "A" and "B" signals is in itself an existential quantifier.

3.3.5 – Informational autopoiesis

The original intent was to empirically investigate the composition of excitatory and inhibitory signals to evaluate selectivity rules (like BCM) and, maybe, discover new rules to sustain informational autopoiesis. The difficulties of implementing stable inhibition made us realize that composition rules, at that level, were required to empirically build the networks we intended to observe.

Requirement 6 – A basic set of rules is required to understand the generation of a semiotic system before we can investigate its autopoietic behavior.

Having added equations to the inhibition part of the model, we now have a framework where it is possible to construct networks on the basis of logical gates and investigate how this construction process could be algorithmized.

CHAPTER IV

Synthetic neuro-cognition

At the beginning of this project, embarking on a quest for Artificial Intelligence, we voluntarily limited the scope to preverbal preconscious intelligence, hoping to avoid the hardest problems of cognition. A preliminary review, unavoidably too restricted, of some cognitive science main streams led us to conclude that intelligent, or should we say cognitive, systems had to be semiotic and autopoietic: two necessary and jointly sufficient conditions to avoid the well-known problems of symbol grounding and zero semantic commitment. Strongly biased by a physicalist (read this as an extrapolation of engineering) background, we favored a connectionist postulate ("*The brain can be algorithmized*") over a computationalist (cognitivist, representationnalist) hypothesis ("*The mind can be algorithmized*"). This postulate tacitly implied the development of artificial neural networks which, we thought, had to be as close as possible to biological neurons in order to have a chance to meet the autopoietic condition.

We then retraced the history of neuron models development and highlighted, again, two different approaches. The first, probably influenced by computationalists (functionalists), was primarily interested in reproducing the logical functions of the mind with oversimplified neuron models. The second, constrained by biological plausibility, focused on the dynamic aspects of spiking neurons. Again, conscious of the biological aspects of autopoiesis, we felt obligated to join the latter.

This led us to the development of an autopoietic semiotic artificial neuron capable of classical binary logic and expandable to fuzzy (analog) logic. Considering the similarities between our agent's behaviors and Braitenberg's vehicles' (1984), we found appropriate to refer to this framework as synthetic neuro-cognition referring to Braitenberg's book subtitle: « Experiments in Synthetic Psychology ». Having elected to stay at the subconscious level, we could hardly talk of psychology, which covers more than intelligence which, in turn, already implies advanced cognition. Talking of neuro-cognition stresses the importance of neurons for the bottom-up development of intelligence. Like Braitenberg, we want to emphasize that cognition, hence intelligence and psychology, develops synthetically by composition of atomic elements (in our case: neurons), hinting by the way that synthesis can well be artificially reproduced. Briefly, we propose synthetic neuro-cognition as the emergence of artificial intelligence.

4.1 - Achievements

What Braitenberg (1984) could do with electrical wires, we can now do with neuronal chains. What von Neumann could do with logic gates, we can now do with neuronal assemblies. While electrical wires and transistors (hence logic gates) have static response curves, DoubleLIF neurons adapt to, and are modified by, the processed information. While the neurons are artificial, the network can self-organize autopoietically. Our objective has not changed, we do not intend to duplicate von Neumann's machines with neuronal logic gates; we are still interested in finding out how such gates can self-organize under external stimulation.

4.1.1 - The simulation

We have a framework allowing us to try different neuronal configurations and investigate how they could self-develop under the influence of external information.

Manipulating these neuronal configurations, it is possible to generate different representational structures and ultimately extract the rules guiding the development of cognitive architectures.

The prototype was developed to demonstrate the validity of the mathematical model. Sets of differential equations cannot be analytically solved and can only be evaluated through numerical approximation. Although it might be only an approximation of the biological elements of cognition, the implemented model could well be a very satisfactory emulation of the basic cognitive functionality.

As developed, the simulation includes more than a model of the brain. The body (sensors and actuators) and the environment (stimulating sources) must also be handled via some physical engine which often ends up being more complex and resources demanding than the brain itself. In reality, including in robotics, the effects of the environment on sensors do not require any calculation, they simply happen. The same can be said about the effect of actuators on the environment. So, the complexity of body and environment is not a significant consideration in our case. It has been kept to a minimum, to the point of being over simplistic, in order to make programming and result analysis easier, not to say simply feasible. The complexity of the brain itself is in fact linear with the total number of synapses with a time constraint due to the integration step of the numerical approximation. This is not trivial since, for a human brain, we would be talking about 10^{14} synapses per millisecond or maybe even 10^{121} synapses per millisecond (10^{11} neurons connected to 10^{11} neurons) if ways of pruning non-existing connections cannot be established. But, for now, we still have a lot to learn with much less neurons and connections and the prototype can run at twenty-something times real-time execution even with the (admittedly over simplistic) physical body and environment simulation.

4.1.2 - The neuronal model

The DoubleLIF neuron model is an extension of the popular "leaky integrate-and-fire" model with a second accumulator to generate non-linear postsynaptic currents (PSCs) from successive spikes taking into account the effect of previous impulses on the postsynaptic cellular potential. DoubleLIF is combining two separate implementations of the conductance-based model connected by a resistance. The first one, identical to the LIF, simulates the entire axon as a single voltage-gated spike producer. The second, reacting to presynaptic spikes, simulates each synapse in the dendritic tree as an individual ligand-gated ionic channel producing a PSC.

The additional accumulator transforms the cellular potential into a true state variable driving the spiking frequency and provides new parameters which develop with time under external stimulation. This metaplasticity allows the simulation of neuron bootstrapping and continuous adaptation in evolving environments.

4.1.3 - The representational system

This neuron is the elementary component of what can become a representational system.

4.1.3.1 – Composability or compositionality

Sensors produce neuronal signals which are analog representations (sensels) of some physical property at some point in the immediate spatiotemporal environment. These sensels (sensory elements) can be composed in more complex representations (percels) by association if they usually exist simultaneously in a common perception field. These percels (perception elements) can also be composed in even more complex representations which we can identify as concepts, assuming that concepts are representations to which we can associate a name, a label, a symbol. Symbolic representation is the lowest level of consciously aware representation, but nonetheless

the result of an already very complex composition process at the subconscious and subconceptual level. The composition of representations is resulting from the physical combination (interconnection) of neurons ("components") into networks. In the preverbal preconscious sensorimotor cognition, representations never reach the symbolic level; the concept might be present, but it is not yet associated with another concept symbolically referring to the same object.

The concept of A-not-B is the neuronal state resulting from a specific set of sensels. It can be used to react to the stimulating sources without having to decode that it means A-not-B. The final response, sent to the actuators, is a global composition of all the available sensels resulting from the immediate spatiotemporal environment. There is no decoding required in the preverbal (subsymbolic) preconscious (subconscious) sensorimotor cognition. Action is encoding all the way; the actuators' response is the final representation of the incoming signals.

4.1.3.2 – Distributed representation

Composability implies the conjunction of multiple signals, but nowhere is the information randomly distributed over a population of neurons. If a light source is perceived by a sensor in the retina, it is also perceived by many other sensors in the retina, but not identically. The signals emitted by the sensors are correlated by their relative position in the retina. Each sensor provides unique information, but this information must be reconcilable with information provided by the other sensors. Between sources and sensors, physical laws apply. At the other end, each fiber in a skeletal muscle, for example, is excited by a unique axon; for efficient cooperation of the fibers, all signals must be perfectly correlated even though they are all different.

4.2 - Regrets

The discovery of the special needs of inhibition came much too late in the project. A lot of time was spent on validating the excitatory behavior of the neuron model, demonstrating that it responded as predicted by our neuroscience postulates (often questioning these postulates), and justifying the role of these postulates in cognition. Inhibition was carried along as a negative stimulation, knowing that its learning and metaplasticity rules had not been fully defined, but expecting they would be variations on the ones being implemented for excitatory stimulation. Unfortunately, at some point, we had to accept that the existing set of equations was not sufficient to generate stable inhibitory connections. We propose an *ad hoc* solution to the problem, but not without raising related questions.

4.2.1 - The numerical method

The proposed solution raised, among others, questions about the numerical method. With respect to the algebraic summation of EPSCs and IPSCs, the order of the spikes within an integration step (1 millisecond) is irrelevant. As long as the stimulation is excitatory, the effect of the spike on synaptic plasticity is also independent of the order of EPSCs; the total change results from the number of spikes. However, when we introduce inhibition, the order of EPSCs and IPSCs becomes significant since any PSC following an EPSC will likely be reinforced while any PSC following an IPSC will be weakened. Now, in our numerical approximation, the neurons are processed serially, rather than in parallel, within an integration step and the order of processing is dictated by the order of neuron instantiation during brain definition, not the actual order of firing within the millisecond step. It is not clear that the problem can be fully addressed at the numerical method level because it might still exist beyond the integration step. Handling spikes as impulses in one integration step makes the signal somewhat noisier, but does not significantly affect the summation process. When it

comes to evaluate the synaptic plasticity changes, the combined effect of multiple spikes might be significantly different from the instantaneous effects of a single spike. The proposed solution spreads the effect of inhibition over multiple integration steps.

4.2.2 - Inhibition

At first glance from a neuroscience perspective, the introduction of inhibition seemed to be relatively simple: inhibitory connections polarize the postsynaptic neurons instead of depolarizing them. The inhibitory "messenger" imagined in the solution last longer and amplifies the inhibition by multiplying the neutralizing factor and by saving the excess power beyond one integration step albeit with a short term decay. Secondary messengers do exist in biological neurons (Kandel al., 2013, ch. 11), but we do not have evidence that they are required for inhibition. A more specifically targeted review of literature would be required to justify such a secondary reaction for inhibition and this might prove difficult since there are much less studies focusing on synaptic plasticity or metaplasticity of inhibitory neurons as illustrated in Maffei (2011). Notwithstanding these considerations, we might have to accept a derogation from our biological plausibility principle to implement an essential functionality.

4.2.3 - The autopoietic network

The proposed solution was tested only in, and consequently only tuned for, a binary representation of a very limited binary environment. This was not sufficient to fully validate the autopoietic capacity of the resulting network. Additional testing and tuning involving multiple sources of different types varying in intensity would be required to fully understand and demonstrate the full range of fuzzy logic for A-not-B gates as A and B vary from 0 to 1. The network's autopoiesis depends on the continuous (analog) property of neuronal representations.

4.3 – Potential developments

First, we have to reconcile the abovementioned shortcomings and confirm that inhibition- based neuronal logic is a fundamental functional requirement of cognition compatible with our postulates: biological plausibility, semiotics and autopoiesis. Then, different avenues will open up to apply and expand the synthetic neuro-cognition framework.

4.3.1 - The simulation

The simulation is a prototype developed to visualize the results generated by the set of differential equations and present different scenarios to evaluate the behavior of an agent responding to external stimulations.

The simulator could be adapted to simulate *in vitro* experiments including STDP protocols and possibly, longer term, investigate *dynamic morphometrics*.

It could also be extended to allow interactive definition of neural networks (agents' brains) and environments for synthetic neuro-cognition experimentations without programming. Statistical tools could be added to evaluate, for example, the covariance of a neuron and all its presynaptic neurons.

Finally, it would be interesting to isolate the brain, embed it in a robot by interfacing all sensors and actuators, and finding ways of developing an architecture strictly by external stimulation. The approach could be modular, since such brains would be clonable and composable at the module level without enforcement of modular encapsulation. Meaning that two modules developed separately could be cloned and merged in a single brain and trained again without any constraint on potential intermodular connections.

4.3.2 - The neuronal model

The possibility of developing experimental protocols for *in vitro* observation of biological neuronal logic A-not-B gates could be investigated in collaboration with electro-neurophysiologists.

4.4 – Conclusion

The heart of the project was the development of a new neuron model, DoubleLIF, including a conductance-based model of ligand-gated stimulation in the dendritic tree as well as voltage-based generation of spikes (ejection of quanta of neurotransmitters) in the axon. This model provides voltage-based synaptic plasticity which is not incompatible with STDP (and easier to apply for online high frequencies) and which is equivalent, for spiking neurons, to the BCM model developed for rate neurons. The parameters of the new accumulator (representing the dendritic tree) can also evolve dynamically and provide an additional level of plasticity (metaplasticity) affecting the basic synaptic plasticity and supporting bootstrapping and continuous selective development as more presynaptic neurons attempt to connect to the dendritic tree.

All these features, which have some biological justification, allow us to claim that the system can be representational and autopoietic. It is representational at a very low level, a semiotic level way below the symbolic level. The combination of neurons entails the composition of representations. The only thing we can consciously experience and describe is the composition of concepts, but this composition is only possible thanks to the causal coupling of physical signals which we cannot consciously experience nor describe. The system is also autopoietic in that the connections between the neurons can develop autonomously under the effect of information.

The intent was to build different networks and observe the constraints and the rules that guided our selection to arrive at a desired functionality. In others words, we wanted to instantiate some algorithms via neuron networks and see if there were rules in the algorithm we were following in our implementation of algorithms: algorithmizing the algorithmization. Unfortunately, we discovered that the rules we had to build networks with inhibitory neurons were not stable enough to guarantee the desired functionality. So, it became a prerequisite to define a logic for spiking neurons.

We expected, maybe hoped, that the algebraic summation of signals would be sufficient to generate stable networks and it was until we started testing the metaplasticity of mixed excitatory and inhibitory neurons. The proposed solution might be more significant than anything we could have done with the original approach. It supports the underlying assumptions that 1) each axon in the brain carries a unique piece of information, 2) this information is the composition of all the pieces of information contributing to its justification including the potential contradictions, 3) the difference between two pieces of information is in itself a new piece of information, and 4) a concept is a set of pieces of information which can be divided in subsets representing more general concepts common to all elements of a given class.

Synthetic neuro-cognition is different from more common cognitive architecture approaches (e.g. Eliasmith 2013) which start from the architectural level and justify, top-down, that the architecture has the properties required for representational structures: systematicity, compositionality, productivity without specifying how far down these properties should apply. For example, Eliasmith proposes convolution-based compositionality at the neuron population level, but does not attribute any meaning (representational capacity) to the individual neurons in these populations. Even to represent a scalar, a population of neurons, generally randomly associated, is

required. Synthetic neuro-cognition, on the contrary, starts at the bottom of the representational structure (sensels) and defines (discovers) rules that apply all the way up to autopoietically generate a cognitive architecture. The spiking neuron logic is the result of such a discovery.

4.5 – Epilogue

We would like to bring to your attention a poster (see Appendix xxx) presented at summer schools[27] in 2010 and 2012 on the evolution of human cognitive systems towards language and consciousness. Because it focuses on the evolution of the representational structure, it is tightly tied to the understanding of synthetic neuro-cognition, but also provides a more global picture of its potential.

---

[27] Institut des Sciences Cognitive, Université du Québec à Montréal (ISC – UQÀM)

APPENDIX A

DoubleLIF

We know from (4) and (6) that

$$C_2 dV_a/dt = (V_c(t) - V_a(t))/R_2 \qquad \text{(A1)}$$

which, for a constant $V_c(t)$, has for solution:

$$V_a(t) = V_c\left(1 - e^{-t/R_2C_2}\right). \qquad \text{(A2)}$$

Knowing that $V_a(t)$ is oscillating between 0 and θ, we can calculate the time required to reach θ simply by replacing $V_a(t)$ by θ in (A2). Rearranging, we find

$$t = -R_2C_2 \ln(1 - \theta/V_c) \qquad \text{(A3)}$$

and consequently since the frequency f = 1/t, we obtain:

$$f = -1/(R_2C_2 \ln(1 - \theta/V_c)) \quad \text{for } V_c > \theta \qquad \text{(A4)}$$

$$0 \quad \text{for } V_c \leq \theta \qquad \text{(A5)}$$

Equation (A4) confirms that the spiking frequency is directly[28] proportional to V_c and that the rate neuron is a particular case of DoubleLIF. In a similar way, it could be shown that the rate neuron is a particular case of classical LIF with constant current input.

---

[28] Or, if you prefer, inversely proportional to the inverse of V_c.

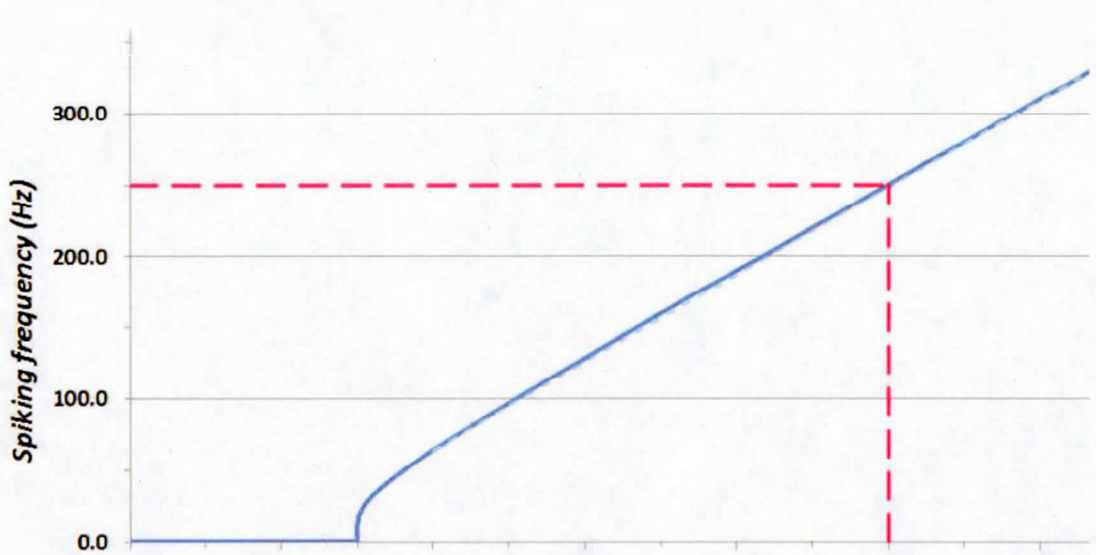


Figure A.1 - DoubleLIF's Frequency to Voltage response

APPENDIX B

Cellular Mechanisms behind STDP

Spike-Timing-Dependent Plasticity (STDP) is a phenomenological model based on an experimental protocol where the strengthening (LTP – Long Term Potentiation) or the weakening (LTD – Long Term Depression) of a synaptic connection is predicted from the tightly controlled time difference between a pair of presynaptic and postsynaptic spikes.

As defined in Sjöström and Gerstner (2010), the experiment requires two neurons (*i, j*) with a synaptic connection (of strength w_{ij}) where an action potential (AP) in *j* induces an excitatory postsynaptic potential (EPSP) in *i*. Both neurons (*i, j*) are under dual whole-cell voltage recordings (V_{ci}, V_{cj}) and possibility to generate precisely timed (sub millisecond) current pulses in one neuron or the other.

The presynaptic spike

When properly sized, a current pulse in the presynaptic neuron *j* generates an AP in its axon which induces an EPSP in the postsynaptic neuron *i*. Normally, a single EPSP will not induce an AP in the postsynaptic neuron, unless its cellular potential is artificially maintained very close to the spiking threshold.

The postsynaptic spike

Similarly, a properly sized current pulse in the postsynaptic neuron i generates an AP in its axon. Clearly, this postsynaptic spike is not, in any way, causally related to the presynaptic spike any more than the pre was causally related to the post. If a correlation can be established between the controlled time delay (t_{pre}-t_{post}) and the changes in synaptic connection strength (Δw_{ij}), which causal mechanisms could be involved?

Action potentials (AP)

As Hodgkins and Huxley (1952) discovered, APs are produced in, and propagated along, the axons when voltage-gated Na^+ channels are opened because the cellular potential (V_c) has exceeded the threshold (~-40 mv) letting sodium ions rush in depolarizing even more the area until voltage-gated K^+ channels open when V_c reaches their threshold (~0 mv) letting potassium ions rush out repolarizing the cell and momentarily closing Na^+ channels. V_c goes then back to its resting value (~-69 mv) after a significant and slowly receding hyperpolarization.

Dendritic action potentials (dAP)

EPSPs are not very different from APs except that they are produced at synaptic connections when neurotransmitters are projected across the synaptic gap by the arrival of an AP at the axon terminal of the presynaptic neuron and ligand-gated Na^+ channels open in the synaptic spine of the postsynaptic neuron. The number of ligand-gated Na^+ channels activated at a synaptic spine by an axon terminal being much less than the number of voltage-gated Na^+ channels in the trigger zone at the axon initial segment, the resulting Na^+ current (EPSC), hence the resulting EPSP, is rarely sufficient to induce, by itself, an AP especially when the synapse is far from the axonal cone. Only in very special cases, portions of the dendritic tree have a sufficiently high density of voltage-gated channels to produce true active APs. This is generally not the case in, and certainly not a requirement for, STDP experiments.

Backpropagating action potentials (bAP)

Whenever an EPSP, strong enough to trigger the opening of voltage-gated Na^+ channels, reaches the axon cone, a full-fledge AP (total depolarization to ~+30 mv followed by hyperpolarization) is produced and actively travels down the axon. At the same time, the AP propagates back into the dendrites, albeit with reducing intensity due to the scarcity (or absence) of voltage-gated channels. The strong (relative to EPSCs) depolarizing current travels further back in all dendritic branches than any incoming EPSP. It should be noted that, although the wave is created by incoming Na^+ ions, these ions do not have to migrate, only electrons are pulled from all branches of the dendritic tree; only the electrons are redistributed in an attempt to re-equilibrate the cellular potential. A new electronic wave refluxes in the opposite direction as soon as the K^+ ions rush in to complete the AP.

Artificial EPSPs

The artificial current pulse generates an artificial EPSP which transforms into a bAP if strong enough to induce an AP when it reaches the axonic hillock. The resulting bAP produces an influx and a reflux of electrons throughout the dendritic tree including over the synapse activated by the neurotransmitters emitted by the presynaptic AP.

The STDP protocol

In an STDP experiment, there are two causal events, the presynaptic and the postsynaptic pulses, separated by a predetermined delay. The objective of the presynaptic pulse is to induce an AP in the axon of the presynaptic neuron to activate the synapse. It is the activation of that synapse (i.e. the opening of ligand-gated ionic channels via the emission of neurotransmitters) that is the real object under study during the experiment: how big an EPSP will this opening produce and how would this EPSP be affected by a postsynaptic pulse? So, an AP in the presynaptic neuron sends neurotransmitters across the synaptic gap and these neurotransmitters unlock

specific ionic channels creating an EPSC if the channels are Na^+ channels[29]. The intensity of the EPSC is proportional to the number of open channels and their efficiency. The rise of the EPSP coincide with the duration of the EPSC and consequently with the opening of the ligand-gated channels. Then those channels close and the potential goes back to its resting value as K^+ ions leak out and Cl^- ions leak in through permanently open channels. Because the number of permanently open channels is small relative to gated channels, the EPSP slowly decays to resting value without any overshoot. On induction of an artificial current pulse, the cellular potential starts climbing until it triggers a true AP at the axonal cone. Then, the cellular potential shoots up until the K^+ voltage-gated channels open depolarizing the cell beyond its resting potential. At that point, all gated channels are closed and the leak slowly brings the potential to its resting equilibrium through the few permanently open channels.

The coincidence

The STDP protocol is designed to study the relative coincidence of these two events: the opening of ligand-gated Na^+ channels and the propagation of a backpropagating action potential over these channels. The experiments allow the measurement of the effect of this coincidence on the changes in synaptic strength (EPSP intensity). In other words, the STDP curves of change vs pre-post delay are the result of a cross-correlation of the number of open channels, *N(t)*, and the backpropagating potential at the synapse, *bAP(t)*. This can be interpreted as the effect of high concentration of positive ions on the reaction rate of protein-building organic anions. Depolarization favors the metabolism (or at least activation) of whichever components facilitate the inward flow of Na^+.

---

[29] For the purpose of this addendum, we will consider only excitatory neurons. Inhibitory neurons (K^+ or Cl^- channels) will be addressed separately.

In summary, it is the average voltage around the channels while (and shortly after) they are open that is defining the potentiation/depression of the synaptic connection. The STDP protocol ensures that a given voltage will be present for a controlled opening of the channels. The opening can be controlled by selecting a delay on a monotonically closing population of channels. The important factor is the voltage level while the channels are open; whether the effect is homosynaptic or heterosynaptic is totally irrelevant.

The model

Error! Reference source not found. shows, in the main section, four curves with a shape very similar to the curves typically fitted from data resulting from STDP experiments (Sjöström 2011). In this case, the curves are not based on experimental data, but they are built from the cross-correlation of *N(t)* and *dAP(t)*.

$$\frac{\Delta w_{ij}}{w_{ij0}} = y(\delta) = \int_{-\infty}^{\infty} N(t-\delta) * \left(bAP(t) + (V_{hold} - V_{\theta})\right) dt$$

where: w_{ij} is the strength of the synapse;

w_{ij0}, the strength at the beginning of the experiment;

Δw_{ij}, the change in strength after a series of pulses;

$\Delta w_{ij}/w_{ij0}$, the change relative to the initial strength ($>0 \rightarrow$ LTP; $<0 \rightarrow$ LTD);

δ is the time delay between the presynaptic (t_{pre}) and the postsynaptic (t_{post}) pulses.

$$N(t) = 0 \quad \text{for } t \leq 0$$
$$Ae^{-t/\tau} \quad \text{for } t > 0.$$

$$dAP(t) = \begin{cases} 0 & \text{for } t \leq 0 \\ B & \text{for } t = 1 \\ 0 & \text{for } t = 2 \\ -B*(1-(t-3)/(50-3)) & \text{for } t > 2. \end{cases}$$

V_{hold} is the resting potential and V_Θ is the voltage threshold between LTD and LTP as defined by the BCM model (same as θ_M). If V_{hold} is lower than V_Θ, LTD can happen without postsynaptic spikes since the conditions are met whenever the channels open. The same should be true of LTP, but the probability of having $V_{hold} > V_\Theta$ without exceeding the spiking threshold of the postsynaptic neuron seems to be much smaller.

N(t) (dashed line) represents the number of channels opened by the induced presynaptic action potential. *bAP(t)* (full line) represents the backpropagating action potential resulting from the artificial current pulse with a sharp positive pulse lasting 2 ms followed by a negative overshoot linearly recessing over 48 ms.

The five smaller drawings, at the top, alphabetically numbered from A to E, illustrate how each point on a curve (in this case, the third curve from the bottom which starts at $\Delta w_{ij}/w_{ij0} = 0$) is calculated.

A- The presynaptic AP happens 50 ms before the postsynaptic AP and all the channels are closed by then.

B- The presynaptic AP happens 20 ms before the postsynaptic AP. There is still a small fraction of the channels open at the time of the postAP such that its positive impulse has some LTP effect followed by smaller LTD impact of the hyperpolarization integrated over the remaining open channels.

C- The presynaptic AP happens just before the postsynaptic AP and all the channels are open for a maximum LTP impact of the full impulse of the postAP somewhat reduced by the integration of the entire hyperpolarization. If the presynaptic AP happens 2 ms later, the

postsynaptic AP and all the channels are open for a maximum the full LTP impact of the postAP impulse is missed and the synapse sees only the LTD effect of the entire hyperpolarization.

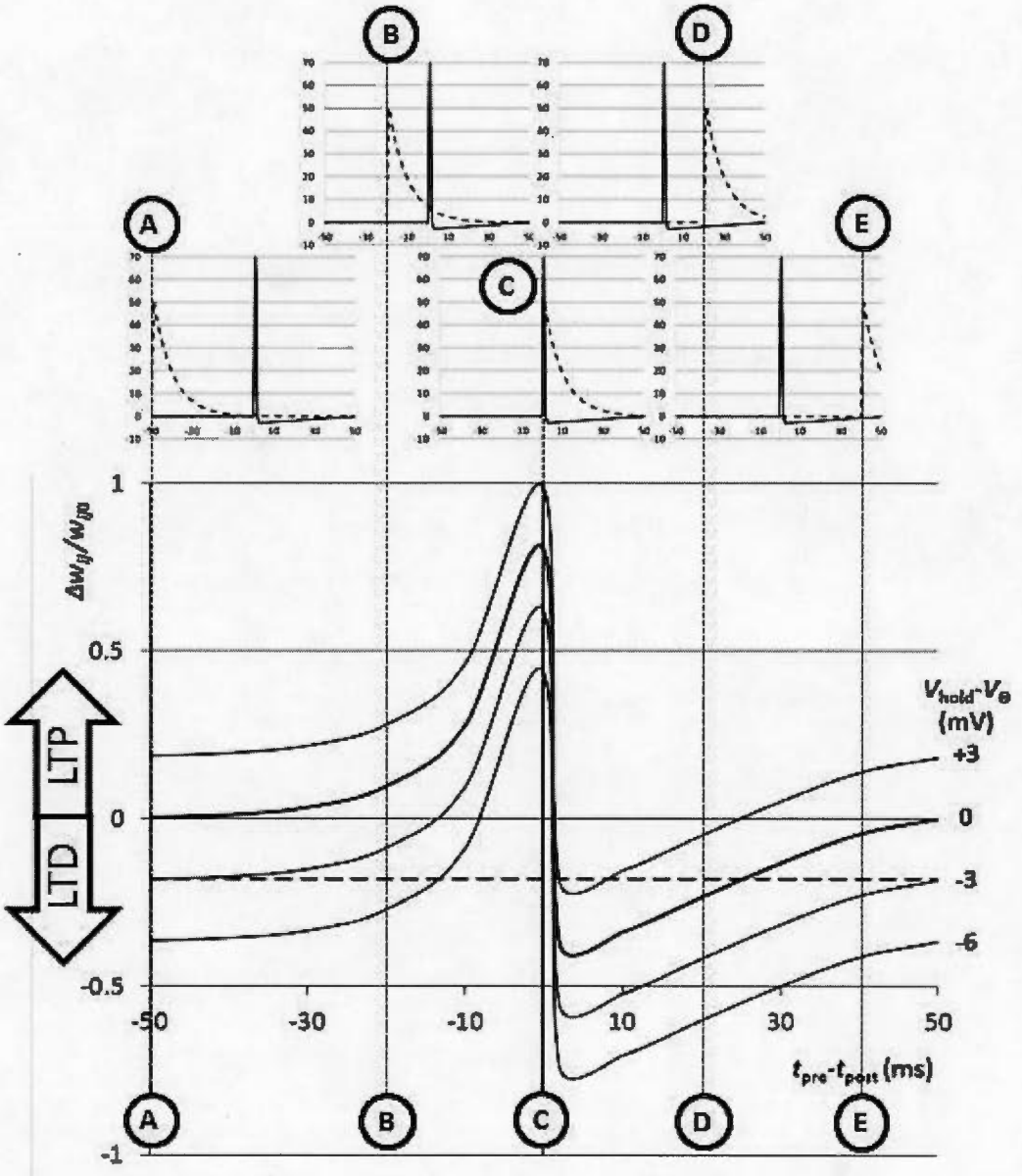


Figure B.1 - STDP cellular mechanisms

D- The presynaptic AP happens just before the postsynaptic AP and all the channels are open for a maximum LTP impact of the full impulse of the postAP somewhat reduced by the integration of the entire hyperpolarization. If the presynaptic AP happens 2 ms later, the postsynaptic AP and all the channels are open for a maximum the full LTP impact of the postAP impulse is missed and the synapse sees only the LTD effect of the entire hyperpolarization.

E- The presynaptic AP happens 20 ms after the postsynaptic AP. The LTP potential of the postAP impulse has been missed and the synapse sees a lesser LTD effect via the integration of the remaining portion of the hyperpolarization.

The presynaptic AP happens 50 ms after the postsynaptic AP and the synapse is unaffected since all traces of the postAP have disappeared by that time.

APPENDIX C

Dirac delta function

For a brief introduction to Dirac delta functions, the reader is referred to Chatfield (1975, Appendix II, pp238-9 [30]). As he says (p238), « It is important to realize that $\delta(t)$ is *not* a function. Rather it is a generalized function, or distribution, which maps a function into a real line. » (original emphasis)

Having defined a set s of t_i

$$s = \{ t > 0 \text{ such that } F(t) = 0 \}$$

we can define

$$f(t) = \int_{-\infty}^{+\infty} g(t) \sum_s \delta(t - t_i)\ dt$$

and assuming that S is finite with N elements, we can say

$$= \sum_{i=1,N} \int_{-\infty}^{+\infty} g(t)\delta(t - t_i)\, dt$$

which also applies to smaller time intervals (Chatfield (1975, p239, statement of exercise 1)

$$= \sum_{i=1,n} \int_{t}^{t+\Delta t} g(t)\delta(t - t_i)\, dt$$

---

[30] Available at https://books.google.ca/books?id=u1D5BwAAQBAJ&pg=PA238&lpg=PA238&dq=dirac+delta+function+time+series+analysis&source=bl&ots=LyXtkI2_Q1&sig=O2PRwaro3x3XxVDDQ_YpFppnqb8&hl=en&sa=X&ved=0CD0Q6AEwB2oVChMIkOmP34iJyQIVypoeCh1lDgMX#v=onepage&q=dirac%20delta%20function%20time%20series%20analysis&f=false

where $i=1,n$ represents the elements of s in this smaller interval $]t, t+\Delta t]$ and is equal to:

$$= \sum_{i=1,n} g(t_i)$$

If the maximum frequency of t_i is carefully selected to ensure that the smallest interval between two t_i always exceeds Δt (e.g. frequency of t_i < 1kH for Δt = 1 ms), there is never more than 1 element in a Δt interval and we obtain:

$$f(t) = g(t_1)$$

Unfortunately, the set s of t_i cannot be defined beforehand because it depends on the very function that we are trying to solve $V_a(t)$, since impulses are produced whenever $V_a(t)$ reaches θ_S (in other words, $F(t) = V_a(t) - \theta_S$). However, our set of differential equations being non-homogenous and non-linear, we already knew that we would have to resort to numerical approximation to find a solution.

In numerical approximation, variables are evaluated at short time intervals (Δt) such that, in the case of our equation 6, we obtain:

$$V_a((k+1)\Delta t) = V_a(k\Delta t) + I_{fire}(k\Delta t)*\Delta t$$

If $V_a((k+1)\Delta t)$ is larger than θ_S, it means that, at some point in the interval $]\Delta t,(k+1)\Delta t]$, $V_a(t)$ reached θ_S and an impulse should have been generated bringing $V_a(t)$ back to 0. So, the value of $V_a((k+1)\Delta t)$ is corrected by subtracting θ_S. Of course the correction is not exact, but well within the precision of numerical approximation methods.

In fact, our set s of t_i is not projecting directly on t, but rather on a bigger set S of t_k

$$S = \{ t = k\Delta t \text{ for } k = 0,\infty \}$$

To make sure that $F(t)$ is equal to 0 when evaluated at $t_{k\Delta t}$, we have to define:

$$F(t) = min(0, V_a(t) - \theta_S)$$

such that $F(t)$ is equal to 0 when evaluated at the first $k\Delta t$ following the moment where $V_a(t)$ reached θ_S in real time. In our real time equations 5 and 6, the *min* function is not required since $V_a(t)$ never exceeds θ_S in real time, but we leave it in the equation to indicate how it should be done in the numerical approximation.

Figure C.1 shows the evaluation of $V_a(t)$ over a few time intervals including one with an impulse generation.

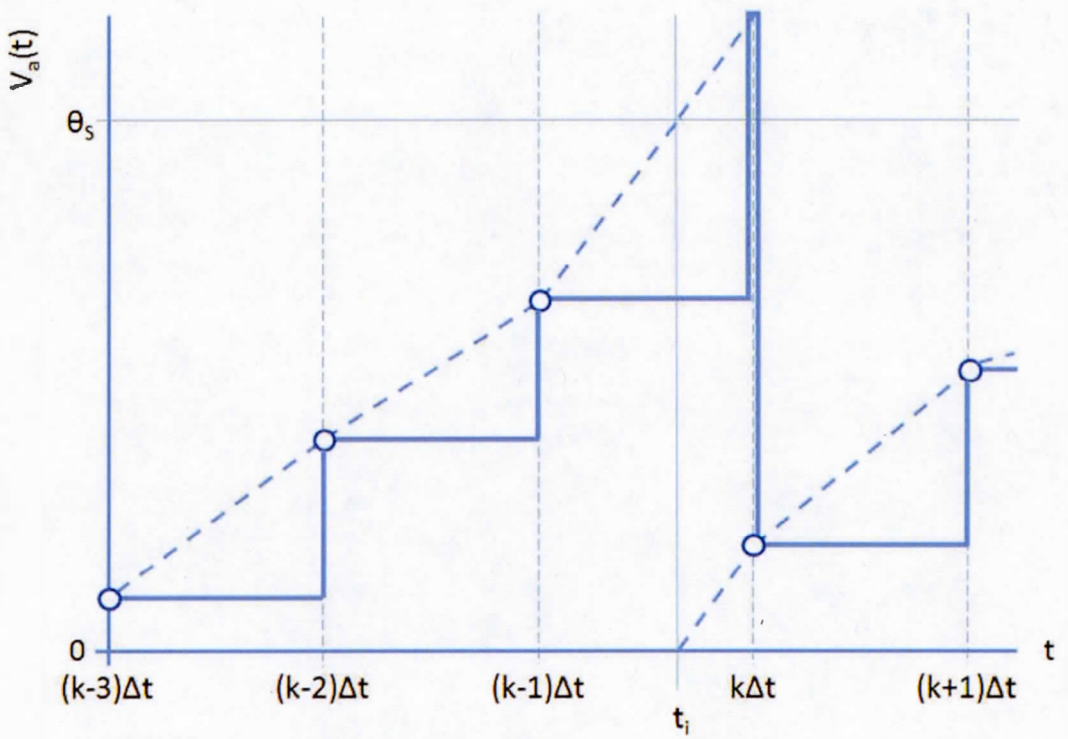


Figure C.1 - Numerical approximation with Dirac delta function

APPENDIX D

Two legs, two hands and the loss of smell

An essay on the origin of language and…the emergence of consciousness

Two legs, two hands and the loss of smell

An essay on the origin of language and... the emergence of consciousness

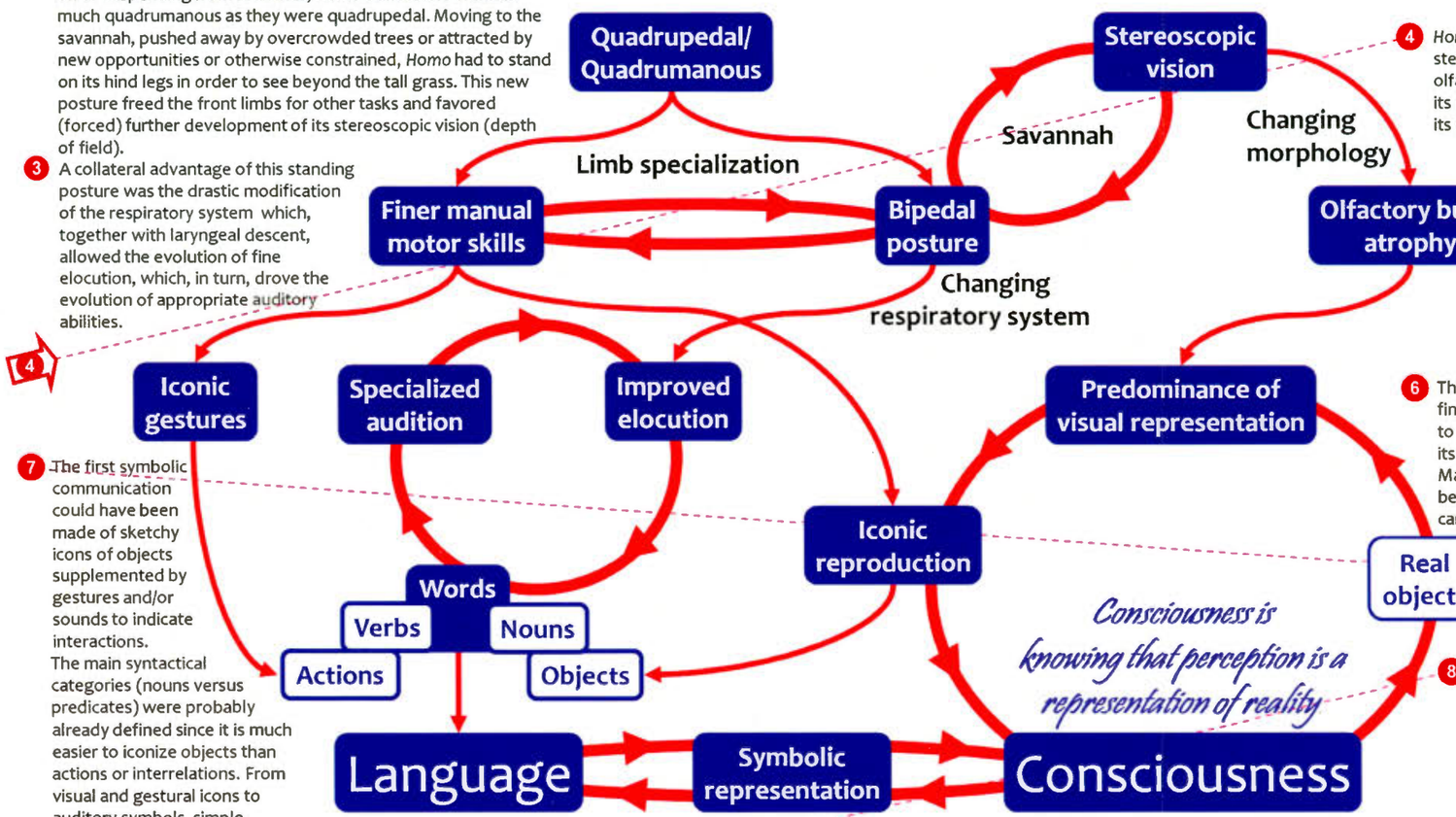
UQÀM
L'origine du langage
Université du Québec à Montréal
Institut d'été en sciences cognitives
Montréal, du 21 au 30 juin 2010

Pierre Vadnais
Doctorat en Informatique Cognitive
Université du Québec à Montréal
vadnais.pierre@courrier.uqam.ca 514-717-4417

1 It is generally well accepted that *Homo* walked away from his primate cousins on his hind legs. Was this fact sufficient to trigger all aspects of the ensuing evolution of hominids, more specifically the emergence of language and consciousness?

2 Having evolved in the jungle, leaping from tree to tree, walking on or suspending from branches, *Homo*'s ancestors were as much quadrumanous as they were quadrupedal. Moving to the savannah, pushed away by overcrowded trees or attracted by new opportunities or otherwise constrained, *Homo* had to stand on its hind legs in order to see beyond the tall grass. This new posture freed the front limbs for other tasks and favored (forced) further development of its stereoscopic vision (depth of field).

3 A collateral advantage of this standing posture was the drastic modification of the respiratory system which, together with laryngeal descent, allowed the evolution of fine elocution, which, in turn, drove the evolution of appropriate auditory abilities.



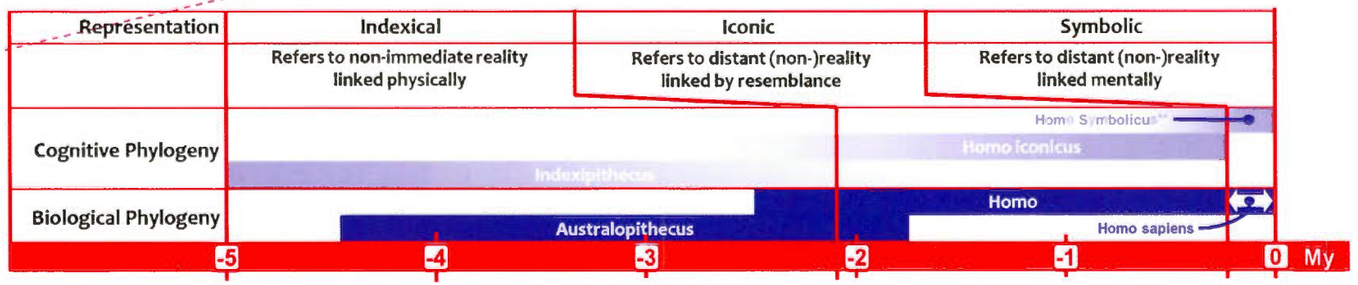
4 *Homo* also inherited from its simian ancestors a stereoscopic vision and the resulting atrophy of the olfactory bulbs. By perfecting the former, to adapt to its new environment (the savannah), it further reduced its sense of smell.

5 This loss of smell overturned the relative importance of the different inputs and representation became predominantly visual.

6 This representational shift enabled *Homo* to use the finer motor abilities of its hands (developing in parallel) to physically reproduce external analogues (icons) of its now mainly visual mental representations. Mammals cannot reproduce their representations because these are predominantly olfactive and odors cannot easily be reproduced.

7 The first symbolic communication could have been made of sketchy icons of objects supplemented by gestures and/or sounds to indicate interactions. The main syntactical categories (nouns versus predicates) were probably already defined since it is much easier to iconize objects than actions or interrelations. From visual and gestural icons to auditory symbols, simple association was sufficient to bridge the gap and symbolic language could evolve. With time, external icons and gestures became redundant and superfluous.

8 The coexistence of a real object and its manually reproduced icon with a mentally identifiable resemblance constitutes the first prototype of Peirce's* semiotic triad. These icons allow *Homo* to refer to spatially or temporally distant real or non-real objects. The physical link required by indexical reference can be replaced by a mental link.



* Peirce, C.S. (1897, 1903) Logic as semiotics: The theory of signs. In J. Buchler, ed., *The Philosophical Writings of Peirce* (1955). New York: Dover Books, 98-119.

** Deacon, T.W. (1997) *The Symbolic Species*. New York / London: W.W. Norton and Company.

UQÀM
Evolution and Function of Consciousness
Université du Québec à Montréal
Institut d'été en sciences cognitives
Montréal, du 29 juin au 11 juillet 2012

2012 addendum

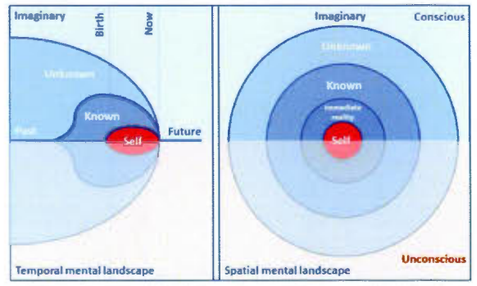
Emergence of consciousness

Preverbal mental landscape

A long long time ago, when words did not yet exist, animals' mental representations of the world were limited to the spatiotemporal immediate reality accessible through their sensors. Of course this restricted reality included proprioception and interoception accessible to each individual, but not others, which we could call a self. Stimuli from this immediate reality, including the self, could also bring back some associated memories.

Iconization (reification of mental representation)

When *Homo* reached the point of *predominant visual representation and finer manual motor skills*, it was able to produce concrete external analogues (physical icons) which, by virtue of physical resemblance, activated some of the same neural regions and pathways as the mental representation of another physical object in memory. The permanence of manually produced physical icons gave time to *Homo* to associate gestural and acoustic symbols (gestures and words) to the shared mental representation. With time, and repetition, these gestures and words became associated with the icon and, by proxy, with the distant object referred to by the icon through resemblance. This reification of representations through production of icons, followed by association of symbols, opened a *mental portal* providing access to a spatio-temporally distant (i.e. currently absent) reality and thereby favored the emergence of consciousness.



Memory became the "known", a set of grounded symbols available to bring back, *at will*, existing memories. A complementary set of ungrounded symbols formed the "unknown" and some concepts created mentally by mixing of known concepts but not groundable in reality formed the "imaginary". The mental landscape, and the physical world, were and will never be, the same because, from then on, words became real physical objects.

BIBLIOGRAPHY

Abbott, L. F. and Nelson, S. B. (2000). Synaptic plasticity: taming the beast. *Nat. Neurosci.* **3**(Suppl.), 1178–1183.

Abraham, W.C. (2008). Metaplasticity: tuning synapses and networks for plasticity. *Nature Reviews Neuroscience*, **9**: 387-399

Abraham, W.C. and Philpot, B. (2009). Metaplasticity, in *Scholarpedia*, 4(5):4894. Available: http://www.scholarpedia.org/article/Metaplasticity

Abraham, W.C. and Bear, M.F. (1996). Metaplasticity: the plasticity of synaptic plasticity. *Trends Neurosci.* **19** (4): 126–30.

Adee, S. (2009). Cat Fight Brews Over Cat Brain on *IEEE Spectrum Tech Talk Blog*, November 23. Available: http://spectrum.ieee.org/tech-talk/semiconductors/devices/blue-brain-project-leader-angry-about-cat-brain

Adrian, E.D. (1926a). The impulses produced by sensory nerve endings: Part I. *J. Physiol.* 61:49–72.

Adrian, E. D., (1926b). The impulses produced by sensory nerve-endings: Part IV: Impulses from pain receptors, *J. Physiol.*, 62, 33 -51.

Adrian, E.D. and Zotterman, Y. (1926a). The impulses produced by sensory nerve endings: Part II: The response of a single end organ. *J Physiol* 61: 151–171.

Adrian, E.D. and Zotterman, Y. (1926b). The impulses produced by sensory nerve endings: Part III: The impulses produced by sensory nerve endings. *J Physiol* 61: 465–483.

Ananthanarayanan, R., Esser S.K., Simon H.D. and Modha, D.S. (2009). The cat is out of the bag: Cortical simulations with 10^9 neurons and 10^{13} synapses, in *Proceedings of the ACM/IEEE Conference on Supercomputing* (Portland, OR, Nov. 14–20). ACM , New York, NY, 1–12.

Arkin, R.C. (1998). *Behaviour-based Robotics*, Cambridge MA: MIT Press.

Armstrong, D. (1968). *A Materialistic Theory of the Mind,* London: RKP

Beck, K. (2000). *Extreme Programming Explained: Embrace Change,* Addison-Wesley, USA and Canada.

Bi, G.Q. and Poo, M. (2001). Synaptic modification by correlated activity: Hebb's postulate revisited, *Annu Rev Neurosci* 24: 139–166.

Bickerton, D. (1990). *Language and Species*. Chicago: University of Chicago Press.

Billard, A. and Dautenhahn, K. (1999). Experiments in Learning by Imitation Grounding and Use of Communication in Robotic Agents, *Adaptive Behaviour*, 7, 411-434.

Bienenstock, E.L., Cooper, L.N. and Munro, P.W. (1982). Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex, *J. Neurosci.* 2:32–48.

Blais, B.S. and Cooper, L. (2008). BCM theory, in *Scholarpedia*, 3(3):1570. Available: http://www.scholarpedia.org/article/BCM_theory

Bliss, T. and Lømo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *J. Physiol.* (Lond.) 232, 331–356.

Brader J., Senn W., Fusi S. (2007). Learning real-world stimuli in a neural network with spike-driven synaptic dynamics. *Neural Comput.* 19, 2881–291210.1162/neco.2007.19.11.2881

Braitenberg, V. (1984). *Vehicles: Experiments in synthetic psychology*. Cambridge, MA:MIT Press.

Brooks, R.A. (1986). A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, RA-2, 14-23

Brooks, R.A. (1989). A robot that walks: Emergent behaviors from a carefully evolved network. *Neural Computation,* 1, 153-162.

Brooks, R.A. (1990). Elephants Don't Play Chess, *Robotics and Autonomous Systems* 6, 3-15.

Brooks, R.A. (1991). Intelligence without representation, *Artificial Intelligence*, 47:139-160

Brunel, N. and van Rossum, MCW (2007). Quantitative investigations of electrical nerve excitation treated as polarization, *Biol. Cybern.*, 97:341–349.

Bullock T.H., Bennett M.V., Johnston D., Josephson R., Marder E. and Fields R.D. (2005). The neuron doctrine, redux in *Science,* 310(5749):791-3.

Chalmers, D.J. (1995). Facing Up to the Problem of Consciousness in *Journal of Consciousness Studies,* 2(3):200-19

Chatfield, C. (1975). *The Analysis of Time Series: Theory and Practice*, Chapman and Hall

Chomsky, N. (1959). A Review of B. F. Skinner's Verbal Behavior in Jakobovits, L.A. and Miron, M.S. (eds.), *Readings in the Psychology of Language*, Prentice-Hall, 1967, pp. 142-143. http://www.chomsky.info/articles/1967----.htm 2006.12.18.

Church, A. (1932). A set of Postulates for the Foundation of Logic. *Annals of Mathematics*, second series, 33, 346-366

Church, A. (1936a). An Unsolvable Problem of Elementary Number Theory. *American Journal of Mathematics*, 58, 345-363

Church, A. (1936b). A Note on the Entscheidungsproblem. *Journal of Symbolic Logic*, 1, 40-41.

Clopath C., Ziegler L., Vasilaki E., Büsing L. and Gerstner W. (2008). Tag-trigger-consolidation: a model of early and late long-term-potentiation and depression. *PLoS Comput. Biol.* 4, e1000248 http://dx.doi.org/10.1371%2Fjournal.pcbi.100024810.1371/journal.pcbi.1000248

Clopath, C., Büsing, L., Vasilaki, E. and Gerstner, W. (2009). Connectivity reflects Coding: A Model of Voltage-based Spike-Timing-Dependent-Plasticity with Homeostasis. Available from Nature Precedings http://hdl.handle.net/10101/npre.2009.3362.1 Published version: *Nature Neurosci.* 13, 344–352 (2010).

Cooper, L.N. and Bear, M.F. (2012). The BCM theory of synapse modification at 30: interaction of theory with experiment in Nature Reviews Neuroscience 13, 798-810 (November 2012)

Cooper, L. N., Liberman, F. and Oja, E. (1979). A theory for the acquisition and loss of neuron specificity in visual cortex. *Biol. Cybern.* 33, 9–28.

Davidson, P. (1993). Toward a General Solution to the Symbol Grounding Problem: Combining Machine Learning and Computer Vision, *in AAAI Fall Symposium Series, Machine Learning in Computer Vision: What, Why and How*, 157-161

Dennett, D. (1997). *Kinds of Minds: Towards an Understanding of Consciousness*, Basic Books

De Robertis, ED and Bennett, HS (1955). Some features of the submicroscopic morphology of synapses in frog and earthworm in *J. Biophys. Biochem. Cytol.* 1:47–58

Dretske, F. I., (1981). *Knowledge and the Flow of Information*, Oxford: Blackwell; reprinted, Stanford, CA: CSLI Publications, 1999.

Dreyfus, H. and Dreyfus, S. (1986). *Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer*. Oxford, UK: Blackwell

Eliasmith, C. and Anderson, C.H. (2003). *Neural Engineering: Computation, Representation, and Dynamics in Neurobiological Systems*, MIT Press, Mass., USA

Elman, J.L., Bates, E.A., Johnson, M.H., Karmiloff-Smith, A., Parisi, D. Plunkett, K. (1996). *Rethinking Innateness: A Connectionist Perspective on Development*, paperback edition 1998, MIT Press, Cambridge, Massachusetts.

Enard W, Gehre S, Hammerschmidt K, Holter SM, Blass T, Somel M, Bruckner MK, Schreiweis C, Winter C, Sohr R, Becker L, Wiebe V, Nickel B, Giger T, Muller U, Groszer M, Adler T, Aguilar A, Bolle I, Calzada-Wack J, Dalke C, Ehrhardt N, Favor J, Fuchs H, Gailus-Durner V, Hans W, Holzlwimmer G, Javaheri A, Kalaydjiev S, Kallnik M, Kling E, Kunder S, Mossbrugger I, Naton B, Racz I, Rathkolb B, Rozman J, Schrewe A, Busch DH, Graw J, Ivandic B, Klingenspor M, Klopstock T, Ollert M, Quintanilla-Martinez L, Schulz H, Wolf E, Wurst W, Zimmer A, Fisher SE, Morgenstern R, Arendt T, de Angelis MH, Fischer J, Schwarz J, Paabo S (2009). A humanized version of Foxp2 affects cortico-basal ganglia circuits in mice. *Cell* 137:961–971.

Floridi, L. (2011a). Semantic Conceptions of Information, *The Stanford Encyclopedia of Philosophy (Spring 2011 Edition)*, Edward N. Zalta (ed.), http://plato.stanford.edu/archives/spr2011/entries/information-semantic/

Floridi, L. (2011b). *The Philosophy of Information,* Oxford University Press

Fodor, J. (1975). *The Language of Thought*, New York: Thomas Crowell.

Galvani, L. [1791] (1953). Commentary on the effect of Electricity on Muscular Motion. RM Green (transl). Cambridge, MA: Licht.

Gerstner, W. and Kistler, W.M. (2002). *Spiking Neuron Models: Single Neurons, Populations, Plasticity*, Cambridge, UK: Cambridge UP.

Gladwell, M. (2005). *Blink: The Power of Thinking Without Thinking,* New York: Little, Brown and Co.

Gödel, K. (1931). Über Formal Unentscheidbare Sätze der Principia Mathematica Und Verwandter Systeme I. *Monatshefte für Mathematik*, 38 (1):173-198.

Harnad, S. (1990). The Symbol Grounding Problem. *Physica* D 42: 335-346.

Harnad, S. (1992). The Turing Test Is Not A Trick: Turing Indistinguishability Is A Scientific Criterion. *SIGART Bulletin* 3(4) (October 1992) pp. 9 - 10.

Hebb, D.O. (1949). *The Organization of Behavior: A Neuropsychological Theory,* NY:Wiley.

Hilbert, D. (1900). Mathematical Problems, Lecture delivered before the International Congress of Mathematicians at Paris in 1900 http://aleph0.clarku.edu/~djoyce/hilbert/problems.html 2011.12.02

Hodgkin, A.L. and Huxley, A.F. (1952). A Quantitative Description of Membrane Current and its Application to Conduction and Excitation in Nerve, *J Physiol.* 117(4): 500–544, Aug 1952.

Izhikevich, E.M. (2003). Simple model of spiking neurons, *IEEE Transactions on Neural Networks*, vol. 14, pp. 1569–1572.

Izhikevich, E.M. (2004). Which Model to Use for Cortical Spiking Neurons, *IEEE Transactions on Neural Networks*, vol. 15, pp. 1063-1070.

Izhikevich, E.M. and Desai, N.S. (2003). Relating STDP to BCM. *Neural. Comput.* 15, 1511–1523.

Izhikevich, E.M. and Edelman, G.M. (2008). Large-scale model of mammalian thalamocortical systems. *Proceedings of the National Academy of Sciences of the USA* 105(9), 3593–3598.

Kandel, E.R., Schwartz, J.H., Jessell, T.M., Siegelbaum, S.A. and Hudspeth, A.J. (2013). *Principles of Neural Science*, Fifth Edition, McGraw-Hill, New-York, USA

Kempter, R., Gerstner, W., and van Hemmen, L. (1999). Hebbian learning and spiking neurons. *Phys. Rev. E* 59, 4498–4514.

Kleene, S.C. (1952). *Introduction to Metamathematics*. Amsterdam, North-Holland

Kohn, A.F. (1989). Dendritic transformations on random synaptic inputs as measured from a neuron's spike train. Modeling and simulation. *IEEE Trans Biomed Eng* 36:44-54.

Lapicque, L. (1907). Recherches quantitatives sur l'excitation électrique des nerfs traitée comme une polarization, *J Physiol Pathol Générale* 9:567–578.

Leibniz, G.W. (1714). *La Monadologie*. http://classiques.uqac.ca/classiques/Leibniz/La_Monadologie/leibniz_monadologie.pdf 2011.11.21

Lisman, J., and Spruston, N. (2005). Postsynaptic depolarization requirements for LTP and LTD: a critique of spike timing-dependent plasticity. *Nat. Neurosci.* 8, 839–841.

Lisman, J. and Spruston, N. (2010). Questions about STDP as a general model of synaptic plasticity in *Frontiers in Synaptic Neuroscience*, 2 (140)

Markov, A.A. (1960). The Theory of Algorithms. *American Mathematical Society Translations*, series 2, 15, 1-14

Maass, W. (1997). Networks of spiking neurons: The third generation of neural network models, *Neural Networks*, vol. 10, pp. 1659–1671.

Maffei, A. (2011). The many forms and functions of long term plasticity at GABAergic synapses. *Neural Plasticity*, Volume 2011, Article ID 254724, 9 pages.

Markram, H. (2006). The Blue Brain Project. *National Review of Neuroscience* 7, 2, 153–160.

Markram, H., Lübke, J., Frotscher,M., Roth,A. and Sakmann, B. (1997a). Physiology and anatomy of synaptic connections between thick tufted pyramidal neurones in the developing rat neocortex. *J. Physiol.*(Lond.) 500, 409–440.

Markram, H., Lübke, J., Frotscher,M., and Sakmann, B. (1997b). Regulation of synaptic efficacy by coincidence of post synaptic Aps and EPSPs. *Science* 275, 213–215.

Maturana, H.R. (1970). Biology of cognition in Maturana, H.R. Varela, F.J. (1980).

Maturana, H.R. Varela, F.J. (1980). *Autopiesis and Cognition: The Realization of the Living*. Boston Studies in the Phylosophy of Science, vol. 42. Dordrecht: D. Reidel.

Mayo, M. (2003). Symbol Grounding and Its Implication for Artificial Intelligence, *Twenty-Sixth Australian Computer Science Conference*, ACSC2003 (Adelaide, Australia), 55-60.

McCarthy, J., Minsky, M. L., Rochester, N. and Shannon, C.E., (1955). A Proposal For The Dartmouth Summer Research Project On Artificial Intelligence http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html 2006.12.17

McCormick D.A., Connors, S.W., Lighthall, J.W. Prince, D.A. (1985). Comparative electrophysiology of pyramidal and sparsely spiny stellate neurons of the neocortex. *J Neurophysiol* 54, 782-806

McCulloch, W. S. and Pitts, W. (1943). A logical calculus of ideas immanent in nervous activity. *Bull. Math Biophys*. 5, 115-133.

Miller, G.A. (1983). Informavores in Machlup, F. Mansfield, U., *The Study of Information: Interdisciplinary Messages*, Wiley-Interscience, pp. 111–113

Minsky M. L. and Papert S. A. (1969). *Perceptrons*. Cambridge, MA: MIT Press

Modha, D.S., Ananthanarayanan, R., Esser, S.K., Ndirango, A., Sherbondy, A.J. and Singh, R. (2011). Cognitive Computing, *Communications of the ACM*, august 2011, 54(8), 62-71.

Nernst, W. (1899). Zur Theorie der elektrischen Reizung in *Nachrichten von der Königl. Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, Heft 1:104–108. http://gdz.sub.uni-goettingen.de/dms/load/pdf/?PPN=PPN252457811_1899&DMDID=DMDLOG_0007&LOGID=LOG_0007&PHYSID=PHYS_0013

Nernst, W. and Barratt, J.O.W. (1904). Über die elektrische Nervenreizung durch Wechselströme. *Zeitschr Elektrochem*., 10(35):664-668. XI. Hauptversammlung der Deutschen Bunsen-Gesellschaft für angewandte physikalische Chemie vom 12. bis 14. Mai 1904 in Bonn.

Newell, A. and Simon, H.A. (1976). Computer science as empirical inquiry: Symbols and search, *Comm. ACM* 19(3) 113-126.

Niebur, E. (2008). Neuronal cable theory. *Scholarpedia*, **3**(5):2674. http://www.scholarpedia.org/article/Neuronal_cable_theory

Oja, Erkki (1982). Simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology* **15** (3): 267–273

O'Reilly, R. C., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience: Understanding the mind by simulating the brain*. Cambridge, MA: MIT Press

Palade, GE, and Palay, SL (1954). Electron microscope observations of interneuronal and neuromuscular synapses in *Anat. Rec.* 118:335–336.

Palay, SL and Palade, GE (1955). The Fine Structure of Neurons in *J Biophys Biochem Cytol.* 1955 January 25; 1(1): 69–88.

Perrett. D. I., Rolls, E. T., Caan, W. C. (1982). Visual neurons responsive to faces in the monkey temporal cortex. *Experimental Brain Research*, 47, 329-342.

Peirce, C.S. (1897, 1903). Logic as semiotics: The theory of signs. In J. Buchler, ed., *The Philosophical Writings of Peirce* (1955). New York: Dover Books, 98-119.

Pfeiffer, R. Scheier, C. (1999). *Understanding Intelligence*, Cambridge MA: MIT Press.

Piaget, J. (1936). *La naissance de l'intelligence chez l'enfant*, Delachaux Niestlé, Neuchatel, Suisse, 5ième édition, 1966.

Pinker, S. (1997). *How the Mind Works*, New-York: Norton.

Putnam, H. (1965). Brains and Behavior, J. Butler, ed. *Analytical Philosophy.* (Second Series). Oxford: Blackwell.

Post, E.L. (1936). Finite Combinatory Processes - Formulation 1. *Journal of Symbolic Logic*, 1, 103-105

Post, E.L. (1943). Formal Reductions of the General Combinatorial Decision Problem. *American Journal of Mathematics*, 65, 197-215

Rall W. (1957). Membrane time constant of motoneurons. *Science* **126**:454.

Rall W. (1959). Branching dendritic trees and motoneuron membrane resistivity. *Exp. Neurol.* **1**:491-527.

Rall W. (1960). Membrane potential transients and membrane time constant of motoneurons. *Exp. Neurol.* **2**:503-532.

Rall W. (2009). Rall model. *Scholarpedia*, **4**(4):1369.

Ramon y Cajal, S. (1909, 1911). *Histologie du système nerveux de l'homme des vertébrés*. Paris: Maloine.

Ramon y Cajal, S. (1995). *Histology of the nervous system of man and vertebrates*. New York: Oxford University Press.

Rosenblatt, F. (1957). The Perceptron--a perceiving and recognizing automaton. Report 85-460-1, Cornell Aeronautical Laboratory

Rosenblatt, Frank (1962). *Principles of neurodynamics,* New York: Spartan

Rosenstein, M.T., and Cohen, P.R. (1998). Symbol Grounding with Delay Coordinates, *AAAI Technical Report WS-98-06, The Grounding of Word Meaning: Data and Models, 20-21.*

Rospars, J.-P. and Lánský, P. (1993). Stochastic model neuron without resetting of dendritic potential. Application to the olfactory system, *Biol. Cybern.* 69:283-294.

Rumelhart, D. E., McClelland, J. L., & the PDP Research Group (Eds.). (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. (2 Volumes) Vol. 1: Foundation, Vol. 2: Psychological and Biological Models*. Cambridge, MA: MIT Press.

Rumelhart, D.E.; Hinton, G.E., Williams, R.J. (1986). Learning representations by back-propagating errors. *Nature* **323** (6088): 533–536.

Saussure, F. de (1913). *Cours de linguistique générale*, Paris, éd. Payot, 1995

Searle, J. R. (1980). Minds, brains and programs. *Behavioral and Brain Sciences* 3: 417-457.

Sejnowski, T.J. (1977). Statistical constraints on synaptic plasticity, *J. Theor. Biol.*, 69:385–389.

Shannon, C.E. (1948). A Mathematical Theory of Communication in *The Bell System Technical Journal,* Vol. XXVII, No. 3, July 1948

Sherrington, C.S. (1906). *The integrative action of the nervous system*. New Haven, CT: Yale University Press.

Shouval, H.Z. (2007). Models of synaptic plasticity, in Scholarpedia, 2(7):1605. Available: http://www.scholarpedia.org/article/Models_of_synaptic_plasticity

Sjöström, P.J., Rancz, E.A., Roth, A., and Häusser, M. (2008). Dendritic Excitability and Synaptic Plasticity. *Physiological Reviews* 88, 769-840.

Sjöström, J. and Gerstner, W. (2010). Spike-timing dependent plasticity. *Scholarpedia*, 5(2):1362.

Skinner, F.K. (2006). Conductance-based models. *Scholarpedia*, **1**(11):1408.

Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences* 11, 1–74.

Song, S., Miller, K.D., and Abbott, L.F. (2000). Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. *Nat. Neurosci.* 3, 919–926.

Sun, R. (2000). Symbol Grounding: A New Look at an Old Idea, *Philosophical Psychology*, 13, 149-172.

Steels, L. and Brooks, R.A. (eds) (1995). *The Artificial Life Route To Artificial Intelligence: Building Embodied, Situated Agents*, Lawrence Erlbaum, Hillsdale, NJ

Stein, R. B. (1965). A theoretical analysis of neuronal variability, *Biophys J.* 5(2): 173–194.

Stent, G.S. (1973). A physiological mechanism for Hebb's postulate of learning, *Proc. Natl. Acad. Sci.* U. S. A. 70: 997-1001.

Steiner, P. (2005). Introduction cognitivisme sciences cognitives. *Labyrinthe*, 20(1), 13-39.

Thorpe, S. T. Imbert, M. (1989). Biological constraints on connectionist modelling. In *Connectionism in perspective*, Pfeifer, R., Schreter, Z., Fogelman-Soulié, F. Steels, L. (Eds.) pp. 63-92. Amsterdam: Elsevier, North Holland.

Thompson, E. (2007). *Mind in Life: Biology, Phenomenology, and the Sciences of Mind.* Cambridge, MA: The Belknap Press of Harvard University Press.

Trappenberg, T.P. (2002). *Fundamentals of Computational Neuroscience,* Oxford University Press

Turing, A.M. (1936). On Computable Numbers, with an Application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, series 2, 42 (1936-37), 230-265

Turing, A.M. (1947). Lecture to the London Mathematical Society on 20 February 1947. In Carpenter, B.E., Doran, R.W. (eds), 1986, *A.M. Turing's ACE Report of 1946 and Other Papers*. Cambridge, Mass.: MIT Press

Turing, A.M. (1950). Computing machinery and intelligence. *Mind*, **59**, 433-460.

van Gelder, T. (1996). Dynamics and Cognition in J. Haugeland (ed) *Mind Design II*, MIT Press, 1997, chapter 16.

Varshavskaya, P. (2002). Behavior-based Early Language Development on a Humanoid Robot, in C.G. Prince, Y. Demiris, Y. Marom, H. Kozima, and C. Balkenius, eds, *Second International Workshop on Epigenetic Robotics: Modelling Cognitive Development in Robotic Systems* (Edimburgh, Scotland), 149-158.

Vogt, P. (2002a). Anchoring Symbols to Sensorimotor Control, *Proceedings of Belgians/Netherlands Artificial Intelligence Conference BNAIC02.* http://cogprints.org/3060/1/bnaic2002.pdf 2012.04.17

Vogt, P. (2002b). The Physical Symbol Grounding Problem, *Cognitive Systems Research, 3, 429-457.*

von Foerster, H. (1974). *Cybernetics of Cybernetics*, Urbana Illinois: University of Illinois

Wittenberg, GM, and Wang, SS. (2006). Malleability of spike-timing- dependent plasticity at the CA3-CA1 synapse. *J. Neurosci.* 26, 6610–6617.