Departamento de Automática, Ingeniería Electrónica e
Informática Industrial

Escuela Técnica Superior de Ingenieros Industriales

A New Foundation for Representation in Cognitive and Brain Science: Category Theory and the Hippocampus

Autor

Jaime David Gómez Ramírez

Directores

Ricardo Sanz Bravo Ramón Galán López

2010

To my wife and my son, with love.

Resumen

Resumen en Español

La tesis presenta un nuevo marco teórico parta estudiar la estructura de sistemas complejos basado en la teoría de categorías. El concepto de estructura es clave en la tesis y se explora formalmente. La estructura no reside tanto en los objetos sino en las relaciones con otros objetos. Las relaciones en la teoría de categorías vienen definidas como morfismos o flechas y la estructura se representa con el diagrama de una categoría.

La teoría de categorías es un reciente campo dentro del álgebra cuyo objetivo es capturar la estructura que prevalece cuando se hacen corresponder distintos objetos matemáticos. Dicha teoría se aplica en la tesis para describir formas generales de equivalencias entre categorías matemáticas y categorías pertenecientes al cerebro y psicológicas, es decir emergentes de la estructura neuronal.

La tesis demuestra que la estructura del cerebro puede estudiarse convenientemente con dicha álgebra. En el hipocampo se encuentran las “place cells” o neuronas que se disparan cuando un agente visita un lugar, y las “grid cells”, células que codifican la métrica del espacio en que el agente se encuentra.

El capítulo 8 detalla como la relación entre estos dos tipos de neuronas puede ser explicada formalmente mediante el concepto de “colimit”. En el capítulo 9 se presenta una teoría de la memoría declarativa y se establece la relación entre la memoria semántica y declarativa basads en los conceptos de “pullback”, “product” y “colimit”.

Summary in English

The thesis establishes correspondences between mathematical structures and brain structures. The virtue of such a correspondence is that it makes available the powerful tools of the latter for the deduction of hypotheses for structure and function in neuropsychology. Such an approach is relatively free from the vagaries of purely verbal reasoning and can offer novel insights into the intrinsic nature of cognitive phenomena.

It is unreasonable to think that purely linear mathematics can be the killer tool to model the complex interactions that take place in the brain. We may indeed, need a new mathematical language for describing brain activity.

It sets the agenda of category theory as the appropriate methodology that provides the necessary theoretical framework in order to understand the structure of complex systems, like the brain, in mathematical terms. Although category theory at first sight may seem too pure and universal, in contrast with the spurious biological realm, where the particular prevails over the universal; it may lead to a new and deeper insight into the structure and the representational power of the brain. The thesis paves the way for a more synthetic methodology in cognitive science, the scale free dynamics hypothesis is studied with a new "categorical" light.

In addition, it provides a theory of hippocampus structure and function based on category theory. In particular, it demonstrates that the co-operation of the grid fields may give rise to a "colimit" which is a place field.

Acknowledgements

First, I would like to thank and acknowledge the academics in my field, who had a great influence upon me and my work; especially Ramon Galán, who took me under his wing at a very special moment in my life; and Ricardo Sanz who I consider my mentor and magister, and through whom I learned so much, within the course our innumerable inspired conversations. I will always be grateful to him for his guidance and wisdom, which sparked a lifelong interest and love of knowledge in me.

Also a very special thanks to my friend and colleague, Manuel G. Bedia, who has been my Virgil, in the final phase of my research, guiding me toward the end of this long process. I want to thank you to Ignacio, Carlos and Iñaki for their genuine camaraderie.

I would also like to dedicate this work to my parents and my brothers and sisters, and especially to my beloved wife Sarah and my precious son Bertrand. My wife has been a fundamental pillar for me through out this process, and without her support and patience this work could not be possible. C’est pour toi!

Preface

> Home is where one starts from. As we grow older The world becomes stranger, the pattern more complicated Of dead and living. Not the intense moment Isolated, with no before and after, But a lifetime burning in every moment And not the lifetime of one man only But of old stones that cannot be deciphered. T.S. Elliot [1].

Manichaeanism, which explains facts based on two opposite poles, is a habitual practice in politics and the media today. With this kind of explanation, we always find there two clearly separated sides, one is good and the other bad; of course the manichean placed himself on the one good side.

Following Jeannerod [2], this polemic way of seeing the world is hardwired in the human brain, so as this thesis has been entirely conceived of and written by a human -no hermeneutical algorithm à la Sokal here- I ask everyone to forgive me for starting with such dichotomic differentiation to make my point.

In a manichaean stance, there are two ways of modeling cognitive systems. One is meant to demonstrate that we understand some particular aspects of the system, and the other tries to discover new structural and functional relationships in the system. The former allows numerical simulation while the latter attempts to capture the organising principles that mediates the behavior of the system. In conclusion the former is data-driven while the second is driven by the creative act of the modeler and ultimately endorsed by the experimental data.

This thesis is an attempt to help in the transition, from a purely descriptive biology to a strongly mathematical one in his methods but also conceptually. Such a transformation has just started and seemingly will keep going on.

The moon-earth system can give us a nice historical introduction to the problem. Ptolomeic´s model was able to predict the position of the planets, so it was a good model in simulating data, which were scarce. The trouble came when the data set became abundant; the experimentalist Tycho Brahe comes to mind. In this new scenario, the Ptolomeic model, in order to keep pace with the new influx of data, became more and more complex, even baroque.

Copernicus and Kepler´s revolution means the substitution of the Ptolomeic model that was unmanageable by a less complex model.

Newton later came up with the law of universal attraction of two bodies in mathematics; this achievement is well known to be based on what was already known on the moon earth position relationships.

The conclusion we can take from this is as follows. Ptolomeic model was able to deal with data, until a certain point, but it was necessary for Copernicus and especially Newton to strive for mathematical structure, to discover the underlying principles or laws that govern the attraction of any two bodies, moon and earth included. Thus, Newton happily solved the two-body problem analytically, but if we include just a single body to our system which now has three bodies, finding a solution becomes impossible! This was true for three hundred years, until Poincaré's breakthrough which consisted in re-thinking the problem with the question, "Is the solar system *stable* for ever?"

Focusing on stability, a systemic property that can not be found in the individual planets but in the system as a whole, allowed Poincaré to succeed where many brilliant minds before him had failed. To predict the position at any time, of any of the three celestial bodies; sun, moon and earth, was possible only when looking into the organisational principles of the system, rather than in the correlation between the properties of its subsystems. This case is a brilliant antecedent of what will come much later, systems biology, which is the science that studies how functional properties arise in the interaction of the components, and in doing so, it reveals the *organisational principles* of biological networks.

In biology, "law" sounds too strong, biologists prefer use *habits* or *organisational principles.* Thus in biology we content ourselves with the search for the *organisational principles* of a system and we leave the search of big words like *laws* to the natural science, for example- Newton laws of dynamics or Maxwell's laws of electromagnetic field. There is a radical difference between biology and physics; while physics deals with stable objects (of course there are exceptions like the radioactive substances), in biology, systems are always getting ready to change to a new state. Biological systems have a certain purposiveness, an arrow in the temporal direction.

In physics, if we represent the transition between two states with the mapping $A \longrightarrow B$, the uncertainty is usually in the final state B, particular boundary conditions for A are imposed in order to get an unique solution for B. On the other hand, in biology, the uncertainty is in the arrow itself. Thus, we can predict that from A that the system will evolve to B, but there is a multiplicity of ways to do so. For example, the output of gene expression, the proteins, may vary substantially depending on the different cells in which they occur, even if they are in the same tissue. In precise mathematical terms, gene expression is a stochastic process.

Degeneracy and redundancy are two characteristics of biological systems that are at the core of the scarcity of laws in biology. For Sydney Brenner, a pioneer of gene sequencing, the prime intellectual task of the future lies in constructing an appropriate theoretical framework for biology.

D'Arcy Thompson, Nicolai Rashevsky or Alan Turing are notable pioneers in the building of such a missing theoretical framework that could put biology and physics at a similar epistemic level. The elegant prose and harmonious sense of aesthetic beauty of the drawings in Thompson's *On growth and form* [3], should not distract us from the fact that his examples only apply to objects with symmetries and clear morphological regularities. Thus, his theory can not be universal. For example the three-dimensional structure of proteins may have a total lack of symmetries. This was discovered in 1958, 51 years after the first edition of Thompson's book, by John Kendrew [4] who determined the first atomic structure of the protein myoglobin.

Rashevsky's relational biology focused on the search of general principles in biology. Rashevsky and later Rosen, attempt to re-think the theoretical biology in mathematical terms, but in practice they failed, because the experimentalists simply did not know what to do with the mathematical abstractions[5].

Turing's search for recursive patterns of organisation is in need of a new paradigm more powerful than the classical computational models that he himself made possible. Super-Turing computation, nanocomputation or analog computing are already on the market [6].

However, we must be prudent when talking about paradigm shifts or Kuhnean revolutions. Mayr suggests that maybe they are not possible in biology, pointing out at the lack of universal laws as the reason [7].

The precise and colourful three dimensional structure of the proteins now obtained in laboratories, does not tell us its function inside the cell. In biology, function is not used as it is in the natural sciences. For example, in engineering the function of a system does not pose a problem because it is the engineer who gives to the system a particular function; the function of an end-effector in a manipulator arm is not to grab things, it is the engineer who confers such power to an object with properties that more or less make it suitable with that task. On the other hand , in biology function is the result of evolution, the function of myoglobin is to store oxygen in muscle cells and that of protease to decompose other proteins.

Francis Crick is known for his discovery of DNA, but this achievement is due in part to his theoretical recognition that in order to study the function of a system, one must first determine its structure. But the structure that Crick had in mind, can not be the morphological configuration that is provided, for example, in the static picture of X-ray crystallography that Kendrew obtained for the myoglobine. If that were the case, the mapping $\textit{Genotype} \longrightarrow \textit{Phenotype}$ would have been already resolved. Of course, we are far from that scenario.

The structure must encompass the dynamic relationship of the components that pre-configure the system as well as its evolving patterns and preferably has to be defined in mathematical terms. The strictly linear causal scheme $\textit{Genotype} \longrightarrow \textit{Phenotype}$ must be replaced by another that encapsulates the multi-level interactions of the system's components and also the changes the environment produces in the system.

Bearing this in mind, we should be able to uncover the structure of the neural system where behavior lies, as we did with the DNA. But this kind of linear reasoning fails because the brain is an infinitely more complex system than a DNA molecule. It might be emphasized that a DNA molecule has to be simple, it is the origin of life, while the brain is the extremely elaborated result of a long history of changes and evolution. The devil is in every detail. But we can always get rid of the devil, by choosing a system not so overwhelmingly complex as the human brain. For example, a simple organism like the soil worm C.elegans, has been cracked down and a complete map of the mere 900 cells that this organisms has, of which 302 are nerve cells, is available to the scientific community.

In science, many problems are still unsolved, otherwise there would not be scientists, but technicians. And many questions that we want to solve today will never have an answer, and not necessarily because they are too complex to be solved, but because they will just disappear from irrelevance.

For example, whether differentiation was a state or a process kept biologists busy for many years. Today it seems completely foolish to pose a question in those terms. There is a major question being posed today that will likely leave us without an answer, this thesis tries in to shed some light on this. The question is as follows: given *i)* the complexity of of the brain; *ii)* that this complexity has been the result of a very long period of evolution, starting from very simple means; *iii)* the major anatomical features of the brain are known, the nervous system is a complex system of neurons that transduce, transmit and process information; how are we going to put this together and understand the brain and its cognitive powers?

This is a too wide question that can not be solved here, we need to narrow the focus. How does the neural system encode mental objects, in particular spatial representation and memory? This is the question that this thesis addresses, and in doing so, it borrows concepts form the mathematical theory of categories. In the end, brain function and behaviour depend upon the interaction of elements (cells, networks, brains). The neural and cognitive sciences, deal with complex components (assemblies of cells, brains) and out of necessity will follow a mathematical strand.

Science moves in little steps, but also makes its progress with revolutionary discoveries and concepts that sweep away whole and entire edifices of thinking and replace them with new theories that explain more with less. However, there is a constant in this march, the strive for mathematisation and unification.

The dynamic nature of biological organisms has to be understood, not in the well-known evolutionary sense of organisms changing over time, but in terms of the persistence of patterns at the different levels of description, from molecular transcriptional networks up to cells, tissues, organisms and ecosystems. Even the socio-economical system (this pair of words is what politics has been transformed into) has an invariance of patterns to be discovered by scientists and exploited for the common good.

When Schöringer wrote in 1944 *What's life?* [8], the biology at that time could be explained without mathematics. He argued that this was not because biology was simpler than physics or chemistry, but quite contrary to this, biology was too complex to be mathematised.

In 2010, it seems foolish to deal with such complexity without mathematics. The new science of complex systems has exploded in the last 50 years. Network science, non linear dynamics, chaos, statistical physics... are now part of the academic curriculum.

Before I conclude, I have to confess a final weakness here, which is that of being a dilettante. But in my defense, I can say that knowing too much about a topic can sometimes be an impediment to doing creative work, discouraging us to take the adventurous hidden road rather than the well signalled route that the majority is taking.

Contents

1 Introduction 1
- 1.1 Introduction 1
- 1.2 YAPS! Yet another paradigm shift? 2
- 1.3 Mathematics as a language and as a modelling tool 4
- 1.4 Plan of the thesis 7

2 State of the Art: Mathematical approaches in brain science 9
- 2.1 Introduction 9
- 2.2 Brain state 11
  - 2.2.1 The search for the meaningful brain states 13
- 2.3 Organising the Mathematical Framework 14
- 2.4 Modeling neurons 15
  - 2.4.1 Examples of models of neurons 16
- 2.5 Modeling populations of neurons 18
  - 2.5.1 Artificial Neural Networks 19
- 2.6 Neurodynamics 19
  - 2.6.1 Neurodynamics, classical dynamics and startistichal mechanics 20
  - 2.6.2 Biological synchronization: the theory of coupled oscillators 23
  - 2.6.3 Neural Field models and neural masses 26
- 2.7 Modeling the whole brain 27
  - 2.7.1 Brain processes as critical state transitions across levels of organisation 28
  - 2.7.2 Synergetics 30
  - 2.7.3 Dynamic geometry 31
  - 2.7.4 Discrete Approaches to Neural Phase Transitions 31

3 The Categorical Imperative: Category theory in Cognitive and Brain Science 33
- 3.1 Introduction 33

3.2 Category Theory 34
  3.2.1 Examples of categories 38
  3.2.2 Definition of some key concepts in the theory of categories 39
3.3 The cat-level avenue 45
3.4 Applications of Category Theory in Cognitive and Brain science 48
  3.4.1 The origins: Rosen's (M,R-systems) 48
  3.4.2 Category theory in Perception 49
  3.4.3 Memory Evolutive Neuronal Systems 50
  3.4.4 Category theory in knowledge acquisition and representation 54

4 Elementary principles in cognitive systems modeling 63
4.1 Introduction 63
4.2 On reductionism 64
  4.2.1 What is reductionism? 65
  4.2.2 Two obstacles for reductionism 66
4.3 On formalisation 67
  4.3.1 The limitations of formalisation 68
4.4 Emergence on Systems Modeling 73
  4.4.1 A few notes on complex systems 74
  4.4.2 A few notes on emergent properties 76
4.5 Three Principles for Cognitive Systems modelling 78
  4.5.1 Principle of Locality 79
  4.5.2 Principle of Hierarchy 80
  4.5.3 Principle of Multiplicity 82

5 The shift towards structure 85
5.1 Introduction 85
5.2 Defining structure 86
  5.2.1 The shepherd's tale 88
5.3 Categorizing structured systems 89
  5.3.1 General systems theory 90
  5.3.2 Structured systems 91
5.4 Linking the mathematical structure of the physical world 94
  5.4.1 An example of theoretical hypothesis in biological systems: The Brouwer theorem 94

6 A general framework for representation 99
6.1 Representation as denotation 99
6.2 Mathematical representation 103
  6.2.1 Measurement Theory 105
  6.2.2 Phase space representation 106
6.3 A theory of representation in cognitive systems 108

6.4 Structural commonality in representation 109
  6.4.1 Representation as a relation of similarity between structured entities 110
  6.4.2 Representation as a relation of isomorphism between structured entities 110
  6.4.3 Representation as a relation of homomorphism between structured entities 110
  6.4.4 Representation implies structural similarity: Homomorphism 111
6.5 Theory of representation based on category theory 113

7 Towards a theory of brain structure, function and dynamics 119
7.1 Introduction 120
7.2 The crisis of a paradigm: Historical account 121
  7.2.1 The old paradigm: Modularism-homuncularism 122
  7.2.2 The brain is not an aggregate system 122
  7.2.3 The brain mapping is of form and meaning 123
7.3 Computational neuroscience and mathematical neuroscience 129
7.4 The levels of organisation in the brain 131
  7.4.1 Brain activity and measurement techniques 133
  7.4.2 Linear vs Circular Causality 134
  7.4.3 Granger Causality 135
7.5 A Theoretical Framework for Connectivity in Biological Networks 136
  7.5.1 Defining Brain Connectivity 138
  7.5.2 Characterisation of Network Topology with Graph Theory 142
  7.5.3 Connectivity and Network Dynamics of the brain 144
  7.5.4 Regular vs Complex network topologies 145
7.6 The search for patterns in the brain 148
  7.6.1 A first approach to meaningful patterns in the brain 149
  7.6.2 Meaningful Patterns in the brain 150
  7.6.3 The multilevel hierarchy of meaningful patterns 151
7.7 The dynamics of the meaningful patterns. Neurodynamics 155
  7.7.1 The I/O Newtonian approach in brain dynamics modeling 156
  7.7.2 Neurodynamics 157
  7.7.3 EEG/ECoG recording and theoretical hypothesis 157
7.8 The K-Sets hierarchy model 161
  7.8.1 K0 163
  7.8.2 KI 163
  7.8.3 KII 163
  7.8.4 KIII 164
  7.8.5 More complex models: KIV and KV 164
  7.8.6 Categorizing K-Sets 166

8 A theory of hippocampus structure and function based on Category Theory . . . 169
- 8.1 The nature of spatial representation . . . 169
- 8.2 The hippocampus as a representational device . . . 171
  - 8.2.1 The cognitve map hypothesis . . . 171
  - 8.2.2 The hippocampus: anatomy and connectivity . . . 173
  - 8.2.3 Place cells . . . 174
  - 8.2.4 Grid cells . . . 176
  - 8.2.5 Head cells . . . 179
- 8.3 A theory of brain spatial representation based on category theory . . . 179
  - 8.3.1 The category of neurons . . . 180
  - 8.3.2 The category of metric fields . . . 181
  - 8.3.3 Functor between Neur and Field . . . 182
- 8.4 A new framework for place and grid cells . . . 184
  - 8.4.1 Place field as colimit of grid fields . . . 186

9 From cells to memories. A categorical approach . . . 191
- 9.1 Types of memory . . . 191
- 9.2 A theory of declarative memory (episodic and semantic) based on category theory . . . 193
  - 9.2.1 Categorical product in acquisition of middle point concept in 1D navigation . . . 194
  - 9.2.2 Categorical pullback in acquisition of middle point concept in 2D navigation . . . 195
  - 9.2.3 Pullback and grid cell formation . . . 196

10 Discussion and Future work . . . 197
- 10.1 Future work . . . 199

References . . . 201

List of Figures

1.1 The diagram displays the small part dedicated to the understanding of the principles compared to those tasks related with the obtention and visualisation of data from the measurement techniques [9]. 5

2.1 Particle m_i rotating about the z-axis with the angular velocity w 10
2.2 A phase portrait is a geometric representation of the trajectories of a dynamical system in the phase plane. The figure depicts the phase space of a simple pendulum, each set of initial conditions is represented by a different curve. The x axis is the phase θ and the y axis the velocity $\frac{d\theta}{dt}$ 12
2.3 Measurement techniques, brain hierarchy level and mathematical models are related as the figure shows. An unique and all encompassing mathematical model of the brain can not be expected. 15
2.4 Amplitude (mV) and duration (ms) determine the action potential of a nerve cell. A. shows an idealized action potential, in B. the recordings of action potentials are often distorted compared to the schematic view because of variations in electrophysiological techniques used to make the recording. The membrane potential under the threshold may also affect to oscillatory activity of neighboring neurons. Image from [10]. 21
2.5 Dynamic model using Newton-euler notation or an open kinematic chain of N links. This system presents transcendental, non-linear systems of differential equations that it is now possible to solve numerically. The analytical solution can be found for particular situations, for example when at least three neighbour system of reference have coincident origins 22

2.6 The simplest coupled cell network is the two-cell. The class of differential equations associated to this network is $\dot{x_1} = g(x_1, x_2), \dot{x_2} = g(x_2, x_1)$ where $x_1, x_2 \in \Re^k$ are the state variables of the cells. Note that the system is as simple as to be defined with one single function g. Symmetry in animal gait shows that for quadruped animals, the only possible symmetry class of CPG networks is the 8-cell network shown on the right side of the figure from [11]. 25

3.1 Samuel Eilenberg, left, and Sanders Mac Lean, right, the two creators of category theory 36
3.2 On the left, Alexander Grothendieck, a very remarkable man, Fields medal in 1966, disenchanted with world politics and academic circles, he moved into the woods, maybe somewhere in the Pyrenees. On the right, William Lawvere, the champion of category theory as a foundation for amthematics. 36
3.3 Category **1** 38
3.4 Category **2** 38
3.5 The figure depicts the categorical construction of a physical object, the gray house on the right. The *mentalae* S and T, on the bottom of the figure are concepts that are reified in a neural organisation, in the category **Neur**. The colimit of the objects represent the *mentale* H (the house). Finally there is a functor between the physical object and the neural category **Neur** 47
3.6 The figure describes the evolution of the neural system during its life. For each t of the timescale is Time, we have the category $NEUR_t$ of neurons at t [12]. 51
3.7 The iterative building of cat-neurons binding patterns of cat-neurons of lower levels, models the computation on mental objects, to form more complex ones. It gives explanatory power for the construction of an algebra of mental objects 53
3.8 The theory T0 defines a simple geometry of points and lines which are primitive types and the relationship between them is given by the axiom "Two-points-define-a-line"[13] 55
3.9 The figure depicts the colimit T_2pL of theories T_1,T_2, and T_L. The colimit theory glues the theories T_1 and T_2 along their common sub-theory T_L. For the dual treatment of theory abstraction using limits see also [13] 56
3.10 Information flow in the genetic system as depicted in [14] 58
3.11 The flow of information from DNA to RNA and protein, is represented with the arrows f^{-1} and g. Thus, given the arrows f and h we can detrmine f^{-1} and g 59

4.1 The structure is prior to its formal and the formalisation captures the structure of the system 73
4.2 A colimit K for the base diagram D. For the sake of clarity in the figure the diagram D has three objects $A_i i = 1..3$. The colimit shows that high level interactions .i.e: the colimit K, may arise without being directly induced from lower levels i.e: the diagram D 81
4.3 The manipulator arm shown in the figure is an example of degenerate system if we follow the definition given by Edelman and Tononi in [15]. In robotics we say that the configuration of the robot is redundant 83

5.1 *Structure* is a kind of *form*, while *form* is one possible kind of *mass*. 90
5.2 Rotating the disk gives a continuous endomap $f : D \to D$. The center of the disk is the fixed point x_0 95
5.3 Points N_1 and N_2 in a topological space Θ can be separated by neighbourhoods if there exists a neighbourhood Θ_1 of N_1 and a neighbourhood Θ_1 of N_2 such that $\theta_1 \cap \theta_2 = \emptyset$. 97

6.1 We can infere properties -e.g: symmetry in the limbs, from this simple drawing that may represent a man 100
6.2 The artist Joseph Kosuth's famous work "One and Three Chairs", which a visual expression of Plato's forms. The piece features a physical chair, a photograph of that chair, and the text of a dictionary definition of the word "chair". 101
6.3 The triadic representation conveys to us a new definition of representation as *the process of construction and relationship of signs within a system of signs.* 102
6.4 The figure depicts a thermometer, the observable ψ, that when in contact with a system, the box, returns a real number **R** that is the state of the system. 104
6.5 "Real" space and Phase space of harmonic oscillator. 107
6.6 Under the case explained above, to state that a mental object is the representation of a neural assembly is ill-founded, any subobject would be a representation of the same neural assembly. 117

7.1 Connection specificity -a neuron is connected with some particular neurons- and neural polarisation -within a neuron the signal is unidirectional- was discovered by Cajal. Drawing of [16] 126

7.2 Linear causality can tell us how a reflex response is caused by a conditioned stimulus -basic sensorimotoric loop. Thus, a stimulus excites a sensory neuron, that excites the sensory cortex, that transmits information to the motor neuron, that finally produces the contraction of the muscle. The mechanisms for perception are not strictly linear but circular. Sensory input entails motoric consequences and likely, motor output affects sensory input. In this light, the brain can be seen as an operator between the world and the body. 128

7.3 The Information Processing Hypothesis IPH conjectures that the stimulus is directly represented in the spatial patterns in the brain. Thus, stimulus and representation of stimulus are isomorphic. 129

7.4 The diagram tries to reflect the focus in the numerical matching without structural accuracy. The wide area between the cell physiology until the organism behavior requires theories based on mathematics that matches structurally the phenomenon and this must be done prior to do any computations [17] 131

7.5 The exchange of information between levels in complex systems, like for example the brain cortex, is not linear, rather the interaction is within the level's components, but also upwards and downwards. 132

7.6 The drawing is from [18] and shows five types of connectivity distributions. Uniform: the distribution is constant, nearest neighbor: nodes have the same degree of connectivity or 0, Small world: high clustering and short path, Power law: Poisson distribution and Exponential 145

7.7 Huber and Wiesel study of the visual field claimed that the orientation columns are 50μm width and represent a shift in angle of 10^o. A complete set of columns subserving the 180^o is called hypercolumn. There are two types of ocular dominance columns (L,R) and two kind of orientation columns. For purposes of illustration, Hubel and Wiesel depicted the two sets of columns are orthogonal but there is not evidence of such relationship between them [19]. 153

7.8 Topological patterns of connection described in [20] (a) Convergence. (b) Divergence. (c) Serial transmission. (d) Parallel transmission. (e) Autofeedback. (f) Excitatory positive feedback. (g) Inhibitory positive feedback. (h) Negative feedback. 162

7.9 The figue is from [21], and depicts the relationship between the hippocampal formation and the sensory cortex KIII sets. In the hippocampal side, DG is dentate gyrus, CA1-3 is cornu ammonis and on the olfactory cortical side, PG is periglomerular, OB olfactory bulb, AON anterior olfactory nucleus and PC prepyriform cortex. Chaotic behavior in KIII is the result of the competition between 3 KII components of the hippocampus, CA1, CA2 and CA3. The dentate gyrus (DG) is modelled as K-I set . . . 164
7.10 The KIV consists of three KIII sets, which model the cortical and hippocampal areas, and the mid-line forebrain, respectively. [22] . . . 165
7.11 K-Sets hierarchy designates a topological hierarchy for neural sets. In [23]. . . . 167

8.1 Simplification of the usual schema of connections modules in the hippocampus. The medial temporal lobe (MTL) is influenced by and influences the hippocampus proper. . . . 174
8.2 Drawn of the hippocampus and its circuitry drawn by Ramón y Cajal. The superficial layers of the EC (layers 1,2,3) provide the most prominent input to the hippocampus, and the deep layers (layers 4,5,6) of the EC, receive the most prominent output. . . . 175
8.3 The picture shows the 36 place fields one of each of 36 rat hippocampal CA1 place cells. Thus, as it is shown in the figure the position of the rat is contained in the firing of these cells. The place fields are cone shapes, this means that that the firing rates increases irrespective of the direction from which the rat arrives. . . . 176
8.4 Grid map of a rat running on a linear track after 10 min [24]. . . 176
8.5 Firing rate vs. head direction plot that shows the preference of this cell around the 180° [25]. . . . 180
8.6 (X, d) is a metric space where X is the set of place cells in the hippocampus and d the euclidean metric distance, (Y, e) is a metric space in the bidimensional plane with identical distance $e = d$. The mapping between the metric spaces $f : X \to Y$ preserves the distances if $e(f(x_1), f(x_2)) = d(x_1, x_2)$. f is said to be an isometry and is immediately a monomorphism (Demo: $x_1 \neq x_2, e(f(x_1), f(x_2)) = d(x_1, x_2) \neq 0 \Rightarrow f(x_1) \neq f(x_2)$). Then an isometry that is an epimorphism is an isomorphism . . 181

8.7 The family of objects A_1, A_2, A_3, A_4 has both a colimit $\mathbf{cP}$ and a coproduct $\coprod_i A_i$. The coproduct is linked by s to the colimit. The link s express the transit from the coproduct to the colimit and embodies the symmetry breaking in the relationship between the family of objects A_i and the colimit . 185
8.8 The figure depicts a colimit where (4) acts as the place field of a place cell(6) in the hippocampus. The colimit is produced by several grid fields (one grid field(1) is produced by one grid cell(5)). (3) is a Functor between the colimit (4) and the physical location (8). (2) are connections among grid fields (1), generating produces the tesellation of the arena (7). 188

9.1 $\mathbf{W_A}$ and $\mathbf{W_B}$ are the two walls that the rat will meet when moving in a single track maze. After reaching both walls, the animal would develop the concept of middle point $\mathbf{P}$ 196

List of Tables

8.1 Place cells and Grid cells, similarities and differences 178

Chapter 1
Introduction

In what follows I address the following two questions. Why has biology not yet attained the same status of physics in terms of rigour and predictability? And in the domain of neuroscience, and why do we still lack a unified theory of cognition? Indeed, we have not even begun to understand the mechanisms that govern critical features of brain function, for example, learning or memory. Walter J. Freeman has addressed this succinctly stating that “Brain science is still in adolescence [26]”. We must stay far from far overarching responses; instead we must setup the basis of a new way of understanding the biological organisation. A conceptual breakthrough in biology is imperative; as a result, a new and deeper insight to the organisational principles of living systems will emerge. The mathematical articulation of a theoretical framework will encourage a new way of thinking in life sciences. The so called medicine based on evidence or the anecdotal accumulation of data in neuroscience will benefit from the new paradigm.

1.1 Introduction

Biology in the 21st century will be for Mathematics, what Physics was in 19th and 20th centuries. This is a well accepted belief among scientists with mathematical training. Some physicists can go even further and deplore the pre-Newtonian stage, in which biology and in particular neuroscience find themselves. Biology may be perceived as descriptive and qualitative, in opposition to physics, which is quantitative and explanatory, thanks to the use of mathematics. Thus, the language of physics is mathematics. Galileo put it memorably, ”Nature’s great book is written in mathematical language”.

However, the postulate that states that biology is descriptive and qualitative, while physics is mathematical and quantitative is a dichotomy too simplistic to be true (as the dichotomies usually are). Undoubtedly, biology is quantitative in its models and results. Indeed, maybe too much. The pace at which new experiments and data are being produced in the life sciences is overwhelming for the specialist and disconcerting for the non-specialist. We simply can not continue throwing more complexity into a literature that is already almost illegible. This can be only detrimental.

It is easily noticeable that in the top scientific journals, for example *Nature* or *Science*, though they are not devoted to any particular field, the articles dealing with biology, specifically molecular biology, make up a huge majority. The very dissimilar rate of production that these publications display between biology and physics, for example, is such that one starts wondering if the non-life science community is getting something done? But the actual reason for this is rooted in the conceptual poorness of biology compared to physics. Physics is built on formal theories embedded in universal laws, rendering particulars and details unnecessary; while in Biology, particular cases are relevant. Here I share the view of the Oxford biologist Denis Noble, "the next stage in the development of biological science will be revolutionary in its conceptual foundations and strongly mathematical in its methods [27]".

However, we must admit that meeting the expectations of the two traditions, the mathematical-logical and empirical-biological, is not a realistic goal, at least in the short term. The goal we are seeking is not to find *the* theory that explains everything, this ideal is rooted in the naive assumption that there can be an unique valid view of a biological system.

This thesis is a search of better ways of understanding brain structure, dynamics and function, introducing new tools and concepts that will aim to the construction of better models of cognition.

1.2 YAPS! Yet another paradigm shift?

Understanding that biology is rich in proper names but poor in concepts is key. Indeed, biology textbooks abound with accurate descriptions of thousands of molecules, organelles and cells, each with their own names and purposes. In contrast, the same books give little or no account for the laws or general principles that govern the interactions between those molecules, organelles or cells.

Mendel's law of inheritance is one, if not the only law, expressed in formal terms in biology. It ought to be noted that Mendel law is not universal but general, because it only applies to eukaryotic and sexuate organisms. This rarity of universal laws in biology marks a clear distinction with physics and poses questions that need to be addressed.

Biology has historically focused on experiments, in order to a posteriori deduce theories and models underlying the phenomena under study; in physics, it is the theory that has lead its development and progress. For example, Newton built a model that captured the relationship between the moon and the earth in the movement of the former around the latter. Thus, through an abstraction Newton was able to understand the structure of the natural phenomena, i.e. the movement of bodies around each other.

It is important to remark here that Newton's theoretical work precedes the recollection of data. As a matter of fact, he built the model not to fit the data; rather the mathematical abstraction that he constructed made it possible to understand the phenomenon, capturing its structure and putting it in mathematical terms[28].

The renowned evolutionary biologist, Ernst Mayr, makes an important observation in [29]. In biology there are not scientific revolutions; at least not in the Kuhnean interpretation of the irruption of quantum mechanics and general relativity in physics. In order to have revolutions or paradigm shifts we need first to have universal laws to be surpassed by new ones, and in biology we lack the kind of the universal laws that exist in physics. Thus, the criterion of falsability of Popper, would apply to physics but not to biology.

For Sydney Brenner, a pioneer of gene sequencing, the prime intellectual task of the future lies in constructing an appropriate theoretical framework for biology[30]. The task is ambitious and precaution is in order; we must first acknowledge the obstacles that we are facing here.

Each scientific community uses its own language, the same word may mean very different things depending on which domain of discourse in which we are. By way of illustration, some examples of words with different meanings between biologists, mathematicians and physicists follow:

i to *differentiate:* in physics is to find the slope of a function; in biology means the changing of function of a cell.

ii to *divide:* in mathematics is the operation that gives us the ratio of two numbers. For a biologist, division is the process of duplication by which a cell replicates its content and split into two, the characteristics transmitted by the mother cell are preserved in the duplicated cells. Division is a key process in biology because all cells, with the exceptions of neurons and cardiac cells, duplicate.

iii *function:* in mathematics -set theory- a function is a relation between a domain and a codomain, each element of the original set or domain corresponds only one element of the target set or codomain. The most extended notation for function is $f : X \rightarrow Y$ where f is a function, X the domain and Y the codomain. In biology, molecules have associated functions, for example, the function of the ribosome is to synthetize proteins, or that of a protease is to decompose other proteins. This interpretation does not apply in physics. In fact, a superconductor does not have a function, but properties. The function of a superconductor is given by the engineer in his strive to operate in optimal conditions of conductancy. Using the words of John Hopfield, the creator of the associative neural networks, "The word function doesn't exist in physics, but physicists are going to have to learn about it"[31]

iv *cell* in mathematics, topology, means edges that connect vertices. The most common use of cell comes from physiology, cells are the smallest living

> things and they have all the properties of life, including reproduction, response to environmental signals, a need for energy, and the release of waste products.

It might be emphasized that having identical words that express dissimilar concepts, depending on the context in which the word being used, is not intrinsically bad nor good. The trouble is, that the scientific disciplines, when missing a common framework, can hardly understand each other. For Kuhn, such framework is possible within the context of evolutionary epistemology of theories [32],[33].

In contemporary academic science, "interdisciplinary" has become a buzz word. Nevertheless the current traditional academic culture does not reward interdisciplinary, rather it is often viewed with suspicion or disdain. Perhaps not without reason, expanding interdisciplinary science prior to build a common epistemological culture, is not the strategy that will drive us to the obtention of important scientific results.

It is important to note here that the paradigm shift or scientific revolution as conceived by Kuhn can not be applied to biology. The paradigm shifts require that laws be challenged, like in the removal of Ptolemaic laws by Copernicus, and biology lacks universal laws.

The two most important revolutions in biology are two, Darwin's theory of evolution and molecular biology outbreak in the 1950's. The former is not a scientific theory in Popper's view, but metascientific; the fittest is only known after it has survived. Molecular biology, on the other hand, has produced dogmas not laws, for example the so-called central dogma in moder biology which forbids the passage of information from the organism and the environment to the DNA [34]. This dogma has been challenged by Stanley Prusiner's work on mad cow disease, in which he demonstrated that the flow of information may go directly form protein to protein. This is contrary to the central dogma in Molecular Biology, which states the unidirectionality in the information flow sequence [35]:

$$DNA \rightarrow ARN \rightarrow protein \tag{1.1}$$

1.3 Mathematics as a language and as a modelling tool

In physics, as Galileo pointed out, the language of nature is a mathematical one. The concepts themselves are mathematical objects and therefore are meaningful only when written in mathematical formulae. For example, the concept of field can not be understood outside the Maxwell equations. On the other hand, in biology, natural language -e.g: plain English, is predominant and the equations marginal. In a biological context, mathematics is "just" a tool to model biological phenomena.

This is a subtle but very significant difference in the use of mathematics that usually goes unnoticed: while the language of physics is mathematics, in biology mathematics is just a tool that may be used to build quantitative models.

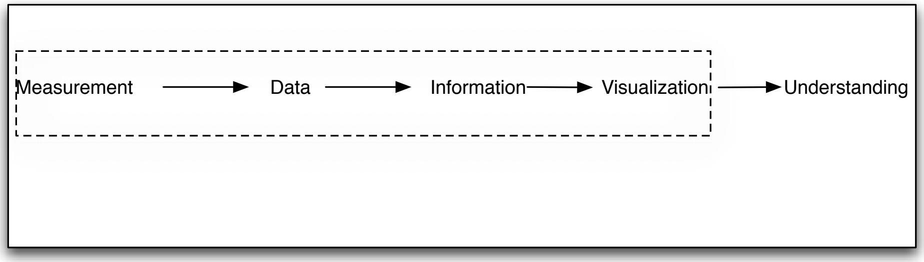

Fig. 1.1 The diagram displays the small part dedicated to the understanding of the principles compared to those tasks related with the obtention and visualisation of data from the measurement techniques [9].

Biology is considered "impure" while mathematics is "pure". This is a common misunderstanding rooted in the fact that in biology, there is a large number of special cases with no analytical solution, because they are too complex or too non-linear.

Formalisation allow us to discover the general principle mechanisms. A mathematical formulation of life itself is open to philosophical and theoretical examinations, suggesting new challenges in both directions, from biology to mathematics and the other way around.

Already in the first paragraph of the popular book "What is life?", Schrödinger announced that mathematics would hardly be utilized in biology. The reason for this, he goes on, was not that the subject was simple enough to be explained without mathematics, but rather that it was much too involved to be fully accessible to mathematics [8].

Thus, life was too complex to be described in mathematical terms. This idea, since the book was written in 1944, has prevailed as a credo that does not need to be discussed, especially if one wants to avoid being accused of reductionism with a flawed vision about what life is.

Rudolf Carnap, the Vienna Circle's philosopher, was the champion of physicalism whose main tenet is that all sciences, including human sciences, were reducible to physics, because every sentence of psychology, biology ... may be formulated in physical language [36].

The extreme reductionism of Carnap is untenable. It is true that the universal laws of physics necessarily apply to living systems, thus physics and chemistry, of course, are valid in biology. But this only proves that nature is ordered and probably hierarchically structured, that being so, a predicate of level i is irrelevant, though true, at superior the levels $i+1, i+2\ldots$ For example, the chemical composition of the ink with which the book is printed, is completely irrelevant for the literary critic, or the ideal gas equation is of not relevance for an environmental biologist. In a similar fashion, as physicists do, system biologists have started asking if there are general physical principles at work in biology.

This is the engineering approach, which does not commit itself with any particular level of description or set of details, rather it focuses in the necessary functional level of description that aims to figure out the general principles that explain why the system behaves in the way it is.

We need to understand the general principles that rule biological organisation. In physics or chemistry, given a system, one may conjecture about its structure. It should be possible to make predictions, experimentally tested, about the future states of the system. Thus, the structure conjectured for the system is refutable as Popper required for a theory in order to be scientific. But in biology, for example when we study proteins, knowing the 3D structure of the protein, is clearly not enough to understand its role played inside the cell. If we modify an enzyme expecting to block its function, another protein could take that role [37]. This is due to redundancy, one of the key organisational principles in biological systems. This principle, together with degeneracy will be conveniently explored in chapter 4.

The historical dimension of biological systems need to be take into account. In biology, each phenomenon is the result of multiple causes, thus history is relevant, unless we summon a Laplacian world, in which the current state of the system completely define its future.

I agree with the oncologist Robert Gatenby when he says,

> Existing mathematical models may not be entirely correct. But they do represent the necessary next step beyond simple verbal reasoning and linear intuition. As in physics, understanding the complex, non-linear systems in cancer biology will require ongoing interdisciplinary, interactive research in which mathematical models, informed by extant data and continuously revised by new information, guide experimental design and interpretation.[38]

The same can be said about mind and brain. The cognitive and brain sciences must necessarily follow the systemic, integrative and mathematical based undertaking referred above. What is needed now in cognitive and brain sciences is to be able to be as successful and rigorous in putting together the pieces that reductionists were able to separate. Such an achievement requires an understanding of the brain at system level.

The time is ripe to take a step back from the details and try to see the big picture.

1.4 Plan of the thesis

I now proceed to give a chapter-by-chapter overview of the topics covered in this thesis.

In this initial chapter 1 we explored the question, why has biology not yet attained the same status of physics in terms of rigour and predictability? and whether we can expect a Kuhnean revolution in brain and cognitive sciences.

In chapter 2 we describe different mathematical tools to model brain structure and dynamics. The microscopic, mesoscopic and macroscopic levels are covered in this review. The spirit of the chapter is to provide the reader with a mathematical toolbox which is able to deal with the non-linearity in brain dynamics.

In chapter 3, category theory is introduced. I will provide a formal introduction to this mathematical theory, including applications in the domain of brain and cognitive science. A new and original theoretical framework in which different tools and approaches can be used in conjunction with one another, will naturally arise in this chapter.

Chapter 4 studies crucial philosophical topics like formalisation, reductionism and emergence from a transversal and systemic perspective. It paves the way for the formulation of *structure* and *representation* in the next two chapters.

Chapter 5 is devoted to *structured systems*. The concept of structure is rigorously explained here. In addition, I use category theory to introduce the concept of *cat-system*, which is a categorical formulation for *structured system*.

In chapter 6, I present a general framework for *representation* based on category theory. The idea is to bring this mathematical formalism into the domain of neural representation of physical spaces, setting the basis for a theory of mental representation, able to relate empirical findings, uniting them into a sound theoretical corpus.

Up to Chapter 6, we have treated the key issues of this thesis, i.e. structure and representation, *in a general setting*. In the last part, the focus is shifted to the brain, in particular the hippocampus. The structure of the hippocampus, and the mechanisms by which mental representation and memories arise from its interactions will be conveniently covered.

Chapter 7 explores two main issues. The first concern is to point out the flawed use of mathematical and philosophical concepts, in our attempts to understand how the brain works. More specifically, I draw attention to the restricted, yet inconclusive quarrel between the different schools of brain organisation and functioning. This chapter, based on empirical evidence, examines the scale-free nature of brain connectivity and activity, and provides a theoretical framework, which is able to model the dynamic nature of biological systems. The second concern is the introduction in brain theory of mathematical descriptions of key concepts, such as pattern or structure, which are at the core of the natural sciences, and should occupy a preeminent position in the neural science.

Chapter 8 describes the effect in injecting the concepts of coproduct and colimit from category theory into the problem of place cell formation in the hippocampus. The hypothesis that will be explored here is that the colimit is the mathematical structure that allow us to encode the emergence of mental concepts related with space such as position, orientation or distance. The main point is that this theoretical postulate will shed light on how populations of grid cells contribute to the formation of one place cell.

In 9 it is described a mathematical formalism based on the theory of categories for modeling declarative memory. The categorical concepts of product and pullback will be explored and mapped onto psychological structures related with semantic memory.

Finally, chapter 10 presents the conclusions and some future works. It ought to be remarked that although this thesis may sometimes embrace oversimplifications and wild speculations. However it is worth further investigation, and its relevance to understanding brain structure and function should be closely examined.

Chapter 2

State of the Art: Mathematical approaches in brain science

As in any other experimental science, in neuroscience, we build models in order to attain a structural match with the natural phenomenon that we want to understand. Models, of course, are made up of data.

In neuroscience, as neuronal mechanisms and processes are not directly observable, models become essential. Indeed, neuronal function is explored through inferences on models or on their parameters, rather than automatically deduced from direct observation. In neuroscience, models not only capture phenomena, they also generate data [39].

2.1 Introduction

I will start this chapter by trying to demonstrate the inescapable use of mathematical tools for the modeling of the physical world (brain included).

It goes without saying that with the following example, I am not reducing the brain to the kind of solid rigidity that is usually depicted in the textbooks of physics. The rationale here is to emphasize the colossal difference having a mathematical model makes in natural science [1].

In particle physics, the angular momentum of a particle with mass m and velocity v, about a rotation axis z at distance r, is given by the formula

$$\mathcal{L}_z = r \times mv.$$

By substituting the angular velocity $v = wr$ where w is the rotational velocity in the axis, one obtains

$$\mathcal{L}_z = r \times mwr = wmr^2.$$

---

[1] I use the term natural science here in the original sense of the German word "Naturwissenschaft", that is, the study of nature by the rigorous and formal investigations as it is prescribed in the scientifc method.

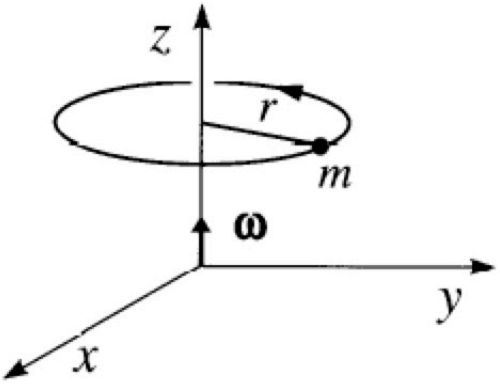

Fig. 2.1 Particle m_i rotating about the z-axis with the angular velocity w

The term mr^2 is called momentum of inertia of mass m in figure 2.1.

$$\mathcal{I} = mr^2.$$

Now, in order to calculate the momentum of inertia of a multi-particle system rotating with identical angular velocity w, we need to sum up the momentum of each particle and the result is the momentum of the system:

$$\mathcal{L}_z = \Sigma_i r_i \times m_i v_i = w \Sigma_i m_i r_i^2$$

Thus, the resulting momentum of inertia of the multi-particle system is now:

$$\mathcal{I} = \Sigma_i m_i r_i^2$$

The question that now arises, is as follows; what happens if we are dealing with an extended continuum system and therefore we can not positively separate its different particles?

In this case, we can not just sum the different particles as we did with the example shown above of particle physics. There are systems in which the particles may not be that easily differentiable, or maybe they are, but to count them all poses problems of computability.

The solution to this deadlock is given by the construction of a mathematical model, only in this case we can licitly assume that we can count all the particles even though we actually do not do it, the trick is that mathematics does this for us.

Accordingly, for a body with a homogeneous mass distribution and density ρ, the differential of the mass is $dm = \rho dV$. The momentum of inertia about a known axis can be calculated by integrating the square of the distance, weighted by the mass density, from a point in the body to the rotation axis:

$$\mathcal{I} = \iiint\limits_{m} mr^2 = \iiint\limits_{V} \rho r^2 dV \tag{2.1}$$

Thus, by integration it is possible to summarize important features of the system without having a complete description of it. The assumption that there is a homogeneous distribution in the system, makes it possible. However, this premise does not apply in most of the real systems. For example, the different cells within a tissue are not interchangeable, they have indeed a variable structure and function that evolve over time.

To deal with such systems we need to understand the global dynamics of a number of elements, strongly coupled between them and with highly non-linear dynamics. This is the essence of **complexity science**, which aims to investigate increasingly complex systems where the relevant features are both local and global in a way that is intrinsically not reducible. It might be remarked here that this science is still in its infancy and is therefore more an amalgamate of methods and ideas than an unified corpus.

This chapter explores some of the most relevant mathematical structures used by brain researchers to unravel brain's structure, function and dynamics. It starts by looking into the concept of brain state. The concept of the state of a system is key in engineering and physics, but maybe it is a not very well understood concept in other relevant fields of brain science like cognitive psychology.

In this chapter, three different levels of the brain are introduced, microscopic, mesoscopic and macroscopic. If we accept that there are different levels of description in the brain (synapses, neurons, columns, modules, etc), each with their own dynamical systems and time flow, then we would expect that there are possibly different mathematical structures at each of the relevant levels.

A concise exploration of the different mathematical structures and how they apply to the different levels in the brain: microscopic, mesoscopic and macroscopic, are sketched in sections 2.4, 2.5 and 2.7.

2.2 Brain state

The brain is a dynamical system of unprecedented complexity, its intricate organisation, is a product of thousands of years of evolution and social interaction.

The state space approach is used by engineers and physicists, and roughly consists of considering the system as a collection of its different states. The states are represented as points or regions in a dimensional space. The dimensionality of the state space is given by the number of independent variables that define the system. Thus, engineers study the behavior of a system based on the trajectory followed by its possible states. The phase space can be also be called state space, and is the space with all possible states of a system. For simple systems, like for example a particle moving along one direction, the phase space is two dimensional. As it is shown in the figure 2.2, the variables position and velocity of the particle are the axis, each state is a pair (position,velocity) for a given instant, and the succession of points represents the evolution of the system over time.

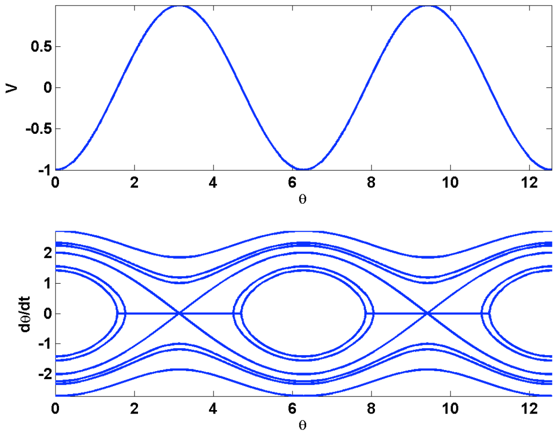

Fig. 2.2 A phase portrait is a geometric representation of the trajectories of a dynamical system in the phase plane. The figure depicts the phase space of a simple pendulum, each set of initial conditions is represented by a different curve. The x axis is the phase θ and the y axis the velocity $\frac{d\theta}{dt}$

The brain is a dynamical system, which we can always be studied using dynamical system theory (DST). DST is basically concerned with the study of the trajectories in the system's state space. This leads us to discuss what a Brain State exactly is.[40]

This attack of the problem is controversial. There are scholars, mainly philosophers who repudiate the idea of brain state. Thomas Polger holds that "brain states ... we don't really even have a clue what such things are"[41], Bechtel and Mundale have argued that "The notion of a brain state is a philosopher's fiction" [42].

Bechtel's rationale is as follows: prior to identifying brain states we need to identify separate brain areas, and as brain areas are very similar across different species, two different species having activity in the same area of the brain, they will necessarily exhibit the same brain state. Of course, this argument is flawed, because the brain areas are not the same, but they just share anatomical and physiological properties and most importantly, the identification of functional area and brain state, in which Bechtel's claim is based, is simply wrong.

But what are the brain states and how can we discern them? Buzsaki defines brain state as a transient equilibrium condition, which contains aspects of past history that are useful for future use[43]. This definition emphasizes the inherent historical aspect of complex system. Complex systems, as Prigogine said, do not forget their initial conditions: they carry their history on their backs.

Synchrony, or more precisely the phase locking of the firing of the neurons may make it possible to distinguish the brain states. For example, we find theta-modulated-gamma waves in a rat exploring its environment, when the animal is at rest or asleep they are replaced by sharp wave induced ripples. The granularity we get with the synchronicity is too wide, using Pare and Llinás words "in electrophysiological terms, waking and paradoxical sleep are fundamentally identical states [44]". We need to be able to discern how perceptual information of the external worlds coming from very different sensorial apparatus is processed, integrated and finally reflected in the synchronous waves measured in EEG. Freeman proposed the analysis of the amplitude modulation as the way to unravel the perceptual content that is conveyed in the wave.

Nunez quantitative theory of neocortical dynamics [45], defines brain state in the contextof the theory, using psychologically based control parameters which also mediate in EEG frequencies.

By way of conclusion, we can state that the main trouble with the concept of brain state is similar to that of some others scientific concepts: the same word may mean very different thnigs. We will see this later with concepts such as structure or pattern. It ought to be said that brain state can not be defined rigorously without having specified a proper context, which is both temporal and spatial scales of a particular experiment of theory [46].

2.2.1 The search for the meaningful brain states

Neural correlates of mental states are points in the phase space of the nervous system that are associated with mental states. We can avoid the hard problem of neural properties and phenomenological states [47], if we focus on the identification and isolation of those points in the phase space in order to discover, under which conditions cognitive conditions arise and evolve in time [48].

One of the most promising strategies for the identification of neural correlates of cognition is the state space approach originating from the analysis of dynamical systems.

It is still unclear how the variables that constitute the neural correlates of cognitive functions, such as perception, memory, language or consciousness, must be chosen.

The study of the nature of this correlation is today a hot topic with a long and honorable history that goes back to Helmholtz and the first Freud. The one-to-one hypothesis that associates EEG measurement with behavior, has been suggested by the first experimental psychologist, who, by the way, is a founder of psychophysics Gustev Fechner. Köhler and the Gestalt school were also favorable to the psychophysical hypothesis, which is scientific in Popper's sense because is falsifiable, but poses methodological problems, some of them

Indeed, as we will see in the next section, depending on which level of the brain we are measuring, we have different tools that will provide different state vectors [49].

2.3 Organising the Mathematical Framework

The models and theories of the brain are empirically tested and arise from signal measurements of brain activity. Undoubtedly, the methodology and technology used, will demarcate the interpretation and application of the formalisations.

For example, single cell recording typically consists on time series of voltage values, trivial to understand for a trained experimentalist. The single cell recording is easy to understand, it is a collection of time series of different voltage values. The action potentials, also called spikes or impulses are, characterised by two parameters: amplitude and duration.

Recording a population of neurons, on the other hand, is far from trivial. The recorded signal is not a point process, rather it represents a summation of several events in the neighborhood of the electrode that is placed in the scalp or directly in the cortex and adjacent regions like the hippocampus or the basal ganglia.

At a fundamental level, neuroscience is a challenging and promising testing ground for developing our understanding of how the macroscopic level emerges from the interaction of large numbers of interacting components. The functioning of the brain at the high level of mental processing is, of course, expected to be the result of the co-operate action of a very large number of neurons.

Neuroscience is an ideal test ground for the statistical mechanics. Other mathematical approaches like dynamical systems theory or physics of phase transitions applied to neuroscience will be explored in this review.

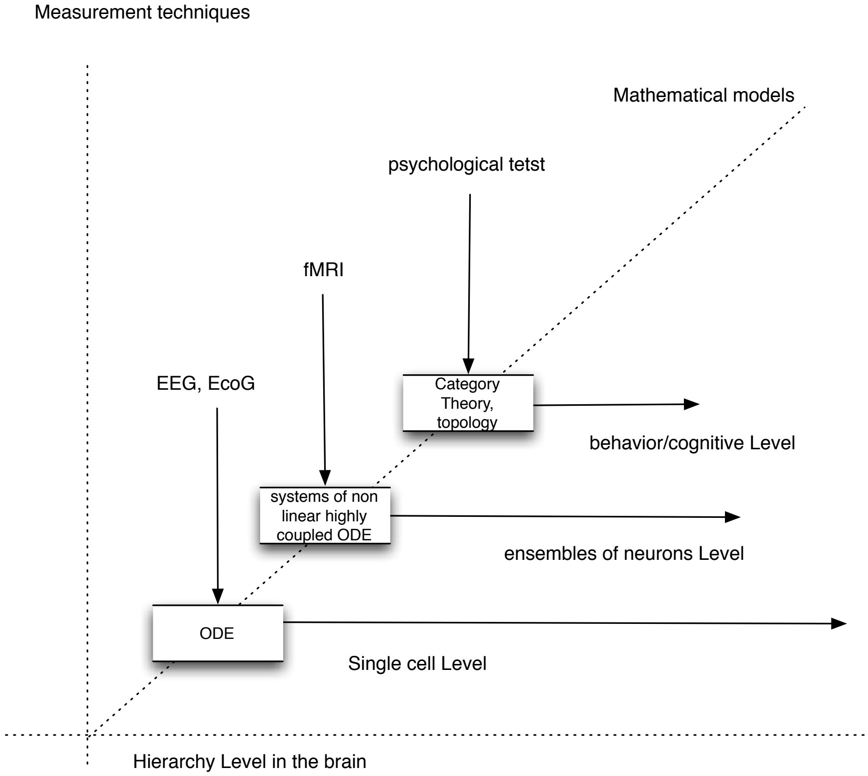

Fig. 2.3 Measurement techniques, brain hierarchy level and mathematical models are related as the figure shows. An unique and all encompassing mathematical model of the brain can not be expected.

2.4 Modeling neurons

In the brain, the microscopic level is that of the neurons and its components. Neurons are not the most numerous cells in the brain, for example there are more glial cells than neurons. The concern of modelling neurons comes from the fact that only neurons can transmit electrical signal over long distances. Hence, to talk about how the brain processes information is, in the long term, to talk about neurons exchanging information [2].

It is important, and maybe illuminating for some scholars, to realize that a neuron is a nonlinear dynamical system, complex enough to be thought in terms of very different aspects. For example, while a biologist is interested in things like the sodium channels, the pharmacologists deals with the chemical

---

[2] just as "in the long run, we are all dead" as JM Keynes once said

blockers of those channels, and a computer scientist would model neurons as digital gates[50].

At the level of a non-interacting single cell, almost everything is very well known, and accurate models of single neuron dynamics are available since Hodgkin-Huxley´s work on action potential propagation in the nerve cells in the early fifties.

2.4.1 Examples of models of neurons

For the sake of clarity, I will make a distinction between two general approaches for the construction of models of neurons.

On the one hand, we have detailed biophysical models of neurons, the Hodkin-Huxley and the Rall models are well known examples [51], [52]. On the other hand, there are more simple models of neurons with less parameters. The "integrate-and-fire" models (I-F) are an example of this type of "simple" model. The I-F models are based on the concept of action potential as a threshold process. In this view, the spatial dimension of the neuron is not taken into account, neurons are seen as unit points, and only temporal aspects of the input-output properties are included in the model.

Let us see some examples of these two kinds of models of neurons.

2.4.1.1 Detailed biophysical models of neurons

In 1952 Hodgkin and Huxley, experimenting on the giant axon of the squid, found the ionic mechanisms underlying the initiation and propagation of action potentials. They discovered different types of ion-currents in the nerve cells: sodium Na^+, potassium K^+ and clorum Cl^-. The ion channels of Na^+ and K^+, control the flow of those ions through the cell membrane, and mediate in the creation of action potentials which are electrical pulses that allow the transmission of information within nerve cells.

Formally, the Hodgkin-Huxley model consists of a equation for the total membrane current, based on the assumption that the cell membrane has an intrinsic capacitance. The ionic currents of Na^+ and K^+ are introduced in the equation which is derived from the Ohm's law. The H-H model aims to put in mathematical terms, the major features of membrane nonlinearity: the voltage-dependent ionic currents that activate and inactivate in time.

$$I_M = C_m \frac{dV}{dt} + I_K + I_{Na} + I_L \tag{2.2}$$

The H-H equation is based on Ohm's law, the left side is the total membrane current, I_M, and the right side includes the capacitance of the membrane C_m, the membrane voltage V, the sodium current I_{Na}, the potassium current, I_K

and I_L, which is the leakage current carried by other ions that move passively through the membrane.

The relevance of this model is unquestionable. Indeed, the H-H model is usually taken as the origin of computational neuroscience [53], for the biologist Denis Noble the H-H model is at the genesis of systems biology [27].

More detailed models came with cable theory. This theory comes from Lord Kelvin's theoretical analysis of signal attenuation in telephone cable. The phone cable is made of copper, which is a much better conductor than the salt solution inside a neuron. However, cable theory has been successfully applied for quantitative predictions about the attenuation of post-synaptic potentials (PSP) as they propagate through a length of dendrite towards the soma [54].

Wilfrid Rall, Denis Noble, Cristof Koch and others have applied cable theory to nerve fibers. The work of Rall is particularly relevant because he demonstrated, using cable properties, that the usual assumption in neural modeling, i.e. neurons are isopotential, was flawed. Rall was the first one to take into account the current flow to dendrites in the models of neurons; before him, the electrophysiological importance of dendrites was ignored [55], [52].

Stochastic Neuron Analysis is being used to model noisy neurons -i.e: noise involved in synaptic transmission. Green's functions, a type of function used to solve inhomogeneous differential equations, could be a possible solution for the cable equation because it is able to take account time delays in neuronal transmissions due to the non-trivial geometry of most neurons in the brain. The development of more powerful analysis methods would be of value to move these results forward, although they already incorporate much of the power in the temporal modifications in neuron transmission due to non-trivial neuronal geometry (specifically the branching dendrites) [56], [57].

2.4.1.2 Models of neurons based on threshold

On the other hand, a more simplified kind of models are integrate and fire models (I-F), which are threshold models of neurons.

Already in 1907, Lapicque [58] stated that action potentials are generated when the integrated synaptic inputs to a neuron reach a threshold value. The integrator-and-fire models are based on this idea.

They I-F models give a phenomenological description of neural behavior, but are not explanatory in the sense that they do not clarify the underlying biophysical causes of electrical activity. The I-F models, compared to the biophysical models seen above, have less parameters, which make them inadequate to predict, for example, how the neuron's behavior is affected by changes in the extracellular temperature or how it reacts to pharmacological manipulations [59].

To conclude, the much simpler integrate-and-fire models seem to be more useful for exploring collective phenomena in neuronal networks [60], but the biophysical models are more accurate than the integrate-and-fire models in describing real single neurons [61]. However, the Achilles' heels of biophysical models is the fact that we still lack a systemic methodology for measuring a large number of parameters. The blue brain project is working in that direction [62]. It is expected that in the next years, technological innovations in imaging techniques, will open new possibilities in quantitative biophysical models of nerve cells and systems. It might be remarked that neurons are dynamical systems, this conception was clear in the original model of Hodgin-Huexley, and it is a key feature that can not be overseen.

2.5 Modeling populations of neurons

To claim that neurons are the computational units in brain circuits, and that the representational power of the brain is dependent on the microscopic level, is a minor statement today. In order to give responses to questions like how the brain perceives the meanings of the stimuli or how the categories of perceptions are built, we need to go further.

This heading addresses the mesoscopic and macroscopic levels, the former is in between the microscopic level of single neuron shown above, and the macroscopic level which deals with the global understanding of dynamics by the study of whole brain areas and hemispheres.

While to measure the electric activity of a single nerve cell needs a much more complex experimental setting than measuring big cell ensembles, to understand the nature of the signal in cell ensembles is way more complex than for single cells. To measure assemblies of neurons, we use techniques like EEG or MEG which record the electric and magnetic activity of hundreds of thousands of units. For example, an electrode placed on the scalp of an individual, records a summation of myriad of electric events in the neighborhood of the electrode. The MEG records magnetic fields. Thus, in order to deal with the number of different neural events involved in the recordings, we need mathematical structures that can cope with the aggregate nature of the signal that is being measured.

Let us now start presenting the different approaches for the study of the dynamics of interacting ensembles of neurons. Complex dynamic function can not be understood by observing the parts in isolation. Integrative functions emerge from the dynamics of entire networks, rather than from linear computations performed at each nodal point of the circuit. To understand cognition we need to understand networks at a larger scale than single cells.

2.5.1 Artificial Neural Networks

One of the most prolific approaches is the Artificial Neural Network (ANN), originally created by the neurophysiologist Warren McCulloch[63]. McCulloch together with the mathematician Walter Pitts, conceived the digital computer as a suitable medium for implementing brain processes. In their view, both the computer and the brain shared a binary nature, the nervous system operations could be reduced to a set of binary states, 0 and 1. The state 1 corresponds to a firing neuron and 0 state to a non firing one.

In cognitive neuropsychology the ANN approach is concomitant with connectionism, which is usually presented as a reaction to the "failure" of the symbolist approach in Artificial Intelligence. The bibliography of the symbolist-connectionist clash is wide, and for the most part poses stimulating philosophical problems, such as the Chinese room argument[64], the symbol grounding problem[65] or the Turing test[66].

The connectionist models use linear algebra applied to point-neurons. It might be remarked here that linear systems analysis is arguably one the major achievements of XXth century mathematics. However, it is questionable that linear analysis is a suitable strategy in brain science. For example, in EEG data analysis, complex phenomena such as the self sustaining activity of the brain can not be entirely explained in linear analysis basis.

In [67], the interrogation: Are artificial neural networks (ANN) a good model for cognition? is answered negatively. William C. Hoffman identifies the weakness of the connectionist approach in the following points: *i)* the omission of the topological structure fundamental for the information processing, *ii)* omission of the morphological structure of neurons, *iii)* the inability to perform memory retrieval from point-neuron "weights" in neurobiological real-time; and *iv)* failure to implement psychological constancy. All these disallow connectionism as a valid approach for a theory of cognition.

Despite the criticisms that charge against the oversimplification in the connectionist approach, the importance of ANN is beyond question, whole research fields like cognitive neuroscience would not exist without them. The range of applications of ANN is massive: decision making, pattern recognition, medical diagnoses, speech recognition or robot modeling and control just to cite a few. The rationale of this must be found in one capital fact: ANN can implement any nonlinear dynamical system [68].

2.6 Neurodynamics

Neurodynamics is the study of neural oscillations, which aims to study brain function by building differential equations that describe neural activity patterns. Neurodynamics builds feedback models that cope with lower levels activity that self-organise into higher level ones, forming mesoscopic states that constraint the activity at the microscopic level into patterns with reduced degrees of freedom from the original miscroscopic configuration.

This area of research is strongly interdisciplinar, it shares tools and insights with other fields such as theoretical neurobiology, nonlinear dynamics, complex systems and statistical physics.

The mechanisms by which changes in the electric membrane potential trigger action potentials in neurons are well known since Hodgkin-Huxley model, which was described in section2.4.1.1.

A sequence of action potentials generated by neurons is called spike train. The spike trains form patterns which may be considered oscillatory activity, and therefore studied using mathematical tools e.g. harmonic oscillators.

Arguably, periodic spiking is at the basis of neural coding. However it might be mentioned that scientists have detected subthreshold membrane potential oscillations that deserve to be studied because they may also contribute to oscillatory activity by facilitating synchronous activity of neighboring neurons (figure 2.4).

Neurodynamics uses concepts borrowed from neural networks. In this light, neurons are conceived as integrators with a threshold. If the sum of the incoming spikes from other neuron, is above the threshold, the neuron fires, otherwise the neuron remains quiescent. The digital (discrete) bias in this conceptualisation of the dynamics of a neuron is very strong.

The simplistic vision of the neuron in the connectionist tradition, with only two possible behaviors, a neuron is at *rest* or a neuron is *firing*, may acquire a more realistic outlook by using concepts borrowed from dynamical systems theory. For example, the *old resting state* would correspond to *stable of equilibrium* and the spiking state can be explained as a *limit cycle attractor*. When the equilibrium is near a *bifurcation*, it is likely that the neuron fires.

2.6.1 Neurodynamics, classical dynamics and startistichal mechanics

In neurodynamics, the final objective is no other than the construction of dynamic models of brain activity; but why do we need to coin a new discipline when we already have classical dynamics, which is committed to the construction of dynamic models?

In classical physics, initial conditions and external forces are controlled by the experimenter, and the system can be modeled using feedforward differential equations that embodies the linear causality between inputs and outputs in the system. For example, the Newton Euler equations of an open kinematic chain shown in figure 2.5.

Neurodynamics deals with brains, which can not be modeled in this way, causality is not always directly proportional. Although traditionally cognitive psychologists have studied brain's response to a external controlled event, like

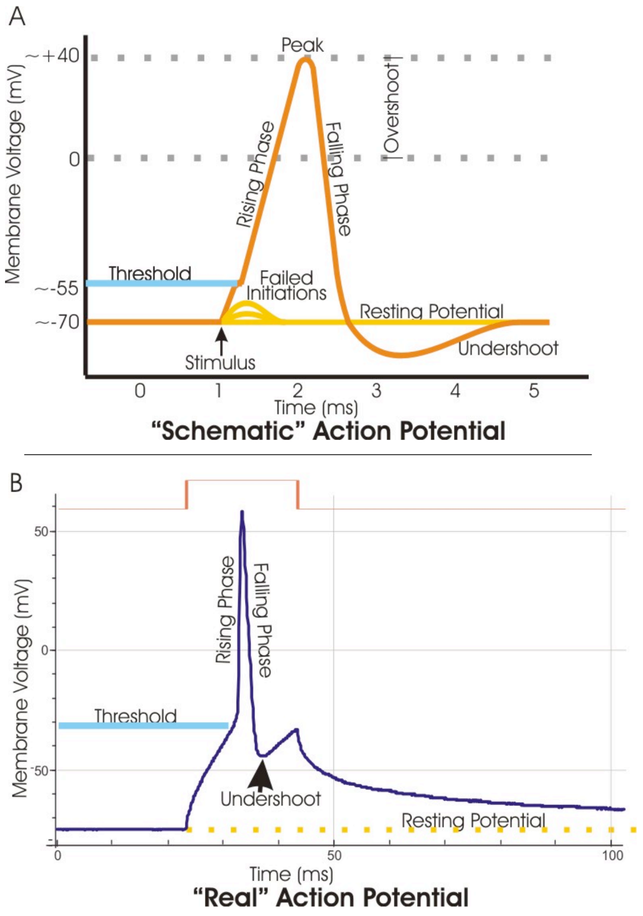

Fig. 2.4 Amplitude (mV) and duration (ms) determine the action potential of a nerve cell. A. shows an idealized action potential, in B. the recordings of action potentials are often distorted compared to the schematic view because of variations in electrophysiological techniques used to make the recording. The membrane potential under the threshold may also affect to oscillatory activity of neighboring neurons. Image from [10].

a particular task or stimulus, this approach is deficient because it does not acknowledge a capital property of brains, brain is very active even in the absence of explicit input or output.

Thus we can distinguished between, *induced activity* which is triggered by external stimuli or motor responses and *ongoing or spontaneous activity* which is the brain activity that can not be associated with external stimuli or specific events. Neurodynamics is capable of incorporating the spontaneous activity in its models, and relate it to mental state like sleep[69], rather than

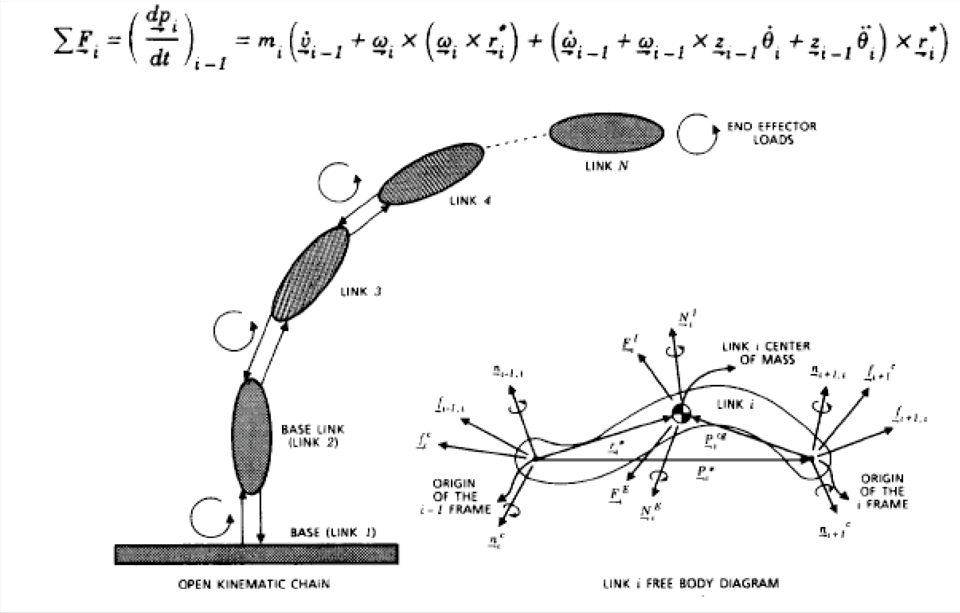

Fig. 2.5 Dynamic model using Newton-euler notation or an open kinematic chain of N links. This system presents transcendental, non-linear systems of differential equations that it is now possible to solve numerically. The analytical solution can be found for particular situations, for example when at least three neighbour system of reference have coincident origins

label it as mere noise, which is commonly done in studies focused on stimulus-related activity.

Neurodynamics and statistical mechanics are related, in order to outline this relationship we need to recollect the concept of phase space which was explained in section 2.2.

Let p be a population of neurons where three relevant properties of the neurons have been selected as follows: the neural membrane resistance R, neural membrane potential V, and time elapsed since last action potential t. The state of each neuron is a point in a three dimensional space given by $(V, R, t) \in \mathbb{R}^3$

A phase space may have as many dimensions as we want it to, then for complex systems, like for example a hypercolumn or a brain functional area, containing millions of interacting neurons, for each variable of every single neuron, a dimension is introduced in the phase space of the system, accounting to spaces of millions of dimensions. Needless to say, this approach becomes soon intractable analytically and computationally.

The statistical field models used in neuroscience, share this initial assumption: the spatio-temporal spike patterns produced by neural circuits convey information about neurons which is where the brain's representations rest. The leaky integrate-and-fire (LIF) model [70], which is variant of the integrate and fire models seen in 2.4.1.2, assume that the non-stationary temporal evolution of the spiking dynamics can be captured by point-like models of neurons [39]. The problem with this kind of model is twofold, it depends on the value of the threshold which is rather arbitrary, and it includes a reset after each spike generation, which makes the whole dynamics highly nonlinear.

One of the most important branches of the science of complexity is statistical mechanics or statistical physics, which has has been increasingly extending its range of applicability outside the domain in which was created, to deal with applications as disparate as metabolic scaling [71] or the analysis of social nets [72].

A word of caution is in order here, statistical mechanics applies to classical systems close to equilibrium, (Planck's constant does not appear), but brains are dissipative and far from equilibrium systems.

2.6.2 Biological synchronization: the theory of coupled oscillators

In this heading, I outline the theory of coupled oscillators, focusing on how this theory deals with coupled dynamical systems, for example, neuronal systems. Phase interactions between oscillators characterise a big variety of phenomena, including biological ones.

Prior to understand coupled oscillators, it is neccessary to understand what an oscillator is and how it works. An oscillator is any system with a periodic behavior, that is to say, a system that after a period T, it returns to the original state T, $f(x,t) = f(x,t+T)$.

When we have two or more oscillators the phase diagram becomes much more complex. Synchrony is a particular case of phase locking. We say that when two oscillators are synchronized the influence between them is minimal [73].

The theory of symmetrical Hopf bifurcation [74] gives the means to classify the patterns for network oscillators. The application of this theory to neural activity rests on the following three assumptions: in the brain there are intrinsic oscillators(1) that are weakly(2) and globally(3) coupled. I comment on the first two assumptions next, and the third one, neural activity is globally coupled, is sketched in the heading dedicated to global brain dynamics in section 2.7.

The weak coupling (2) of neurons is uncontroversial. As everyone knows, neurons are chemically and electrically coupled by neurotransmitters and synapses. However, it is important to have always present that computational models of neurons showing periodic activity, where coupling is a parameter that can be tuned to get the desired results, are not realistic models of brain activity.

The first assumption(1) deserves more attention: neurons are intrinsic (self-sustained) oscillators. In neurobiology, it has been reported the existence of intrinsic oscillators. Some examples are the neurons in the inferior olive nucleus or the neurons in the nucleus reticularis in the thalamus [75] which are able to generate a rhythmic activity in the same range of the sleep states, 8-14 Hz [76]. Caplan et al.[77] have shown that theta oscillations in neurons in the hippocampus are involved in spatial learning and navigation. However, in the nervous system, these are the exception more than the norm. Major characteristics of brain dynamics like transience, variability and non-stationarity in neural firing, are a limitation to the use of oscillator theory.

Oscillations are real in brain activity. Neurons have different firing patterns, depending on the cognitive state. The different behavioural and perceptual states of the brain are associated with different brain rhythms. Buzsaki's Rythms of the brain [43], is a very enlightening textbook for the study of rhythms in the brain. Resonance is a property that describes the ability of neurons to respond selectively to inputs at preferred frequencies. For a general study of the relationship between electrical oscillations and resonance in neurons, see [78]

In [79] a more detailed case of intrinsic resonance frequencies are shown; the stellate cells in layer II of mEC. It may be remarked that in less controlled scenarios, the noise and rapid fluctuations in brain activity makes the search for the intrinsic oscillators in the brain extremely complicated, in either experimental settings or mathematical modeling.

However, this kind of research should not be underestimated as scientific naiveté, and as Pérez-Velázquez suggests in [80], it may be still possible to ascribe "functional" intrinsic frequencies for specific brain states. In order for the theory of coupled oscillators to be relevant in neuroscience, the empirical reality of the three assumptions just mentioned above must be reported.

The theory of coupled oscillators has been applied to brain coordination dynamics[81], [82]. The coordination is empirically observable in terms of the synchronisation of the neural firing in neural populations. One of the most striking patterns that can be found in the literature are the standard quadrupedal gaits -walk, trot, pace, etc. In [83] is conjectured that the symmetry in the gaits, suggest a structure for locomotor central pattern generators (CPG). This hypothesis states that a a population of neurons may exist, that governs the rhythms associated to the various locomotive gaits. For an overview of models that exploit the symmetrical characteristic of the animal gaits: [84], [85]

In [11], Golubitsky discusses the importance of symmetry in neuroscience. This a very important work from a theoretical-mathematical perspective, because it gives a mathematical description of the symmetries found in the different phenomena relevant for natural scientists: animal locomotion, the connectivity of visual cortex and the vestibular system are described.

The network of oscillators is described as a graph, whose nodes represent neurons and whose arrows indicate which neurons are coupled to which. Each neuron is a system of ordinary differential equations ODE.

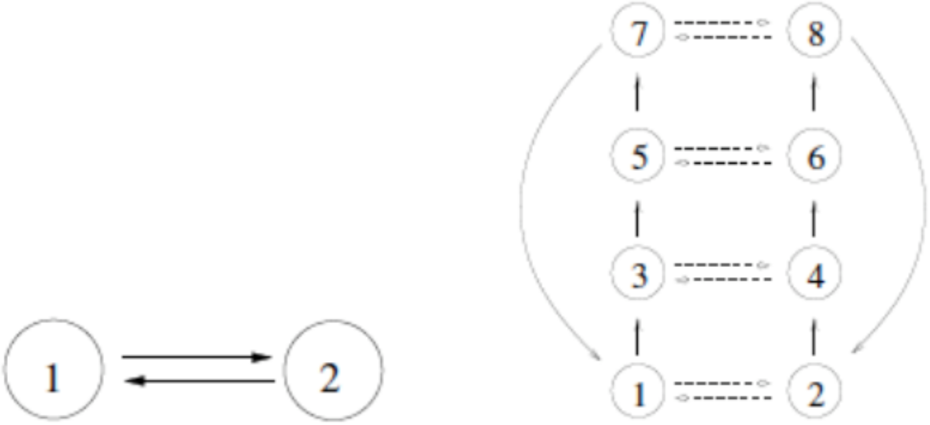

Fig. 2.6 The simplest coupled cell network is the two-cell. The class of differential equations associated to this network is $\dot{x}_1 = g(x_1, x_2), \dot{x}_2 = g(x_2, x_1)$ where $x_1, x_2 \in \Re^k$ are the state variables of the cells. Note that the system is as simple as to be defined with one single function g. Symmetry in animal gait shows that for quadruped animals, the only possible symmetry class of CPG networks is the 8-cell network shown on the right side of the figure from [11].

The use of symmetry in neuroscience is a very promising direction. Some interesting examples of this line of research are given below. In [86], the mathematical basis of the leech heart, are unveiled. The symmetry in the leech's heart is tackled with ODE and Grupoid theory. Theunissen et al. have developed a mathematical formalism on the neural coding in the cricket cercal sensory system [87]

It ought be remarked here, that the works described in this heading, are primarily theoretical. The existence of CPG, ensembles of neurons that orchestrate the rhythms involved with gaits or other periodic activities, in systems not as simple as the brain of a leech, is more of a mathematician's dream than a biological reality.

The rest of the section is devoted to models that study synchronous oscillations in neural assemblies, and therefore are under the umbrella of neurodynamics. The neural field theories are described next, including neural mass models.

2.6.3 Neural Field models and neural masses

Neural field models use mean-field approximation, describing the activity of large number of neurons by building models with "simplifying assumptions and empirical priors [88]". The mean-field approach enables to sidestep the problematics of huge dimensional spaces shown above.

The basis of this kind of models rests on the probability density over ensembles of neurons. The Fokker-Planck equation is used to calculate the density dynamics. Thus density-dynamics and mean-field approximation are the basis of neural-field models. It might be noted that while this kind of models have shown to be well equipped to describe the steady-state regime in a system, their appropriateness in non stationary behaviors is under debate [89].

Neural field models are a generalisation of neural mass models, or put in a different way, the latter are a particular case of neural field models. As we will see below, the Fokker-Planck equation allows to model populations of neurons using the probability density; when in the model the density becomes a point-mass over the space of states of a population of neurons, the model is called neural-mass model. On the other hand, when the location of the mass or expected state of the ensemble of neurons is not a point, but a function of both time and position on the brain, the model is referred as neural field. In the rest of the section the basis of neural field model and the Fokker-Planck equation are introduced. Moreover, references of major results in neural mass models are provided.

Neural field models borrow concepts from statistical mechanics. The main idea is to reduce the degrees of freedom in a population of neurons, reducing the state space dimensionality. In this context, a degree of freedom is a parameter (a variable) measured for each neuron. This is a well known procedure in physics, for example, Boltzmann's theory of gases, gets to reduce the individual degrees of freedom of the single particles, by means of the introduction of field terms, like temperature or pressure.

Operating in a similar way, a large population of spiking neurons can be reduced to a function of distribution that captures the neural state of the whole population for any given instant. Thus, for a population of neurons of any size, we can calculate the density of the phase state $\phi(V, R)$, where V is the potential and R the membrane's resistance.

The Fokker-Planck equation gives the density dynamics of large populations of either molecules in a gas or neurons in a brain area. The equation, irrespective of the dynamics of the particular components, is linear and quite simple. What is gained with this method is that even though the dynamics of a single neuron may be highly complex, even chaotic [3], the density is simple, linear and deterministic.

---

[3] It is worth mentioning that Chaos Theory is not a stable theory as can be quantum theory or electromagnetic theory, but it provides a set of mathematical tools that may be used

The equation, in the neural context, describes the time evolution of the probability density function of the state of a neuron. For example, if we define the state of a neuron as $q = (V, I, t)$, where V is the membrane voltage, I the current, and t the lapsed time since the last action potential; then the density of the phase space defined by the ensemble of neurons is $p(Q, t)$, where $p : \mathbb{R}^n \times \mathbb{R}_+ \rightarrow \mathbb{R}_{0,1}$, is a scalar function that returns the probability density at each point in the phase space.

The Fokker-Planck equation enables us going from models of individual spike neurons (points in the phase state) into electrical waves (fields), that convey the spikes of myriad of neurons. Thus it is possible to describe neuronal population dynamics in terms of the Fokker-Planck equation.

As an example of this approach, Breakspear et al. [90], have constructed neural dynamics model based on Fokker-Planck equation that tackle the multiscale hierarchy of levels in the brain. Neural mass models capture the dynamics of a neural population. The term was coined by W.J. Freeman in his 1975's book "Mass action in the nervous system [20]".

The theory of K-sets developed by Freeman in the 70's, is based on a nested hierarchy of models that describe the dynamic behavior of neural masses at the mesoscopic level. A neural mass refers to a neural system around 10^4 neurons with 10^8 synapses. These models captured the dynamics of neuronal ensembles in the olfactory bulb [91]. Neural mass models are generalized in [39] by furnishing wave equations that describe the spatio-temporal evolution of the expected neural states which are here, a function of space.

2.7 Modeling the whole brain

A common feature of the models described above in section 2.6 is that the critical step is to identify the mesoscopic state variables that refer to local masses of neurons, and then write the equations that express how the activity of self-organise neural ensembles produces global macroscopic order.

In this section I will review the macroscopic level. We need to understand the spatio-temporal patterns of brain activity. I will sketch some mathematical tools that promise new outcome to the understanding of the brain at a system level.

In order to cope effectively with the brain emergent properties that mediate in high cognitive functions, it is necessary to have an understanding of the interactions between different brain regions. The dynamics of the brain is high dimensional and non linear, therefore we need mathematical tools able to cope with the extreme complexity of brain organisation.

In EEG and EMG activity, fast and robust sensory processing has been observed through frequent phase transitions [92]. A new paradigm is emerging,

---

to model non linear phenomena. A non linear phenomenon does not mean indeterministic phenomenon.

providing a new understanding in terms of transient activity that connect different brain states or attractors, rather than in the classical approach of linear dynamics, that expects that brain activity will be confined in an attractor as a response to controlled perturbation induced from the outside.

Thus, the focus is put on the passage among attractors, rather than the study of equilibrium states, where the system will be confined after the perturbation occurred. This opens a new window of comprehension of complex systems and in particular the brain. In particular category theory is proposed as new methodology that could fuse the different existing frameworks into a coherent picture of brain structure function and dynamics. Given the importance of category theory in this thesis, it will be fully introduced in chapter 3, and used all along the thesis.

Let us start now by sketching some mathematical tools that are tackling the problem of brain dynamics at mesoscopic and macroscopic scales.

2.7.1 Brain processes as critical state transitions across levels of organisation

The internal brain dynamics can not be entirely explained within the linear analysis paradigm, the rationale for this limitation is two-fold. First, the linear-stimulus response paradigm is powerless with mental phenomena such as phantom limbs or hallucinations, which are processes that arise from internal brain state rather than triggered from external stimuli.

On the other hand, brain activity is characterised by abrupt state transitions. Ordinary differential equations (ODE), have the serious limitation that they can not describe phase transitions, they deal well describing what happens in the system before and after but not across the transition. The function of a cortical area can not be understood in isolation, as a matter of fact it depends on its relation with other cortical areas. Brain function would result form the transient coordinated activity of ensembles of neurons.

The theory of coupled oscillators that was sketched in 2.6.2 is a theoretical perspective that, so far, can not deal with the empirical evidence that shows that the rich internal brain dynamics is characterized by stochastic and chaotic properties.

2.7.1.1 Theory of attractors in brain dynamics

To define the brain as a dynamical system without describing its attractors is like building a treehouse without a tree.

It ought to be remarked here that deterministic phenomena are easier to model than stochastic ones. Thus, to claim that chaotic or other deterministic regimes may be valid candidates to model brain activity is due more to the convenience or usability of the method, than its real relevance. For a critical discussion of how realistic the theoretical dynamical regimes for brain dynamics are, see [80].

An attractor is a set towards which a dynamical system evolves over time. An attractor can have several geometric forms such as point, plane, ... or even a recurrent fractal structure, in which case it is called strange attractor. The study of attractors in neuroscience is rooted in the idea that stable, persistent activity is a key aspect for understanding neural computation. The formal definition of an attractor as a closed subset, is not a straightforward interpretation in physiological recordings of time series. As a consequence of this vagueness in the interpretation of the mathematical concept of attractor in the physiological realm, the term has been used in a rather loose way. In [80] it is argued that in neuroscience, we should speak of attractor-like systems rather than attractors.

In [93], temporal correlations between stimuli are converted into spatial correlations between attractors. Cossart et al.[94] have investigated the spatio-temporal dynamics of neocortical activity, concluding that coordinated activity of groups of neurons may correspond to circuit attractors-emergent features of feedback neural networks.

The most spectacular finding related to attractors is the cells that display persistent activity in memory tasks. As early as 1949 with Donald Hebb's inescapable book "The organization of behavior"[95], short-memory memories are thought to be attractor states of neuronal representations. Memories could be fixed point attractors. More recently, attractor dynamics has been used to explain context-dependent episodic memories in the hippocampus [96].

It is important to realize that the empirical demonstration of an attractor is, roughly speaking, what engineers want to know when looking at how a system works: perturb the system in order to study how it comes back to its resting state. EEG activity is enormously robust, it is questionable that epilepsy[97], [98] or Parkinson's disease[99] can be explained in terms of attractors. For example, epileptic seizures, are hard to destabilise once they have started.

From the fact that we have failed to "clearly" find invariant attractors, it does not follow that they do not exist, for [80] the reason is that we do not have the adequate theoretical/methodological means to address the *invariance* in brain activity, especially at the higher levels, including the behavioral one. To focus on simple systems like invertebrate nervous systems may be able to help us understand the robustness and rhythmicity found in some organisms [100].

2.7.2 Synergetics

Whole brain dynamics is high-dimensional. However, low-dimensional dynamics may be found in specific brain areas, like for example the theta oscillation [4] in the hippocampus of rodents when they are exploring [101], [102], [103].

The low dimensional dynamics, has lead some scholars to hypothesize that brain dynamics is chaotic [104],[105]. We must be prudent here; features of chaotic systems like sensitivity to initial conditions and short-term predictability should be demonstrated beforehand.

Hermann Haken, a laser physicist, is the creator of the interdisciplinary field of research called synergetics [106], [107]. Synergetics is the study of the general principles of self-organization in systems, irrespective of the nature of their individual parts. In Haken's synergetics, causality has two directions, downwards and upwards. The upward causality is local-to-global, novel systemic dynamics emerge from the interaction of the components. The original aspect of synergetics comes from the downwards causality, which is global-to-local. The slaving principle states that there are global order parameters that enslave the constituents of the system and effectively govern local interactions between them. This principle may help to explain the low-dimensionality of neural ensembles in the brain. Because the cooperation of the individual parts, one speaks of circular causality. At a critical point, a single order parameter may undergo a non-equilibrium phase transition (see bifurcation) with symmetry breaking, critical slowing down and critical fluctuations.

In synergetics, co-operation rather than competition, is hypothesised as the key mechanism that enables the existence of order parameters that determine the behaviour of the individual parts of the system. The order parameters are collective or systemic variables that arise when the system is in the brink of instability. The order parameters obey low dimensional dynamics and characterize the system macroscopically.

In neuroscience, synergetics and in particular the slaving principle is being applied to brain function. In this approach, the brain is conceived as a self-organizing system operating close to instabilities where its activities are governed by the order parameters that enslave the individual parts or neurons[108]. Moreover, higher cognitive functions, like consciousness, for example, has been modeled in terms of order parameters.

---

[4] Theta oscillation can be used with two different interpretations. In rats and mice, theta oscillation is a specific type of regular oscillation seen in the hippocampus and several other brain regions connected to it. Thus in rats we say “hippocampal theta oscillation”. The original sense of theta is used in human experimentation meaning: EEG waves falling into a frequency range of 4-7 Hz, regardless of where in the brain they occur or what their functional significance. Therefore “theta oscillation” and “hippocampal theta rhythm” are different because in the first case the EEG waves are obtained through electrodes glued to the scalp measuring several areas of the cortex, while in the latter case is implicitly assumed that the EEG wave comes from the hippocampus, using invasive techniques in rodents .

2.7.3 Dynamic geometry

The tensor network theory was proposed by Pellionisz and Llinás [109], it leads to a geometric interpretation of brain function. The relation between the brain and the external world is defined in terms of an interactive geometrical relationship between sensory and motor expression.

The theory deals with the "internal" functional space that can interact successfully with the "external reality". Llinás explains this internal space in terms of a dynamic geometry. Dynamic geometry is the mapping between the event in the external world and the event in the internal space.

Prior to defining any metric tensor one needs to address a well-defined distance function. As the anatomy of the brain does not present a smooth and linear representation of the external world, we need to define a distance function which will be functionally isotropic.

The multidimensional spaces of the central neural system are, for the most part, not definable in well-known geometries such as Euclidean or Riemannian spaces. There must be a stochastic metric -distance function- able to deal with the stochastic nature of the functional space.

The neurological reasoning is as follows, local chaotic oscillations may give rise to synchronous oscillations on a larger scale [110] [111]. Thus, from the local chaotic behavior and its associated stochastic metric, arises a smooth metric structure at the global level, for which there exists synchronous oscillations.

Roy and Llinás have recently upgraded this theory [112]. The novelty is the use of a stochastic metric tensor, which is defined as five-dimensional space-time where the fifth dimension is a probability space. The stochastic metric tensor in dynamic geometry is defined as $g_{ij} = h_{ij}(x, \xi)$, where x is the four dimensional Minkowski space (3 for space + 1 for time), and ξ is a fifth dimension. The tensor is applied to discern between different dream states. In REM sleep $g_{ij} \neq 0$ because the internal degrees of freedom increase in the dream state. In deep sleep, the degrees of freedom seem to be frozen, $g_{ij} = 0$.

2.7.4 Discrete Approaches to Neural Phase Transitions

In this section some tools used to study of the transitions between levels of neuronal organisation are briefly described. I present some tools that tackle the problem of dimensionality reduction for large scale brain modeling. As it was pointed out in the heading 2.7.1 methods based on differential equations have the serious limitation that they can not describe phase transitions. Transitions can be controlled by coupling strength between components connections.

The phenomenology of phase transitions in neural processes are being developed using new mathematical tools. For example, Coupled Map lattice (CML) [113][114], Phase transitions in random graphs [115] and Neuropercolation [116]. In particular, Neuropercolation tries to generalize the cellular automata behavior to simulate the complex behavior of neural populations. Neuropercolation extends the concept of phase transitions to large interactive populations of nerve cells. The construction of a coherent framework that fuse the different approaches presented here has not been achieved. However, empirical findings like neuronal avalanches [117], [118] or percolation transitions [119] seem to back this research area.

Chapter 3

The Categorical Imperative: Category theory in Cognitive and Brain Science

Traditionally, mathematical formalisms in cognitive science have been confined to toy model world descriptions [120], [121]. In the absence of a theory written in mathematical terms, the separation between the different disciplines that form the cognitive sciences, will be progressively more acute and an understanding between them unattainable.

It is unreasonable to think that purely linear mathematics can be the killer tool to model the complex interactions that take place in the brain. We may indeed, need a new mathematical language for describing brain activity.

As Mac Lane pointed out, the fact that mathematics is protean, means that the same mathematical structure has many different empirical realizations, this is why mathematics is about patterns or forms, where each of these forms describes different aspects of the external world. Mathematics is therefore the part of science that deals with more than one empirical context. This same relationship between mathematics and the external world, suggests a similar one between category theory and mathematics: all the mathematical fields can be organised according their structure by specific categories, and such specific categories can be organised using the notion of general category as provided in the axioms of category theory as it will be shown this chapter.

This chapter aims to set the agenda of category theory (CT) as the appropriate methodology that provides the necessary theoretical framework in order to build cognitive models in mathematical terms.

3.1 Introduction

The structure and dynamics of the human brain are overwhelming, 10^{10} neurons with 10^4 connections each one. This gives an idea of the complications we are facing here.

We can identify two major difficulties in modelling cognitive systems as complicated as the brain.

1. Undeniably the brain is a very malleable system, but the theories and models of the brain do not have to be that malleable. The proliferation of fMRI experiments are not necessarily producing a better understanding of brain morphology and functioning.
2. One of the biggest challenges in science today is to decipher the map of the brain. In Crick's words, "there is little hope to understand how the brain works until we have a map of the neural wiring in the mammalian brain". The view that the brain configures metric maps that reflect objects and processes from the world out there, needs to be revisited with a meaningful and formal concept of structure[122],[123].

Despite the daunting difficulties in understanding how cognition arises from neural interactions, it is possible to make advances. What is needed now more than ever in the brain sciences, is to developed model-related analysis strategies that enable us to propose hypothesis that can be contrasted to empirical investigation. In Michael C. Reed words[124]:

> experimental neurobiology, like experimental physics, needs input from deep and imaginative theorists.

In this thesis is conjectured that category theory could provide the necessary concepts to bridge the gap between the different levels of brain activity: from microscopic neuronal activity, to macroscopic aspects, like for example, emotions or abstract thoughts, which are defined in a rather vague fashion in the psychological literature.

Category theory proposes a radically new view in system modeling, which shifts the focus from the elements to the relationships between those elements.

This chapter has three main objectives that will be explored, as indicated bellow, in the three following sections.

- to introduce the mathematical theory of categories, providing formal definitions of key concepts (section 3.2).
- to outline the basis of a new theoretical framework based on CT that may cope with the binding problem (section 3.3).
- to provide a state of the art of the actual applications of this theory in the domain of cognitive and brain science (section 3.4).

3.2 Category Theory

Category theory is a new branch of mathematics i.e. abstract algebra, that was created during the 40's by Mac Lane and Eilenberg, as a way of studying and characterizing different kinds of mathematical structures in terms of their admissible transformations.

I will make a historical sketch of the birth of category theory and the first categorists. It ought to be mentioned that in science, history tends to be dismissed as irrelevant, the exclusive focus on the present and the last technical achievements is, I believe, nothing but detrimental. A historical perspective can strengthen our comprehension of the current concepts, improve our understanding of the problem, and help to foresee future developments.

The birth of category theory can be dated to 1945, with the paper by Eilenberg and Mac Lane entitled, "General theory of natural equivalences", in which the notion of natural isomorphism is defined and used to capture the shared structure in homology groups. By "natural" in natural isomorphism, Mac Lane and Eilenberg intended a common "phenomenon", for example a property existing in different contexts. In their original paper, group homomorphisms were "natural".

Today it is well known that category theory deals with structures and the relationships between them. Each kind of structure, for example, algebraic or topological, forms a mathematical category. Indeed, categories in mathematics arose of the need of a formalism to describe the passage from one type of mathematical structure to another [125].

It might be said that Mac Lane and Eilenberg underestimated the potential of their own creation, they did not see category theory as "a field for further research effort, but just as a language of orientation"[126]. It was subsequently especially thanks to the efforts of Kan and Grothendieck, that category theory was perceived as a radically different way of doing mathematics, far more important that the "handy language to be used by topologists and others" as Mac Lane initially originally intended. Gothendieck defined abelian categories through axioms that, importantly, were not like the axioms for abelian groups, but they constituted the description of abelian groups and also of other similar categories. The rationale of Gothendieck's approach anticipated the current understanding of category theory: we pay no attention to the nature of objects and arrows, but to the patterns of arrows that connect the objects.

Kan, in 1956, created the concept of adjoint functor which is now considered as the most significant contribution of category theory to the broader arena of mathematical thinking [127], or a conceptual tool of the first magnitude on the par with the idea of continuous functions [128].

Ehresmann and Lawvere deserve special mention as well, in the origins and development of category theory. Ehresmann, in the 40's worked out the notion of local structures and later made major contributions with the concepts of groupoids and structured categories.

Lawvere's doctoral thesis in 1963 is a major milestone in the history of category theory. He proposed a new theoretical framework for mathematics itself, in which the set-theoretical framework is replaced by the axiomatic system provided by category theory. Lawvere's work is intrinsically related to the construction of a proper foundation for mathematics: the category of categories may provide such foundation.

Mac Lane, influenced by Lawvere, changed his initial view on category theory as a mere "handy language". He followed with interest the Lawvere's vision of category theory as a true and definitive framework for mathematics.

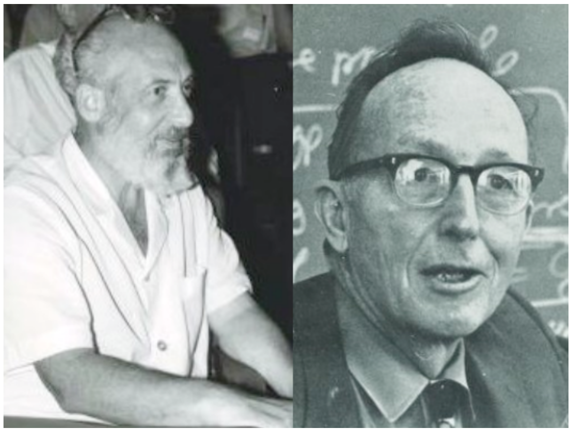

Fig. 3.1 Samuel Eilenberg, left, and Sanders Mac Lean, right, the two creators of category theory

However, it must be said that Mac Lane does not commit to categorical-theoreic (Lawvere) foundations of mathematics neither a set-theoretic (Bourbaki). In his beautiful book “Mathematics Form and Function” [129], he concludes that there is, as yet, no simple and adequate way of conceptually organizing all mathematics. Set theory, he points out, is a handy vehicle, but its constructions are sometimes artificial; while category theory, on the other hand, is almost everywhere in topology and algebra, but “do not as yet relate very well to most analysis”.

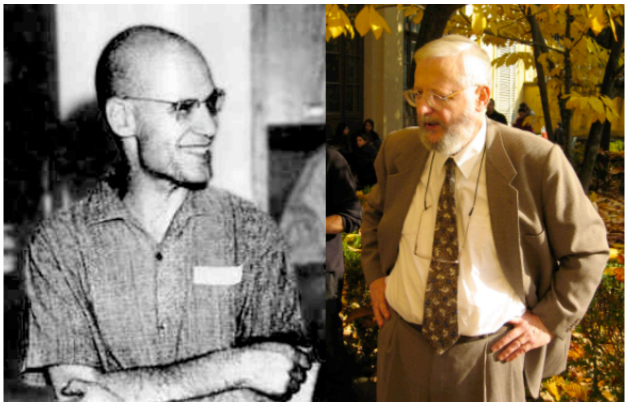

Fig. 3.2 On the left, Alexander Grothendieck, a very remarkable man, Fields medal in 1966, disenchanted with world politics and academic circles, he moved into the woods, maybe somewhere in the Pyrenees. On the right, William Lawvere, the champion of category theory as a foundation for amthematics.

Category theory can be seen as the “mathematics of mathematics” because one can make definitions in a broader context, that is to say, without reference to the particulars to which the definition might be applied [130]. In contrast to the classical atomist approach of set theory that builds all its notions in terms of the elements, CT does it in terms of arrows. Indeed the motto of CT, “thinks in arrows”, means that CT is able to reformulate key concepts in terms of arrows rather than the elements themselves. As we will see later, most constructions in mathematics may be defined by universal properties, expressed in categorical terms.

Category theory is inherently iterative: there are transformations between categories: namely functors, and transformations between functors called natural transformations. It is interesting to note that Mac Lane and Eilenberg originally defined “category” in order to be able to define “functor”, and “functor” in order to be able to define ”natural transformations”.

Thus, the theory of categories, functors and natural transformations, which is the actual name for Mac Lane and Eilenberg’s theory, provides a characterization of the notion of a structure-preserving transformation, and thereby of a species of structures admitting such transformations.

There are some excellent textbooks in CT. A classic is “Categories for the working Mathematician” by Mac Lane[131] which is a very clear and elegant text, though not easily accessible for readers not familiarised with category theory. More basic introductory texts are: “Conceptual Mathematics” by Lawvere and Schanuel[132], which is written in a very didactic style, and addressed for general reader, “Category theory” by Awodey[128] and “Arrows, Structures and Functors” by Arbib and Manes[133], which is a strongly recommendable book for beginners in CT with some mathematical background.

“Basic Category Theory for Computer Scientists” by Pierce[127] is a short and and clear introductory text. Computer science students may find in “Category Theory for Computing Science” by Barr and Wells[125] an extended monograph with solutions to exercises. An “Introduction to Category theory in for easy movements” by Schalk and Simmons[134], contains also solutions to exercises. Now it is time to give a definition of category.

Definition 3.1. A category K is a collection of objects $Obj(K)$, where for each pair of objects A, B of $Obj(K)$, there is a collection of morphisms, $K(A,B)$, being A the domain and B the codomain that satisfies *i* and *ii*:

i associativity of composition: for any pair of morphisms $f : A \to B$, $g : B \to C$, there is a composite morphism, $g \circ f : A \to C$, such that

$$(h \circ g) \circ f = h \circ (g \circ f) \tag{3.1}$$

for all $f : A \to B$, $g : B \to C$, $h : C \to D$

ii identity: for each object A, there is a given arrow $1_A : A \to A$ called identity arrow. Thus, for any arrow $f : A \to B$

$$1_B \circ f = f \text{ and } f \circ 1_A = f \tag{3.2}$$

Categories are nothing more than an algebra of arrows equipped with the composition operator. A category is *anything* that satisfies the definition above, the objects can be sets, groups, monoids, vector spaces... or neurons in the hippocampus. Categories provide a domain of mathematical discourse, characterized in a very general way[133]. In this vein, classical logic or set theory are particulars in the more general and universal theory of categories.

3.2.1 Examples of categories

Let us see more examples of categories. Trivially, the category **0** is the empty category with no objects and no morphisms.

The category **1** has no objects and one morphism which is necessarily the object's identity arrow 3.3. Category **0** and category **1** are called *discrete* *categories*, because the arrows are just the identity arrows, one for each object in the category. The category **2** has two objects and three morphisms, id_A, id_B and f, as the diagram 3.4 shows.

$$\begin{tikzcd}
A \arrow[loop above, "id_A"]
\end{tikzcd}$$

Fig. 3.3 Category **1**

$$\begin{tikzcd}
A \arrow[loop left, "id_A"] \arrow[r, "f"] & B \arrow[loop right, "id_B"]
\end{tikzcd}$$

Fig. 3.4 Category **2**

We first of all must note that the categories shown above can be expressed as graphs where the objects are nodes and the morphisms are edges.

More interesting are the *concrete categories*, in which the objects are sets, possibly with some structure, and the arrows are functions or homomorphisms when the objects have structure. In the textbooks, categories are usually defined in terms of ordinary set theory. The category of sets, **Set**, is the category whose objects are sets and whose arrows are functions between sets, and the associativity of composition and identity hold.

Other examples of *concrete categories* are: the category of pre-order sets **Poset**, whose objects are posets and the arrows monotone functions, the category of monoids **Mon**, whose objects are monoids and the arrows functions that preserve the monoid structure, or the category of vector spaces, **Vec**, whose objects are vector spaces and the arrows linear maps.

3.2.2 Definition of some key concepts in the theory of categories

In this heading I will be formally defined some concepts that will be used in this thesis.

Let us start by the concept of diagram. Everyone, by glancing a category theory book, can realise its diagramatic nature. Diagrams, are the way in which this theory proves theorems and states properties of categorical objects. The categorists do this by exploiting the commutativity of the diagrams. Commutative diagrams, work as the equations in category theory, and provide a geometric, and therefore more "visual" description than the purely algebraic insight that is given by the equations.

Let us start by defining the "language of category theory", that is to say, diagrams. Diagrams are defined upon the well known concept of directed graph.

Definition 3.2. A directed graph is a class of nodes, together with a class of oriented edges for each pair of nodes. Thus, a directed graph is a category without composition or identities.

Remark 3.1. A directed graph can be defined without using category theory as a 4-tuple $G = (N; E; \alpha_s; \alpha_t)$ where E is a set of arrows, N is a set of nodes and α_s is a function defined in $E \rightarrow N$ that gives the source of each arrow and tghe function α_t that gives the target of each arrow. Thus $\alpha_s \circ f = domf$ and $\alpha_t \circ f = codf$.

Definition 3.3. A directed graph in a category K, is a diagram whose nodes $X_i, X_j \ldots$ are labelled by the objects D_{X_i} of K and whose edges $X_i \rightarrow X_i$ are labelled by morphisms in $K(D_{X_i}, D_{X_j})$.

Definition 3.4. A diagram in a category K is commutative if for every pair of nodes X,Y in the diagram, all the paths in the diagram yield the same arrow in the category K.

$$\begin{tikzcd}
X \arrow[r, "f'"] \arrow[d, "g'"'] & Z \arrow[d, "g"] \\
W \arrow[r, "f"] & Y
\end{tikzcd}$$

Thus the path in the diagram $X \to Z \to Y$ shown in the following diagram, which is the arrow, $g \circ f'$, in K, is equal to the path $X \to W \to Y$, which is the arrow $f \circ g'$ in K. The two paths between X and Z compose the same arrow $g \circ f' = f \circ g'$, because for every argument, they have the same value.

For example, the associativity of composition that all categories must have based the definition of category given above, means that in the following diagram, the two paths or composition of arrows, $(h \circ g) \circ f$ and $h \circ (g \circ f)$ are identical, $(h \circ g) \circ f = h \circ (g \circ f)$, as the the following diagram shows.

$$\begin{tikzcd}
A \arrow[r, "f"] \arrow[dr, "g\circ f"'] & B \arrow[d, "g"'] \arrow[dr, "h\circ g"] & \\
 & C \arrow[r, "h"'] & D
\end{tikzcd}$$

The notion of dual is very important on category theory, because the definitions in category theory can be dualized, that is to say, the definitions come in pairs, which obviously simplifies the work of defining and understanding concepts in a theory.

For example, we have initial objecategory theory/terminal object, product/coproduct, limit/colimit, cone/cocone, monomorphism/epimorphism.

The dual of a category K, K^{op}, is obtained just reversing the arrows.

Definition 3.5. The dual K^{op} of a category K, is such that the objects are the same, $Obj(K) = Obj(K^{op})$ and there is a one-to-one correspondence between the arrows. For every arrow $f : A \to B$ in K there is a dual arrow $\bar{f} : B \to A$ in K^{op} with composition and identity properties defined in the normal way.

Definition 3.6. A cone to a diagram D is an object C, and a family of arrows $c_i : C \to D_i$, such that for every arrow $\delta D_i :\to D_j$ in D, the following diagram commutes. C is called apical object and the morphisms $c_i : C \to D_i$ leg morphisms.

$$\begin{tikzcd}
C \arrow[r, "c_j"] \arrow[d, "c_i"'] & D_j \\
D_i \arrow[ur, "\delta"] &
\end{tikzcd}$$

The cones for D form a category of cones **Con**(D).

Definition 3.7. The dual of a cone is a cocone. Thus a cocone of a diagram D in a category K is an object C and a family of arrows $c_i : D_i \to C$, such that for every arrow $\delta D_i \to D_j$ in D, $c_i \circ \delta = c_i$ which is equal to predicate that the following diagram commutes.

$$\begin{tikzcd}
C & D_j \arrow[l, "c_j"'] \arrow[dl, "\delta"'] \\
D_i \arrow[u, "c_i"] &
\end{tikzcd}$$

3.2.2.1 Universal constructions in categories

Category theory starts with the realisation that most constructions in maths may be defined by universal properties. The basic notions of category theory are six: object, morphism, domain, codomain, associativity and identity. However, strictly speaking, there is only one: morphism, because the other five: object, domain, codomain, identity and associativity, rely on the concept of morphism.

Following Kan and Grothendieck investigations, the concepts of object, morphism, domain and codomain can be described in terms of adjoint functors [135]. Here we will focus on first order concepts of Category theory, thats is to say, concepts that strictly belong to categories and functors. Higher order relations, like natural transformations and adjoint functors will be defined later, when needed in section 3.4.

Before starting with the survey of universal constructs, we need to precise what is intended here by universal. An universal construction describes a class of objects and accompanying arrows that share a common property. The universal properties allow us to do not have to deal with all the particular details of a construction, but to prove if the construct verifies an universal property. In that case, we can forget the details because all what there is to know about the construct is already contained in the universal property This is, in actual fact, the basis of all scientific modelling.

The rationale behind universal properties is to obtain definitions of objects as general and descriptive as possible. It has been suggested that universals of the mind may be expressed using universal properties in the theory of categories [136]. Category theory can encompass these two apparently divorced terms ,generality and descriptiveness, because allows to capture the structure of the object in the most general way.

Definition 3.8. A morphism $f : A \to B$ in a category K, is a monomorphism if for every pair of arrows $g, h : X \to A$, $f \circ g = f \circ h$ implies $g = h$. Put it diagramatically, f is a monomorphism if whenever the following diagram commutes, implies $g = h$.

$$\begin{tikzcd} X \arrow[r, shift left, "g"] \arrow[r, shift right, "h"'] & A \arrow[r, "f"] & B \end{tikzcd}$$

Remark 3.2. Note that the predicate "for every pair of arrows $g, h : X \to A$, $f \circ g = f \circ h$ implies $g = h$" states an universal property.

Definition 3.9. A morphism $f : A \to B$ in a category K, is an epimorphism if for every pair of arrows $g, h : B \to X$, $g \circ f = h \circ f$ implies $g = h$. Put it diagramatically, f is an epimorphism if whenever the following diagram commutes, implies $g = h$.

$$\begin{tikzcd} A \arrow[r, "f"] & B \arrow[r, shift left, "g"] \arrow[r, shift right, "h"'] & X \end{tikzcd}$$

Definition 3.10. A morphism $f : A \to B$ in a category K, is an isomorphism if there is a morphism $g : A \to B$, such that $g \circ f = 1_A$ and $f \circ g = 1_B$. The objects A and B are said to be isomorphic because it exists and isomorphism between them, and since inverses are unique g is the inverse of f, $g = f^{-1}$. Trivially, an isomorphism in K is also an isomorphism in K^{op}.

Definition 3.11. An initial object of a category K, is an object 0 in $Obj(K)$, such that for any object X in $Obj(K)$ there is one and only one morphism $0 \xrightarrow{f} X$.

Thus, the idea behind initial object is the smallest thing with all the structure in question. For example, in the category of sets **Set**, the empty set $\emptyset$ is a terminal object, in the category determined by one poset[1], an initial object is the absolute minimum for the poset (in the case it exists)

The concept of terminal object is dual of initial object.

Definition 3.12. A terminal object of a category K is an object 1 in $Obj(K)$ such that, for any object X in $Obj(K)$ there is one and only one morphism $X \xrightarrow{f} 1$.

In the category determined by a poset the terminal object is the absolute maximum for the poset (in the case it exists).

Two initial objects (or two terminal objects) in a category are uniquely isomorphic. That is, for two initial objects (or terminal onjects) I, I' there is one and only one isomorphism $I \xrightarrow{iso} I'$. Thus multiple initial objects in a category are not distinguishable because they are isomorphic. The same holds for terminal objects.

---

[1] Note that the category determined by one poset is not the same than the category of posets **Pos**

Definition 3.13. A product of two objects A, B in $Obj(K)$, is an object P equipped with two morphisms, $P \xrightarrow{p} A$ and $P \xrightarrow{q} B$, such that given any pair of morphisms, $X \xrightarrow{f} A$ and $X \xrightarrow{g} B$, there is an unique morphism h making the following diagram commutes.

$$\begin{tikzcd}
& X \arrow[ddl, "f"'] \arrow[d, "h"] \arrow[ddr, "g"] & \\
& P \arrow[dl, "p"] \arrow[dr, "q"'] & \\
A & & B
\end{tikzcd}$$

A coproduct is the dual of product, so it can be obtained reversing the arrows in the above diagram. Formally:

Definition 3.14. A coproduct of two objects A, B in $Obj(K)$, is an object C with two morphisms, $A \xrightarrow{p} C$ and $B \xrightarrow{q} C$, such that for every object $Obj(K)$ and all pairs of morphisms, $A \xrightarrow{f} X$ and $B \xrightarrow{g} X$, there is an unique morphism h making the following diagram commutes.

$$\begin{tikzcd}
& X & \\
& P \arrow[u, "h"] & \\
A \arrow[uur, "f"] \arrow[ur, "p"'] & & B \arrow[uul, "g"'] \arrow[ul, "q"]
\end{tikzcd}$$

The definition of coproduct can be generalised to n representations.

Definition 3.15. A coproduct of a family of n objects Ai consists of an object $\coprod_i A_i$ together with a family of arrows $(A_i \xrightarrow{\pi_i} \coprod_i A_i)$ such that for any object C and family of arrows $(A_i \xrightarrow{\alpha_i} C)$ there is an unique arrow $\coprod_i A_i \xrightarrow{h} C)$ such that the following diagram commutes.

As it will be shown in section 3.4.4.3, Philips et al. [137], have pointed out that a fundamental cognitive principle such as transitive inference or class inclusion, involve the capacity to compute coproducts.

Products and copropucts can be defined in a more general way. Indeed, a product is a form of limit and coproduct is a form of colimit.

$$\begin{tikzcd}
C & \\
\coprod_i A_i \arrow[u, dotted, "(h)"] & A_i \arrow[l, "\pi_i"'] \arrow[ul, "\alpha_i"']
\end{tikzcd}$$

Definition 3.16. A limit for a diagram D is a terminal object in Con(D). Thus the cone $(c_i : X \to D_i)$ is such that for every cone in D $(c_i : X' \to D_i)$, there is an unique morphism $\delta : X \to X'$ such that the following triangle commutes.

$$\begin{tikzcd}
X' \arrow[r, "\delta"'] \arrow[dr, "c'_i"'] & X \arrow[d, "c_i"'] \\
 & D_i
\end{tikzcd}$$

Definition 3.17. A colimit for a diagram D is a cocone with the universal co-property that for any other cocone there is a unique arrow such that the following diagram commutes. Initial objects are colimits of empty diagrams.

$$\begin{tikzcd}
X' & X \arrow[l, "\delta"'] \\
 & D_i \arrow[u, "c_i"'] \arrow[ul, "c'_i"]
\end{tikzcd}$$

The importance of limits and colimits relies on the fact that they are general mechanisms for combining structures. For example, in [138], Ronald Brown and Timothy Porter suggest that colimits may give useful analogies to the way complex systems operate. The concept of colimit may help to explain how individual parts provide a working whole, that is to say, it captures the structure of the system, making irrelevant the choice at intermediate stages in the dynamics of the system.

Ehresmann and Vanbremersch have been pioneered in finding the relevance of colimit to biological development[12]. The dynamics of the category of neurons produces evolving structures which can be captured with colimits. This notion of colimit allows to "put together" different neural process into a coherent whole.

In chapter 8, colimits are presented as the main modeling tool for spatial representation in the brain hippocampus.

As noted above, links between objects in a category are called morphisms or arrows. There are also links between categories, these are called functors. Let A and B be categories, a functor $F : A \to B$, maps every object A of category A to an object $F(A)$ in B and each morphism in A to those of B. Functor is a structure preserving map between categories. Formally:

Definition 3.18. A functor $F : A \to B$, from category A to category B verifies:

1. $F : Obj(A) \to Obj(B)$ such that for all $a \in Obj(A)$, $F(a) \in Obj(B)$
2. For each morphism $f : a \to a'a, a' \in A$ exists a morphism $F(f) : F(a) \to F(a') \in B$ such that for all $a \in Obj(A)$ and all composable arrows f,g in A:
   a. $F(1_a) = 1_{F(a)}$
   b. $F(g \circ f) = F(g) \circ F(f)$

3.3 The cat-level avenue

The brain is an ultra complex system with the additional difficulty that its behavior is very hard to analyse in vivo. Although the trees are well known(detailed models at microscopic level), there is still missing an unified theory of the wood(macroscopic models). Unfortunately, it seems that in brain sciences, descriptive understanding is divorced from the understanding of underlying mechanisms. Put it in a different way, explain the higher level properties like for example: memory or language, in terms of lower level properties, like action potential of firing rate of single neurons, requires far more than inferences or intuitions. But there are good news in all this obstacles, the brain dynamics is, at least in principle, amenable to mathematical representation.

In chapter 1, it has been described a number of different mathematical tools applied to neuroscience. Depending on the level of organisation in the brain, we find different approaches and techniques better suited than others. We thus have a whole host of mathematical structures and frameworks to integrate.

In order to acquire a global understanding of brain dynamics, we need to create a new methodology, able to fuse, or at least vertebrate, all these different approaches. It might be remarked here, that I am not claiming for an unified theory of cognition of any sort, but to bring in a new methodology, able to modify in a considerable manner the existent frameworks so allow for new results to be obtained.

This approach based on category theory, tries to relate the levels, as well as making clearer the mathematical structures needed at each level, and their efficiency in obtaining relevant explanations of what is observed in the brain.

This is roughly how category theory cope with the hierarchy of organisational levels.

1. To detect the localisation of the meaningful patterns of components at $level_i$
2. To build the colimit that embodies an object at $level_{i+1}$. The colimit is obtained as a result of the binding process of components at level $level_i$ and is a category of objects of $level_i$
3. Go to step 1 in order to continue the construction of categories of higher level i+2...

The mathematical theory of categories may help to set the basis of a novel framework to cope with the phenomenon of emergence in neuroscience.

In particular, it gives a common framework for studying the different levels of description. In neuroscience, these levels are typically assumed to be two: microscopic and macroscopic, or three: microscopic, macroscopic and the mesoscopic level, which lies between those.

There are probably many more levels, each feeding into and being controlled by networks at other levels. We need mathematical objects able to deal with these networks.

However, from the microscopic $level_1$ where the category **Neur** is placed all the way up to $level_N$ of mental objects **ImO**, there are N-1 levels that need to be addressed. All those N-1 levels are in truth multiple scales of organization that go from local network of neurons to the scalp topographic scale of the EEG and the statistical correlation of MRI.

In this vein, the category of neurons **Neur** and the category of mental objects **ImO** are models of the brain at very different hierarchical levels, namely the microscopic and macroscopic. Thus, the category **Neur** is composed of neurons, and models the physical structure of the brain and its elementary neural dynamics.

Category theory may provide the answer to the fundamental question of *how cognitive functions, including higher mental processes, arise from the structural connectivity of the brain?* Categories of increasingly complex objects, are built and formally specified, using terms like for example product, colimit or functor that will be defined in section 3.2.

In [139], E.C. Zeeman holds that a mathematical explanation of how the brain works has to rely on the concept of isomorphism.

> *First develop a piece of mathematics X that describes the permanent structure (memory) and the working (thinking) of the mind, and another piece of mathematics Y that describes the permanent structure (anatomy) and working (electromechanical) of the brain; then, from hypothesis based on experimental evidence, prove an isomorphism* $X \cong Y$

Of course, we must be cautious with *experimental evidence* about brain studies mentioned in a paper of 1962. However, the approach and the emphasizes given to the concepts of structure and morphism are totally relevant. There are two main issues to be considered from the above quote.

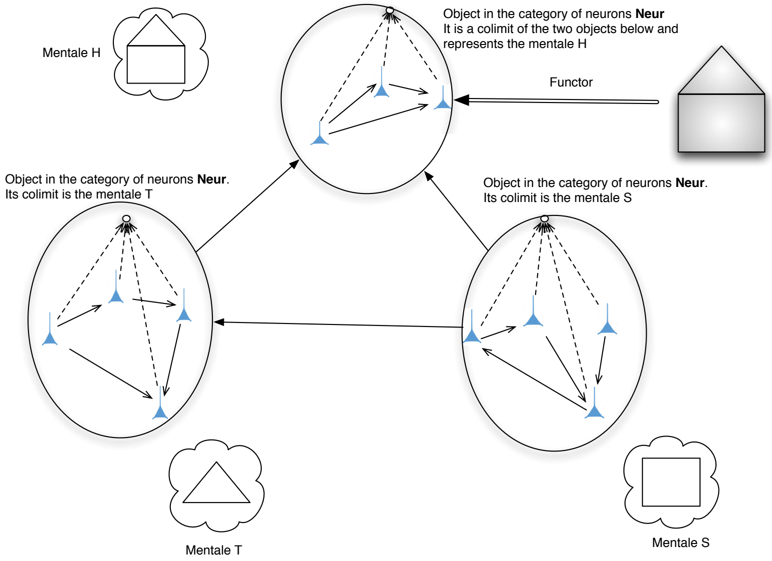

Fig. 3.5 The figure depicts the categorical construction of a physical object, the gray house on the right. The *mentalae* S and T, on the bottom of the figure are concepts that are reified in a neural organisation, in the category **Neur**. The colimit of the objects represent the *mentale* H (the house). Finally there is a functor between the physical object and the neural category **Neur**

First, if category theory is the mathematical theory that studies the structure, we are going to necessarily need such theory, when we try to model the structure and function of the brain. The focus in not in what the objects may be or do but what arrow-patterns exists between the objects. The arrows reveal the structure and any universals that pervade the structure.

Second, the relation of representation between two systems, involves the establishment of a mapping, ideally an isomorphism, or in the actual case, a homomorphism, as it will be shown in chapter 6. In category theory we can establish morphisms between objects within a category, but also, in the next level up, we can find morphisms between categories, these comprise functors, which are structure-preserving maps between categories as defined in 3.18.

3.4 Applications of Category Theory in Cognitive and Brain science

In this section I provide a review of the most relevant uses of the mathematical theory of categories in empirical disciplines, like for example, biology, cognitive psychology and neuroscience. There is a long and fruitful tradition of using category theory in software engineering, for example, in programming language semantics, modular specification or formal software validation. It would take too long to include that relation here, for those interested in category theory in computer programming can see [140].

It might be remarked, that at present, 2010, category theory is not a standard mathematical tool in any of the empirical sciences. The only exception to this can be found in quantum mechanics, where category theory is a well known theory for those working in quantum gravity and related fields[141]. However, quantum mechanics can not be considered an empirical science *tout court*, the empirical tests in quantum gravity is still an ongoing process. The construction of large energy particle accelerator, will help in this direction.

3.4.1 The origins: Rosen's (M,R-systems)

The application of category theory to biological systems began as early as the 50s, with the attempts of Robert Rosen within the frame of relational biology.

Rosen's work is often cryptic, which unfortunately has helped to portray the idea that it is indeed a very interesting fotmalism but too abstract and hard to understand, let alone to apply. Rosen's theory aims to transcend the reactive paradigm of simple physical systems, to a new post-Newtonian paradigm, able to deal with complex natural systems, or organisms using Rosen's terminology.

The metabolic, repair systems, (M,R)-systems, is a way of characterising systems quite independently of their physic-chemical constitution. In Rosen's words

> What remains then is an abstract pattern of functional organization, which has properties of its own, independent of any particular way it might be materially realized. Indeed, it is what remains invariant in the class of all such material realizations, and hence characterizes that class.[142]

Rosen was the first in seen the power of category theory to express, in a purely mathematical realm, patterns of relations between models, and between models themselves. This fact was exploited by Rosen who was looking for a theoretical framework for modeling in general, and in particular for living systems.

The theory of categories provides the necessary mindset to model the (M,R)-systems, a kind of systems that are independent of how they are physically realized. The (M,R) systems characterise the minimal organisation a material system would have to manifest, in order to realise a cell. Rosen acknowledged, of course, that there exist also in biology systems that depend upon the material details of a particular realization. In order to represent the (M,R)-systems it is need a general class of diagrams that the theory of categories provided.

In essence Rosen's theory is as follows. A completely autonomous living organism needs to encode all the information about the state of all of its catalysts (the M in (M,R)-system), and when necessary, makes the necessary replacements itself (the R in (M,R)-system).

The only way this information can be encoded, is in the connectivity of the network itself. The network must be constructed in such a way, that its entire connectivity must be uniquely encoded i.e. the natural(the living organism) and the formal system(the model of the organism) are isomorphic [143].

The formal system is relational, consisting of a category of interrelated components. A component is defined by a mapping $f : A \to B$. For example, the mapping f can represent an enzyme that converts a substance a A into a product B. Both metabolism and repair processes are modeled as morphisms, $M : A \to B$, $R : A \to H(A,B)$ respectively, where H represents the class of all the metabolisms from A to B, $M : A \to B$.

Rosen, with the (M,R)-systems, tries to provide a solution to the infinite regression in the iterative mappings M,R and H. They key issue is how the network produces the selector map g in order to avoid that vicious circle.

$$\begin{tikzcd} A \arrow[r, "f"] & B \arrow[r, "g"] & H(A,B) \end{tikzcd}$$

where the operator $g : B \to H(A,B)$ is produced by the network when the system is capable of autonomous invariance. The mapping $\beta : H(A,B) \to H(B, H(A,B))$ encapsulates the notion of metabolic enclosure and provides the solution to the vicious circle. For β to exit, given a metabolism f, the corresponding selector g must be such that $g(f) = f$.

$$\begin{tikzcd} A \arrow[r, "f"] & B \arrow[r, "g"] & H(A,B) \arrow[r, "\beta"] & H(B,H(A,B)) \end{tikzcd}$$

3.4.2 Category theory in Perception

In [144] Z. Arzi-Gonczarowski deploys a basic category theory tool for perception modeling purposes. A perception is a 3-tuple $< E, I, \rho >$ such that E and I are finite disjoint sets and ρ is the arrow $\rho : E \times I \to \{t, f, u\}$. The set E is made of elements of the external world, and I are mental concepts or internal connotations of the external world. Therefore the predicate $\rho : E \times I \to \{t, f, u\}$ is 3-valued. A mental concept i, can be a true, false or unknown, connotation of an external object.

Let E be an environment and $P_1 =< I_1, \rho_1 >$ and $P_2 =< I_2, \rho_2 >$ two perceptions over E. The mapping $h : P_1 \rightarrow P_2$ is a perception morphism (p-morphism) iff h is a mapping between the connotations I_1 and I_2 and definite truth values (t,f) are preserved by the p-morphism. Note that p-morphisms are the categorical morphisms of the category based on the collection of all perceptions with the same environment E.

3.4.3 Memory Evolutive Neuronal Systems

In order to understand the theory developed by the mathematician Andree Ehresmann and the physicist Jean-Paul Vanbremeersch, it is necessary to understand first, the approach to systems modeling based on category theory in which the theory is rooted.

In chapter 2 it was introduced the concept of state space, which is the space with all possible states of a system. The state space approach is non structured in the sense that a state space is merely a collection of points. For example, the state of a pendulum is given by the quantities representing the position and the velocity like in $\{(0,0), (pi/4, 0.1), (pi/2, 0.3)\}$.

This is in contrast with the configuration category approach, which imposes the structure of categories to the system's representation. The identity, and very importantly, the composition which is most important distinguishing feature of a category[13], are taken into account.

The theory of Memory Evolutive Systems (MES), produces a mathematical model for natural open self-organizing systems, such as biological, sociological or neural systems. In particular, the theory has been applied to neural systems: Memory Evolutive Neural Systems, MENS, which is a particular case of MES. Thus, MENS is a mathematical model based on the theory of categories for cognitive systems, and MES is a more general model, that deals with problems like emergence and self-organisation for autonomous complex hierarchical systems, in a formal-categorical way. MENS will be sketched in this heading.

It is important to note what the term *evolutive* means here. A system is evolutive when it evolves in time, which is precisely the definition of dynamical system. However, MES are dynamical systems but not in the same sense that physicists use. Indeed, by the choice of using the term evolutive, rather than dynamical, we must expect that the term evolutionary is being used within the context of Evolutionary Theory.

As it has been already said, category theory is based upon the notion of morphism, this is in contrast for example with set theory which relies in the concept of membership of a quantity x, in a collection or set X, $x \in X$. A morphism, for example, $f : a \rightarrow b$ in a category C, expresses one of the many possible ways in which the object a relates to the object b, in the context given by the category C.

We can easily find the underlying graph G provided by the morphisms in a category C. The objects in C are the nodes in G and the morphisms C are the edges in G. However, it might be emphasized that a graph and a category are not the same thing. In fact, a category can be seen as a (multi)-graph but equipped with something more: a composition law. In a category, compositions are calculated when two morphisms have identical domain and codomain. Category theory exploits this symmetry, by stating that different paths whose compositions have the same domain and codomain, may have the same meaning.

MENS starts with a multi-directed graph that models a nervous system, which is the lower model in the hierarchy of categories that is constructed in this theory. At such level, a graph will map the neural assembly as follows: for every real neuron, a node is created in the graph, and for every synapse between two neurons, there is an edge in the graph that connects the two nodes that represent the presynaptic and postsynaptic neurons.

A sub-system **Neur** is introduced in the theory to model the neuronal system. Of course, the nervous system is dynamic or evolutive in Ehresmann's words, and neurons and synapses may be created or disappear.

We can now start to model the system using the categorical approach that was referred above. The configuration category of such system at a given time t, is given by the category of neurons $\mathbf{Neur}_t$, and the trajectory of the configuration from the instant t to another instant t', is given by the partial functor $F : Neur_t \rightarrow Neur_{t'}$, which maps the configuration of neurons and synapses existing in **Neur**, at two different times.

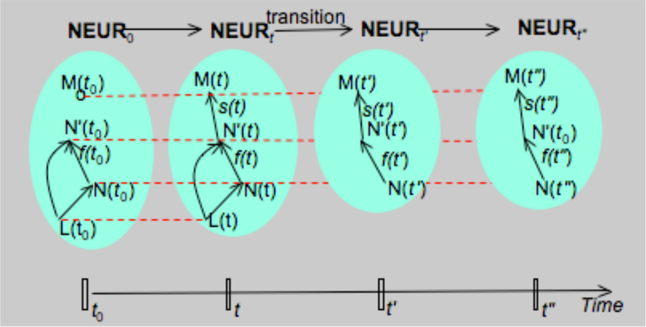

Fig. 3.6 The figure describes the evolution of the neural system during its life. For each t of the timescale is Time, we have the category $NEUR_t$ of neurons at t [12].

Now, we need to find the relevant patterns for **Neur**. A synchronous assembly of neurons P may work as a whole on another neuron N. In this case we say that N is a coordination neuron whose activity corresponds to the synchronous assembly of neurons P. It may happen that N can not be found, but there is a colimit **cP** at a superior level than that of **Neur**, that is to say, a cat-neuron **cP** of an immediately superior level may emerge. In that case, the cat-neuron **cP** becomes the colimit of P and also of other patterns P' synchronised with P.

Thus, MENS begins with a sub-system **Neur** that is initially modeled as a graph, which will suffer transformations that will be modeled through partial functors in the configuration category space, given by the category of neurons at different given instants $\mathbf{Neur}_t$. The system will evolve, and by the complexification process, higher cat-neurons representing more complex brain processes will rise.

It is important to remark that a cat-neuron in MENS is conceived as a multifold object because it admits several decompositions patterns P,P' . . . which may be not connected. This point is crucial because it establishes a powerful mathematical tool to explore the *binding problem:* How different functional brain areas for particular items, are coordinated in such a way to form widely distributed assemblies of neurons that represent complex items

Using Ehresmann's words "the cat-neuron characterizes the invariant that the different assemblies it binds have in common [12]". In this light, the cat-neuron colimit is a formalisation of the multiplicity principle, which is a generalisation of what neurobiologists call the degeneracy problem in neuronal code: given a single group of neurons they can participate in more than one representative function. The multiplicity principles will be fully explored in chapter 4.

The fundamental question of *how higher mental processes arise from the functioning of the brain?* is approached by the formation of increasingly complex objects. In this vein, neurons (**Neur**), category of neurons and mental objects (**ImO** Image of object O) are models of the brain at different hierarchical levels.

This is roughly how the complexification of a category of neurons is formed: category theory cope with the hierarchy of organisational levels

1. build the subsystem **Neur** which is represented as a graph having as nodes neurons and as edges synapses. The nodes or neurons in are cat-neuron of level 0.
2. synchronous assemblies of neurons form patterns and a cat-neuron of level 1 is created when a colimit to those patterns can be found.
3. By induction a cat-neuron of level k is the colimit of a series of patterns of lower cat-$\text{neurons}_{k-1,1}$

As the figure 3.7 shows, the category **Neur** is composed of neurons and models the physical structure of the brain and its elementary neural dynamics. The binding of a pattern P of neurons in a category of level 1 (**Neur**), becomes the mental image ImO of an object O. Thus, ImO is a *cat-neuron* of level 1. Progressively, the construction of a cat-neuron of higher level, 2 and so on, is established by the mental image of an object C formed by the juxtaposition of several objects O_i that the animal can already recognize.

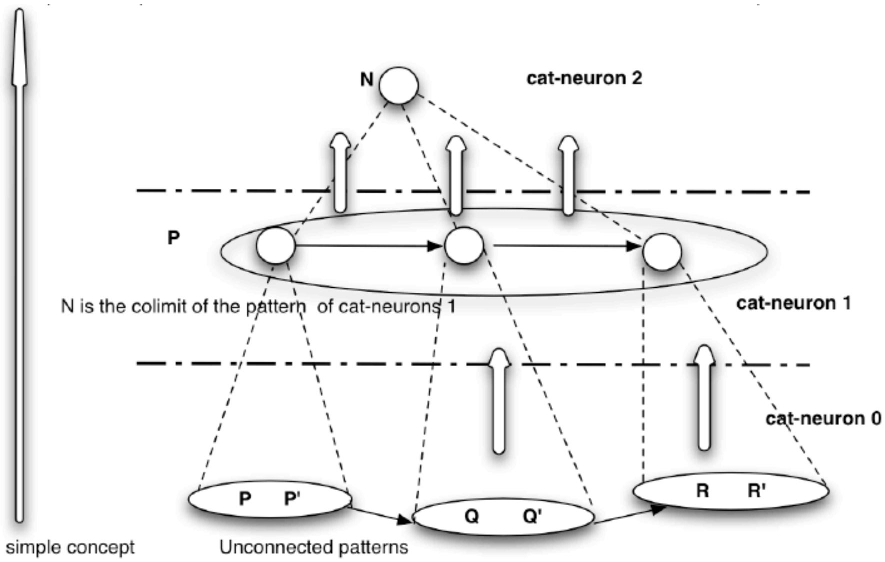

Fig. 3.7 The iterative building of cat-neurons binding patterns of cat-neurons of lower levels, models the computation on mental objects, to form more complex ones. It gives explanatory power for the construction of an algebra of mental objects

MENS does not explain how the patterns of neurons are formed, or how the category of neurons of higher complexity are created however, but it does provide fresh theoretical insights that deserve to be explored by either empirical tests or computational simulations. Indeed it tackles fundamental conceptual issues like, the phenomenon of emergence or the degeneracy principle in biological systems, and does so by formulating new mathematical structures, adapted to the structure and functioning of the brain/mind.

What is necessary now is for experimentalists to validate the theoretical hypotheses by measuring neural and cognitive activity. Only then, will it be possible to establish the real value of the theoretical claims and the mathematical structures proposed in MENS. Of course, the empirical measurement of purely mathematical objects in an ultra-complex system like the brain, is an extremely hard mission. As an example; since neither every neuron nor every stimulus can be studied at once, experiments that merely record the activity of a group of neurons upon presentation a stimulus, cannot definitively prove that only that stimulus creates the activity of the neurons that the stimulus seems to activate.

This thesis intends to make progress in that direction, and in chapter 8, it will be described a testable theory, that is able to explain how place cells in the hippocampus are created.

3.4.4 Category theory in knowledge acquisition and representation

In [135], Lawvere points out that category theory will be a necessary tool in the construction of an adequately explicit science of knowing. In this chapter I sketch how this idea has progressed since Lawvere's claim. For example, in this heading we will see how "the initial plausability" claimed by Lawvere has been transformed into real implementations, for example, in the construction of ontologies by adaptive neural networks 3.4.4.1, in categorical models for family resemblance and contextuality, as well as in semiotics 3.4.4.2, or in the field of cognitive psychology, for example in concept formation and inferential abilities in humans 3.4.4.3. Let us see some of these implementations.

3.4.4.1 Category theory for ontology building

Healy et al.[145] describe knowledge as "a category representing an ontology formalized in a modular hierarchy of closed knowledge fragments, or theories... Expressing an ontology as a mathematical category of theories allows the machinery of category theory to be harnessed in understanding the semantics of adaptive, distributed systems".

Category theory is used here as a mathematical language for ontology building. The rationale is as follows: *i.* concepts form a structure, so it is possible to build the category of concepts and *ii.* structures can be mapped onto each other, so we can find a functor that preserves the structure of two different category concepts. Thus, a model is a functor where the domain is the classifying category and the codomain any category with similar structure.

```
Concept T0
  sorts Points, Lines
  op on: Points*Lines →Boolean
  Axiom Two-points-define-a-line is
    forall(x, y:Points) ((x not = y) implies
      (exists L:Lines) (on (x, L) and on (y, L) and
        ((forall m:lines) (on (x, m) and on (y, m))
          implies (m = L) ))
end
```

Fig. 3.8 The theory T0 defines a simple geometry of points and lines which are primitive types and the relationship between them is given by the axiom “Two-points-define-a-line” [13]

This categorical approach to ontology building, consists of defining the category of concepts **Concept**. In figure 3.8 it is shown how the concept T0, which is a geometry theory, can be used for forming an ontology within the category **Concept**. Morphisms in **Concept** relate theories, for example $s : T0 \to T1$, expresses the partial association of the syntax of a concept $T0$ with concept $T1$.

Colimits are used in the ontology in order to provide greater specificity, as figure 3.9 shows. The dual of colimits, limits, provide the opposite direction, that of abstraction. But so far, only syntactic properties of theories expressed in predicate calculus have been tackled. In [13], Healy at al. go further, and in the second part of the paper, they provide a mathematical semantic model, in which these theories are grounded to network components. Thus, the syntax of the geometric theories in the example above is now given in terms of the architectural structure of neural networks.

A network architecture A, formed of a set of neurons, together with an array of the connection weight values w of that set, is modeled as the category $N_{A,w}$. An object of $N_{A,w}$ is defined by a pair (p_i, w), where set $p_i = 1, 2, ...n_k$ is the nodes of A and w represents the set of output values for p_i connections. A morphism $m : (p_i, w) \to (p_j, w')$ of $N_{A,w}$ is defined by a set of connection paths (synapses) between the set of nodes (neurons) and their weight states, (p_i, w), and the nodes (neurons) and their weight states, (p_j, w').

Functors now transport the invariant structure across the category **Concept** and the category $N_{A,w}$, $M : \textbf{Concept} \to N_{A,w}$ The main idea is that

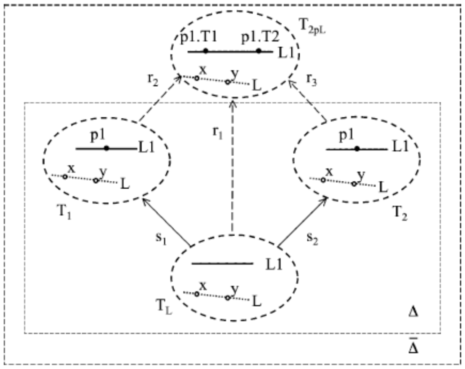

Fig. 3.9 The figure depicts the colimit T_2pL of theories T_1,T_2, and T_L. The colimit theory glues the theories T_1 and T_2 along their common sub-theory T_L. For the dual treatment of theory abstraction using limits see also [13]

learning can be modeled as a transition between categories. A functor is used to model the structure-preserving associations between categories. On the other hand, colimits express the learning of more complex concepts through the re-use of simpler concepts already represented in the connection weight memory of a neural network. The categorical approach to concepts it makes possible to model the declarative semantics of systems, such as neural networks, which are distributed over many components.

3.4.4.2 Category theory in semiotics

In [146], Neuman and Nave address two major problems in modeling concept formation: family resemblance and contextuality, borrowing concepts from category theory, in particular pushout and its dual, pullback. The formal theory for concept formation is based on two steps, first there is a bottom-up process that is modeled as a pushout, and then a top-down process which is modeled as a pullback.

Let us see, how the authors explain the concept formation of “Dog” in a child. Sign-mediated concept formation starts in the child when the mother assigns the same sign to at least two different objects. For example, the mother may refer to the Great Dane and the Chihuahua as “Dogs”. We can express this with the following diagram.

Next, a third object, for example, a German Shepherd, is introduced to the child by the mother. The resulting diagram tells us that a Chihuahua is like a German Shepherd, and a Great Dane is like German Shepherd. These are local similarities, but there is an obvious global property, the three objects are all dogs, can not be inferred directly from the diagram bellow.

$$\begin{tikzcd}
 & \text{Great Dane} \arrow[d] \\
Chihuahua \arrow[r] & Dog
\end{tikzcd}$$

$$\begin{tikzcd}
\text{German Shepherd} \arrow[r] \arrow[d] & \text{Great Dane} \arrow[d] \\
Chihuahua \arrow[r] & Dog
\end{tikzcd}$$

In order to model the global property of family resemblance we need to define Dog*, which may be constructed trhough the pullback. The dual top down process is modeled with the following pushout diagram.

$$\begin{tikzcd}
Dog{*} \arrow[rrd, bend left] \arrow[rd, dotted] \arrow[rdd, bend right] & & \\
 & Dog \arrow[r] \arrow[d] & Chihuahua \arrow[d] \\
 & Greatdane \arrow[r] & Germanshepherd
\end{tikzcd}$$

The same authors have applied category theory to understand the mechanisms of coding that living systems posses, and in particular genetic coding [14].

This work may be labeled as biosemiotics, which is the field of theoretical biology that studies the nature of coding and communication of signs, in living organisms.

The informational approach followed in Neuman's work, is different from the engineering approach based on Shannon's theory. The divorce between information and meaning, in the engineering view, is clearly expressed in Shannon's words: "Frequently the messages have meaning; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem [147]".

For Neuman, biological systems are not information-processing systems but "meaning-making systems". Neuman defines meaning as "the effect that is produced via semiosis through interaction between at least two parties". This approach is sympathetic with the view defended in this thesis in chapter 7, that states that brain coding is of form and meaning.

One of the dogmas in biology avows that the flow of information is only one way: from DNA to RNA to protein. However, we know today that the dogma is not entirely true. It has been found a "reverse transcription" flow of information, from RNA to DNA, that could be at the basis of the mechanism by which, for example, RNA tumor viruses infect normal cells and turn them into cancer cells.

In addition, it has been hypothesized [148], that under certain experimental conditions, DNA can be translated into protein without the mediation of RNA.

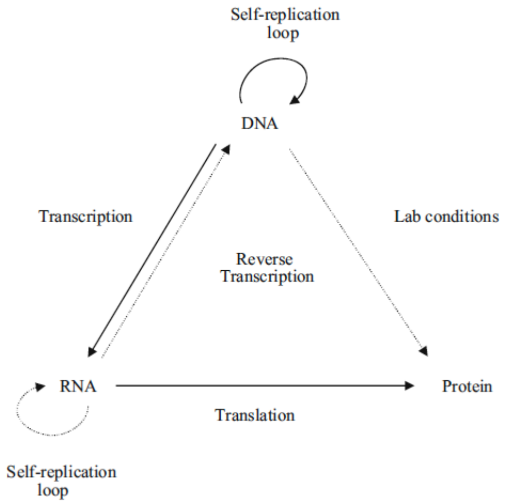

Fig. 3.10 Information flow in the genetic system as depicted in [14]

The paper produces a mathematical solution using category theory to the following problem: how systems translate information encapsulated in DNA into proteins along two different paths: (1) from RNA to DNA; and (2) from RNA to protein.

The mathematical approach to this problem relies on the concept of isomorphism. By the definition of isomorphism given above, we know that a map f is an isomorphism if there exactly one map g such that $h = g \circ f$ where $g = h \circ f^{-1}$

For a genetic system, it is possible to establish an isomorphism between the four DNA bases and the four RNA bases which is represented in figure 3.11 with the arrows f and f^{-1}.

Moreover, the above figure 3.11 allows the authors to reformulate in mathematical terms the question of how information flows in the genetic system. In this view, the DNA does not contain information, in the sense given by Shannon, rather it has value. The value is "the effect, mainly the generation of protein, through the symbolic mediation of RNA, which constrains the different possible interpretations of the DNA for self-replication".

$$\begin{tikzcd}
 & \text{DNA} \arrow[dl, shift right, "f^{-1}"'] \arrow[dr, dotted, "g"] & \\
\text{RNA} \arrow[ur, shift right, "f"'] \arrow[rr, "h"'] & & \text{Protein}
\end{tikzcd}$$

Fig. 3.11 The flow of information from DNA to RNA and protein, is represented with the arrows f^{-1} and g. Thus, given the arrows f and h we can detrmine f^{-1} and g

3.4.4.3 Category theory in cognitive psychology

There are psychological categories, that at first sight, do not have too much in common with the mathematical categories described in this chapter. In "To cognize is to categorize revisited: Category theory is where Mathematics meets Biology [149]", I briefly sketch the possible parallelism between both kinds of categories. Let us look at this with more detail now.

On the one hand, a mathematical category is an algebra of mappings among things that can be composed and where composition is associative. On the other hand, in psychology, categorization is referred to as the process by which distinct entities are treated as equivalent. The output of the process of categorization are categories, which are concepts that we necessarily build in order to cope with reality.

Within this view, category and concept are assumed to have identical meaning. Mental concepts or categories are created based on the *principle of similarity:* an object falls under a concept if we can determine that the object is sufficiently similar to the other category members. Thus, an instance i falls under a concept or category C *iff* $\forall featuref \in C, f(i)$is true. This theory of concepts, as categories of similar items, has been predominant in theory of mind, since the seminal work of Eleonor Rosch [150], [151]. For a detailed description of this view and its criticisms see [152].

William C. Hoffman[136], shows that the basic structure is the same for both kinds of categories. He establishes a correspondence between psychological and mathematical categories, based on equivalence. Such an association allows us to set the basis for a mathematical investigation of cognition, perception and emotion, which are the fundamental components for consciousness. The mathematical theory of categories, makes available tools for the deduction of hypothesis for structure and function in neuropsychology. The intrinsic categorical nature of the mind is clearly stated in this quote:

> the computation involved in basic human affairs is primarily topological or categorical rather than computational in the traditional computer-science sense of that term. Every percept is a geometric object in the mathematical sense. And concepts and emotions, to be sufficiently general, must be categorical.

Hoffman uses the category of sets **Set** and the category of topological spaces **Top**, the former models "trains of thought" and the latter is able to cope with percepts, concepts and emotions that occupy the mind.

Philips et al. [137], use category theory to show that, human inferential abilities like transitive inference and class inclusion, involve the dual category theory concepts, product and coproduct respectively. Children around five years of age develop what is called *Transitive Inference* which is for example, given that Bertrand is taller than Ludwig, and Ludwig is taller than John, one may infer that Bertrand is also taller than John. Class inclusion develops later in children and consists of the ability to discern between the cardinality of classes and subclasses, for example there are more animals than cows.

Category theory shows that these abilities can be formally connected. Transitive inference can be modeled with product, and Class inclusion with its dual, coproduct. This fact would explain that these two reasoning abilities have "similar profiles of development", because they involve related sorts of processes, namely product and coproduct

In a later paper, [153], Philips et al. use category theory to explain systematicity in human cognition. Systematicity is defined as "the property of human cognition whereby cognitive capacity comes in groups of related behaviours". In this work, systematicity is perceived as a necessary consequence of a higher-order theory of cognitive architecture, this would make either the symbolist and the connectionist approaches inadequate to explain systematicity. While connectionist and symbolist theories are both first-order, category theory, and in particular adjunctions, allows us to deal with the higher order structure, which is more appropriate for cognitive functions modeling.

Definition 3.19. An adjunction consists of a pair of functors, $F : C \to D$, $G : D \to C$, and a natural transformation $\eta : I_C \to (G \circ F)$, such that for each C-object X and C-arrow $f : X \to G(Y)$, there is a unique D-arrow $f : F(X) \to Y$, such that the following triangle commutes:

$$\begin{tikzcd}
X \arrow[r, "\eta_X"] \arrow[dr, "f"'] & G(F(X)) \arrow[d, dotted, "G(f)"] \\
& G(Y)
\end{tikzcd}$$

An adjunction is a natural transformation between functors, natural here means particular realization of the functors. By analogy, systematicity "is natural in that it does not depend on a particular representational scheme.

Hence, the explanation does not depend on ad hoc assumptions about internal representations”.

In their conclusions, the authors suggest that a more integrative theory of cognitive architecture could be attained by including acquisition of representations which is not tackled in the paper. Indeed, why an object belongs to a particular class and not to another is not explained but assumed. Finally, the incorporation of category theory into the Bayesian theory is stated as a promising future line of research. Undoubtedly the creation of a category of Bayesian models will need of ingenious mathematicians.

Chapter 4
Elementary principles in cognitive systems modeling

This thesis is about thinking, in all its ontological, epistemological and operational facets. It deals not exclusively with the process of thinking, but also with conditions which must be created and laws that must be fulfilled, in order to think about the process of thinking.

Arguably, human cognition may be largely a emergent phenomenon. How cognitive functions "emerge" from the nested hierarchical architecture of neural systems, is a question that has been posed with a myriad of approaches.

Concepts that traditionally belong to the philosophy of science like emergence, reductionism or principle, have started to receive some attention from researchers in complex systems.

4.1 Introduction

In this chapter I propose a framework for modeling in cognitive science. The brain (and mind) is a ultra complex system, we can not expect to find the definitive model that encompasses the countless factor that mediates in mental processes such as like language, memory or perception.

Section 4.2 explores the concept of reductionism. I will define reductionism and I will show that, for the purpose of modeling complex systems, the reductionist approach can only give us a fragmentary picture of the system. Section 4.3 discusses formalisation and its limitations.

Emergence is introduced in section 4.4, this concept is explored in a rigorous way, trying to provide a formal description of emergent properties within a context relevant for cognitive scientists. Section 4.5, describes three principles for modeling in cognitive systems -i.e: locality, hierarchy and multiplicity.

Well known problems on modeling in cognitive and brain system, for example frame problem or degeneracy problem, are shown here as inherent aspects of laws or principles that certain systems hold, and need to be formulated before to the modeling act.

It ought to be remarked that in this chapter, modeling is not seen as quantitative prediction data, that kind of models were described in chapter 2. Here, I present a survey of modeling in a more basic or fundamental way. Modelers should not be satisfied to merely demonstrate a fit to data, the goal of modeling is to increase our understanding of systems [154].

4.2 On reductionism

Reductionism is a term that has gained considerable bad press within certain *cultural milieu.* However, reductionism can be seen as just another *-ism*, carried out by the reductionists, who study complex phenomena by over simplifying them.

There is a well-known fear of reductionism in the social sciences. Admittedly, reductionist statements ornamented with some obscure mathematical terminology by some, has served to brutalize social reality, and minimize environmental influences for the most self-serving reasons.

As an example of this, there is a very colourful dispute, published in the journal *Mathematical Intelligencer*, between the mathematician and number theorist, Neal Noblitz and Herbert Simon, Nobel prize in Economy[155]. At the origin of the argument was the politologist and Simon's friend, Samuel P. Huntington, the same that wrote later on, in 1993, The Clash of Civilizations, an essay that was very influential in G.W. Bush's State Department.

In his 1968 article [156], Hungtinton's collateral effects only caused the irritation of some sensitive readers with mathematical background.

Hungtinton defined the following equations in order to support his conclusions about the impact of modernization in developing nations:

$$\frac{SM}{ED} = SF, \frac{SF}{MO} = PP, \frac{PP}{PT} = PI$$

where SM is *social mobilization*, ED is *economic development*, SF is *social frustration*, MO is *mobility opportunities*, PP is *political participation*, PT is *political institutionalization* and PI is *political instability.*

The problem is that Huntington never formally defined what the variables used in those formulae, *social mobilization*, *political instability*, exactly mean. He neither provided the reader elementary facts like the units in which the variables must be measured or their range of possible values.

Hungtinton's conclusions, supported by trivial and badly formulated mathematics, lead to surprising consequences. For example, Philippines was a "stable" country and France was not, or South Africa, during the Apartheid was a "satisfied society" [157]. If we acknowledge the big mismatch between the mathematical equations and the social and economical facts that the formulae pretend to explain, the conclusions are not surprising at all, but the consequence of bad mathematics. This was pointed out by Neal Koblitz in "Mathematics as Propaganda" [158].

Despite the evident bad formalisation practised by Huntington, Herbert Simon came to the politologist's defense with a paper on "Some Trivial But Useful Mathematics"[159]. A vehement exchange continued and other mathematicians like Saunders Mac Lane and Serge Latig, joined Koblitz against Hungtinton's simplistic reductionism [160].

Reductionism and mathematization, are a danger only when used with spurious interests and limited knowledge of the mathematical structures that are introduced in the explanations. Hungtinton's case is a conspicuous example of this; alas it is not an unique case. The psychoanalyst Jacques Lacan, suggested, and convinced some people with no mathematical background, that Don Juan was a compact topological space [161].

Having shown "the dangers of reductionism and mathematization in social sciences',' we must not misunderstand them. In social sciences and in cognitive science too, the ambiguous and strictly verbal account of the facts still prevail.

There seems to be some sort of irrational or pathological fear of reductionism and mathematization within the realm of social science and cognitive psychology. The following unattributed quote expresses this superbly

> If in the physical sciences, one is able to successfully reduce complex phenomena to a simple rule or model, one is awarded the Nobel prize, the reward for a similar attempt in the social sciences is to be pilloried in the New York Review of Books.

With this in mind, we can readily see the necessity of reconfiguring the concepts of reductionism and formalisation on more rigorous ground.

4.2.1 What is reductionism?

A reductionism approach explains a phenomenon C, through a simpler one, S. It is by that account that reductionism gives reductive explanations. It might be reminded that there is no understanding without reduction.

In the philosophy of science, to reduce is to express laws and rules of secondary sciences in terms of primary ones [162], i.e. the secondary science of Thermodynamics was explained using concepts and laws belonging to Newtonian mechanics, the primary science.

In mathematical logic, to reduce is considered as equivalent to deduce. Given a set of premises P we deduce a new statement s, however, s is not really new because it was already contained in P.

Reductionism can be understood in several ways. For example higher/lower or partial/total reductionism, both cases will be explored next. Bunge[163] establishes a differentiation between higher reductionism and lower reductionism. Indeed, the etymology of to *reducere* is to lead back, and not necessarily toward smaller scales. One example of higher reductionism is to state that a calculator is intelligent because it is fabricated by an intelligent person. An example of lower reductionism is to explain the intelligence of an animal as a set of automatic mechanisms.

On this view, emergent phenomena can be more or less complex than the elements of the substrate that give rise to them[164]. *Behaviorism* could be seen as lower reductionism: a complex mental function, -e.g: emotion, may be reduced to a set of preconfigured states, e.g. the states of an automaton, which are much simpler than the mental functions.

Both approaches are reductionist, but it is important to note that the higher reductionism is, epistemologically speaking, more untenable than the lower one, because it leads to obscure explanations about complex phenomena. In other words, it explains something by means of something else even more complicated *ignoti per ignotum*, as in the above example: a calculator is reduced to something more complex, a person.

The lower reductionism, on the other hand, constitutes the very basis of the modern science, which roughly consists on the study of any object by reducing it to its elementary components and properties. A distinction between total and partial reduction is appropriate at this point. For example, particle mechanics is deducible from quantum mechanics or solid rigid kinematics is deducible from solid rigid dynamics. Of course, the total reducibility is not always applicable. As an example of this, biology can not be reduced to physics, but partially reduced. Theories that are totally reducible, are isomorphic [165].

In short, let *Phy* and *Bio* be two theories, and let A a set of assumptions not included in either *Phy* or *Bio*. Then *Bio* is fully reducible from *Phy*, iff *Bio* follows logically from *Phy*. *Bio* is partially reducible from *Phy* iff *Bio* follows logically from $Phy \cup A$ and not from *Phy* alone.

4.2.2 Two obstacles for reductionism

There are two main problems concerning reductionism. One is of ontological nature or *what are the elementary components belonging to the complex studied object?* and the other is the epistemic problem or *how well the lower components of a complex object reflects its nature and essential properties?.*

These two problems are indeed almost perennial matters in the history of philosophy and science. Therefore no magic recipe or simplistic formula can be expected here, rather there is a need for a theory, within a general framework, where the questions exposed above can be conveniently treated.

In order to understand a complex phenomenon, is by no means always enough to decompose it in its atomic components. Atomism, as a matter of fact, subscribes a strong bottom-up causation placed in a temporal framework where time is unidirectional.

In complex systems, for example, biological ones, not only the physical factors have to be considered. Ecological factors are necessary for the obtention of powerful explanations. Even though the environmental or cultural aspects are, of course, ultimately sustained in physical causes, such causes could be inaccessible or irrelevant. Causality is not always exclusively unidirectional, there is also a downward causal flow. This point was sketched more fully in chapter 1. We are very close in science to reduce as much as we can, the next step must be a work on synthesis. A full understanding of complex systems will come only form integrating the reductionism methodology into a more all-inclusive approach.

4.3 On formalisation

Descartes' invention of analytical geometry supposed an astounding conceptual revolution. The universal order, with its objects and laws, could be mapped and studied within the cartesian concept of space.

This happened a long time ago, but this should not make us to take for granted the dramatic change that supposed the inclusion of algebra in geometry. The geometric forms, which were thought of, in the pre Descartes era, as synthetic mental ideas, were now able to be formulated as mappings in coordinate axis. This paved the way, three centuries later, with the work of Lobachevski and Bolyai, to non-Euclidean geometry and its application in quantum physics [166].

Having said this, the cartesian axis seems a very feeble skeleton to sustain the whole body of the universe laws. In words of David Bohm [167], reality is a multi-dimensional reality and only under very concrete conditions can be satisfactorily simplified to three dimensional reality.

Reckoning this as true, it exists the danger to misunderstand ontology with epistemology, or said plainly, between on the one hand, the things that populate the world, and on the other the methods to know those things.

The thesis defended here is that formal systems, expressed by means of diagrams or equations, can be an extraordinarily efficient and successful mechanisms to explain phenomena and represent features of the world. The point to be retained is that formalisations are methodological tools and not ontological simplifications. For example, a map of London is not the same as the city of London, however, the map is a formal representation of a physical system, or to put it a different way, there is a homomorphism that preserves significant aspects of the system that is being modeled and the model itself. The fact that the map captures the structure existing in the city is evident in this.

A map has to be simpler than the world it models. Of course, this means that every so often the map may be wrong, in that case, the mapping is not homomorphic and the structure is not preserved.

When we succeed in representing an external object of the real world in our head; a map, or to be more precise, a homomorphism that preserves the common structure of both the external object and the mental object has been established. Using the theory of categories, if we specify the category M of mental objects and their homomorphisms and the category E of external objects and their homomorphism, a functor $F : M \to E$ constitues a model of the external world.

Let P a physical system, given by a domain P_D and a set of relationships P_R, and let M a mathematical system, given by a domain M_D and a set of relationships M_R. A homomorphism h, is a mapping from P_D to M_D that preserves the structure in an appropriate way [157]. This psychophysical homomorphism constitutes a small scale reproduction of the natural laws that rule the physical world. In the words of Dehaene, "we carry in ourselves a universe of mental objects whose laws imitate those of physics and geometry" [168].

When moving from the *hard sciences* -e.g: physics, to *less hard sciences*, e.g: psychology or sociology, the formalist credo does not see as well received as it is in physics. Needless to say, putting into our daily experience and intuitive concepts, into formal terms is not an easy task by any means.

However, the formalist endeavor in science -hard or soft ones- is unstoppable, and as formal descriptions of the phenomena studied progress, ordinary accounts are abandoned. Clearly, this replacement of the ordinary for the mathematical-formal is a constant in the books of the history of science. Quoting R.D. Luce:

> *and as honest scientists, as far as we disentangle the structure we need to begin to describe it in formal terms.*[169]

Although the fear to reductionism is not totally irrational as it was shown in 4.2, the unease, or aversion of some cognitive scientists with the formalisation of the cognitive sciences is unjustifiable and counterproductive.

To know and be able to formalize the laws and the essential properties that rule the cosmos does not make it less beautiful. Said in Lee Smolin words

> *Scholars in social sciences seem to be comfortable with the belief that biology and natural selection successfully delivered homo sapiens into the upper paleolithic and then was abandoned to the ministrations of culture [170].*

4.3.1 The limitations of formalisation

Next, I sketch six general obstacles to formalisation. In what it follows I hope to shed some light in a crucial point: formalisation has, not only an epistemological dimension that can be understood as a method to gain some knowledge about the world, but also an ontological dimension; formalisation is an inescapable cognitive activity, intrinsically related to perception.

4.3.1.1 Definitions in the dictionary are circular

When asked by a child, for example, about the definition of one word, we easily realize that the very simple questions may require long and complicated answers.

Let us obviate here all the difficulties rooted in the process of categorisation of everyday concepts and in its formulation in natural language[152], and focus in the definitions of the concepts.

The definitions that one finds in the dictionary are sometimes circular and ambiguous. In natural language, a circular definition is a description of the meaning of a word that assumes a prior understanding of the term being defined. For example:

fast -adj. : *swift*; *quick*; *speedy*
swift -adj. : *fast*

Or another comical but realistic example is:

see -verb : See *see*

The basic component in a dictionary are the lexemes (words) which as it is shown above, are sometimes ill-defined with circular definitions.

In mathematics is possible to start with more basic concepts than lexemes. The indecomposable elements in mathematics are objects that can be seen as elementary particles with a minimal amount of ambiguity in their definition.

The indecomposable elements hold the Indecomposability Principle, established by Brouwer that states that the continuum cannot be partitioned into two nonempty pieces [171].

The idea I am defending is simply to promote the use of formal languages -e.g: mathematics rather than natural languages -e.g: English, in cognitive and brain science. A good reason for taking this strand is that the basic concepts of formal languages are infinitely closer to indecomposability than natural languages. [172]

It must be emphasized that this approach is radically different than the language of thought hypothesis (LOT), proposed by the philosopher Jerry Fodor, that postulates that thought and thinking take place in a mental language[173].

LOT is linguistically bias, thinking is seen as the product of syntactic operations defined over representations.

The problem with LOT is that as it focuses exclusively in linguistics, and obviates other languages like mathematics, which is an extremely efficient source of new concepts and tools to understand reality (thought included).

4.3.1.2 Obstacles to formalisation: The plurality of definitions

Formal logic definitions are not always enough to fully describe a mathematical concept, indeed in the textbooks of mathematics, the definitions of new concepts are self-included in definitions, but they usually need of lemmas and examples in order to be properly defined.

It ought to be noted that in mathematics, unique definitions for mathematical concepts do not exist, not even for those as important as function or structure. This is because these are concepts used in several areas of mathematics, in many different ways, so no single definition of function has been universally adopted. For example, the concept of morphism which recurs in much of contemporary mathematics and in particular in this thesis, has different denotations. For example, in set theory, morphisms are functions; in linear algebra, linear transformations; in group theory, group homomorphisms; and in topology, continuous functions.

This lack of unity in the definition, even for some of the most fundamental concepts, ought not to drive us to wrong conclusions about the useless or impossibility of the formal program.

Au contraire, the obtention of formal definitions provides a common setting where everyone knows what the others are intended to say, even if restricted to some particular domain situation or context.

4.3.1.3 Obstacles to formalisation: The sensitive coding

Another problem alleged against formalisation is the sensitive coding (we use the term coding borrowed from informatics as equivalent to formalisation).

Obviously, when encoding any process or set of data, some relevant part of information can be lost. This occurs because the encoding does not capture the invariance of the object that is being coded, in consequence the structure is not preserved through the codification process.

The coding process can be seen as a morphism, in order to capture the invariance, such morphism should be an isomorphism.

4.3.1.4 Obstacles to formalisation: The Gödel's incompleteness theorem

The first and second Gödel's incompleteness theorems are used for some scholars as the kill all argument against formalisation[174].

Gödel's first incompleteness theorem says that any axiomatisable theory that is adequate for arithmetic is incomplete. This implies that any attempt to capture some arithmetical aspect of the world using a formal system, will not be able to prove all the truths about that aspect of the world. In the same vein, Skolem's paradox shows that no formal axiomatised system can capture our intuitive notion of set.

This is far from surprising, especially if we are aware of the simple fact that the formal systems that we generally construct cannot completely capture the aspect of reality that they describe. This is analogous to the incompleteness of arithmetic, since our formal axiomatisation of arithmetic cannot capture everything about the mathematical system itself.

One main lesson to get from Gödel's incompleteness theorem [1] is that the foundational relation of maths to mathematical logic is ill-founded. Indeed, in order to verify the correctness of a theory T, it is necessary to build a stronger theory T'. This leads to an infinite regression because the completeness of T needs of $T' \supset T$ which needs of $T'' \supset T'$ and so on. The Gödel's theorem demonstrates the impossibility of establishing the non contradiction of arithmetic solely by methods borrowed from arithmetic or weaker theories[175].

The Gödel's deadlock means that every time we build a formal system, in order to test its completeness, one stronger formal system has to be additionally built. As there is not a finite method to guarantee the completeness of any formal system, the gap between formalisation of a system and the system itself seems abyssal.

But this is not any impediment at all. In both Set Theory and in Computer Programming textbooks, formal definitions abound, that exemplify this infinite regression. The axioms of Peano for the natural numbers is a good example of what we are saying, the set of natural number N is such that satisfies:

$$\begin{aligned} & 1 \in N \\ & \nexists n : succ(n) = 1 \\ & \forall n \in N : succ(n) \in N \end{aligned} \tag{4.1}$$

The formal definition of the natural numbers, one of the most basic structures, is recursively defined.

In addition, it is worth noting that the concept of formalism proposed in this thesis, is wider and more general than the that used in the logical tradition that focuses on formal consistency and completeness, which are purely syntactic concepts.

Formalisation can be seen as the cognitive process that necessarily mediates in the grasping of the regularities existing in the world; and formalism as the output of such process, bringing together the mathematical entities, that reflect the structural invariance between the real system and its formalisation.

---

[1] *First Incompleteness Theorem*: Any adequate axiomatizable theory is incomplete. In particular the sentence "This sentence is not provable" is true but not provable in the theory. *Second Incompleteness Theorem*: In any consistent axiomatizable theory (axiomatizable means the axioms can be computably generated) which can encode sequences of numbers (and thus the syntactic notions of "formula", "sentence", "proof") the consistency of the system is not provable in the system.

4.3.1.5 Obstacles to formalisation: Formalisation as opposed to intuition

It is necessary to abandon the platonic idea of mathematical forms and theories as an extramundus thing. Gödel himself is a champion of the Platonism in mathematics, not in geometry, which he conceives as physically grounded in the world, but with Set theory. The objects of Set theory, are Platonic Forms because they are non-physical yet objective. Due to the intuition faculty, it is possible to perceive truths about these objects with which we have no physical interaction.[176]

Let us for a moment assume Gödel's platonism credo for set theory. The Platonic Forms, to which we have access only through intuition, are committed to the law or order that we perceive in nature, it follows then that those intuitions must be governed by rules as well. Both intuition and set-theoretic objects, are mutually constrained by the morphological laws or legality using Jean Petitot's notation. [177]

It is vulgarly assumed that intuition covers all that is not formalised. At this point is convenient to shed new light on the concept of intuition. Intuiton and formalisation are not on opposing sides, rather intuition is at the core of formalisation, intuition naturally converges into formalisation. There is a tendency to understand formalisation as a process that progressively limits the field of the intuition which we can define as that that is not (yet) formalized. This is a dangerous myth that must be debunked. First, even in the case that formalisation was *une chôse pour des elites*, this elite should be listened to, at least, as much as those lacking the skills that the formalists posses. Second, there is no such a thing as unstructured cognitive state. Any mental content, in order to be attainted, needs of the intermediation of perceptual schemas (forms) that configures the agent's perception.

Mistakenly, formalisation is seen as a prerogative of a small elite composed by those that find themselves comfortable in mathematical and logical arena in opposition to intuition, which as a natural gift that everyone has, no pedigree nor skills are required. It might be remarked here, that all explanations or descriptions, are rooted in some context or theory. The definition of any object makes sense, only as far as is related to a theory, which must be formalised (mathematised) to be operational.

The cognitive agent is always mapping the outer repeating patterns into its own inner structures. However artificial it may seem, formalisation resides naturally and primitively in the cognitive agent, formalisation supersedes intuition. The world out there, is nomologically (legally in Petitot's terms) organised, and the only option for us, as cognitive agents, is to interact with it.

4.3.1.6 Natural thought is circular and equations are linear

Formal scientific explanations propose, contrarily to ordinary ones, are general and perfectible models.

Linearity of natural language and proof theory in logics is confronted with structural circularity. Through effective construction of "ideas", "signs", "percepts"... the cognizer can open the door of the universe. Mathematics or any other formalism, becomes relevant to science when the scientist is able to disentangle the structure in the observed phenomena. The symbols that the formalism is made of, are necessary to keep track of those structures captured in the scientist's head. The scientific study of cognition will, of necessity, require a formal language able to seize and to express, in an systemic and comprehensible way, the structural invariance residing in the phenomenon of cognition.

The words of the father of cybernetics will put an end to this section.

> "power to operate with temporary emotional symbols and to organize out of them a semipermanent, recallable language. If one is not able to do this, one is likely to find that his ideas evaporate from the sheer difficulty of preserving them in an as yet unformulated shape." [178]

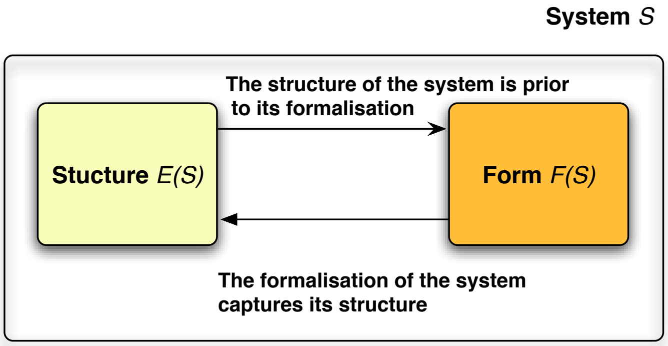

Fig. 4.1 The structure is prior to its formal and the formalisation captures the structure of the system

4.4 Emergence on Systems Modeling

The modeling cycle in science, necessarily begins with a reductionist approach, creating the simplest possible model. The modeling process generates an understanding of the underlying structures, as components are represented with mathematical structures. The minimal model then grows in complexity, phenomenological descriptions can not be captured by hypothesis that solely rely on the components' properties. The informal notion of "complexity" has a number of approaches and definitions. For example, Kolmogorov complexity for one object, is the amount of information required to describe the object.

Here we are interested in organizational complexity. The organizational complexity, in order to be studied, needs a scale hierarchy representation of the system. In this view, a larger scale system may impose boundary conditions upon the faster dynamics of lower scale systems[179],[180].

Emergence may be seen as a manifestation of complexity. However, emergent property do not necessarily necessarily arise in complex systems. There are many examples of emergent phenomena, some of the most common are transitions between solid, liquid and gaseous states, honeycombs, life, markets, economies, cities.... Interestingly, in the Wikipedia, intelligence is included as an emergent phenomenon.

As this thesis is about cognition, we will explore the connection between emergency and complexity. The mind can be studied as an emergent phenomenon that arise, in last term, from purely physical processes that occur in the brain, which is a system with a tremendous organizational complexity.

4.4.1 A few notes on complex systems

In order to be effective in complex system modeling, we can not rely on disquisitions about how complex a system is by setting up which properties are global and which are local. The main problem with this procedure is that it assumes three things that we can not take for granted,

- the components of the system have clear cut borders
- the frontier between the system and the environment is a clear cut border
- the properties of the system are always observable and measurable

Perhaps for a better understanding of complex systems, we must outline the work of scholars -e.g: System theorists, who use system as a major theoretical construction in their explanations.

Ludwig von Bertalanffy, father of general systems theory, defines a system as a "set of elements standing in interrelations". In this view, Mario Bunge, has accomplished a colossal work, writing a veritable opera magna in eight volumes that goes from Ontology to Sociology.[165]

Bunge, leaded a formidable project[2] called Exact Metaphysics. In essence, Bunge is a systemist, the motto "everything in the universe is a system or a

---

[2] This project has not been completely abandoned. Indeed the Society for the exact philosophy organises every year a congress

part of a system" and "every system has global properties that its components lack", has been formally defined in Bunge's work with some recent variations and adjustments as we will see next.

Given a system S, a model of the system, $m(S)$, can be expressed as follows:

$$m(S) =< C(S), E(S), S(S) > \tag{4.2}$$

, where $C(S)$ is the set of components of system S, $E(S)$ is the set of components of the environment relevant for S, that is to say, items that act on or upon items of $C(S)$, and $S(s)$ is the structure of system S, which is the collection of relationships between the parts of $C(S)$ and between $C(s)$ and $E(s)$.

Moreover, $S(S) = S_I(S)US_E(S)$, where $S_I(S)$ or endomorph structure, represents the internal structure of system S, and $S_E(S)$ or exomorph structure which is the structure done by the relationships between internal items in $C(S)$ and their relevant neighbors in $E(S)$.

In 2003, Bunge[164] included the mechanism $M(S)$ to the model, resulting:

$$m(S) =< C(S), E(S), S(S), M(S) > \tag{4.3}$$

However, this does not resolve one of the main objections of this model: the distinction between the system and the environment is too clear cut.

It is well known that one of the main features of a complex system are the indeterminate nature of its boundaries. The decision about where the system finishes and where the environment begins, is ultimately made by the observer which is not included in the model.

Nils Baas has developed a model that deals with emergent properties in complex system including the observer[181]. The definition of emergence rests on the existence of an observer with its own observational mechanisms. Thus, one property is emergent when it cannot be observed in one system S_1 but it may be positively observed in the resulting system S_2, which is the new structure generated through the interactions of the components inside S_1.

Then,

$$S_2 = R(S_1, Obs^1, Int^1) \tag{4.4}$$

where S_1 is the initial system, Obs^1 are the observational mechanisms and Int^1 are the interactions between neighbor components of S_1.

Therefore p is an emergent property iff can be observed in S_2 and not in S_1.

On the other hand, the inclusion of the observer in the model is problematic, because it may lead to infertile disputes(at least in the purely mathematical and scientific arena) , for example the realism/anti-realism in analytic philosophy.

The realists subscribe the existence of a world of objects independent of anyone's linguistic practices or conceptual schemes. On the other hand, anti-realism, states the predominance of the language over the thought and of the thought over the reality [182].

According to anti-realism, it is useless to try to answer whether one sentence is true or false and this is because in actual fact, the realist and the anti-realist support rival systems of logic.

Dummett, an anti-realist, suggests that a definitive resolution of such metaphysical debates can be obtained only if we give predominance to the language. Consequently, the confrontation between realists and anti-realists vanishes, because each opponent would be concerned with a particular type of language - so one might be an anti-realist about arithmetic but a realist about the past [183].

To sum up, the inclusion of the observer, poses some metaphysical complications which it would be preferable to avoid.

At this point, before we proceed deeper in the analyses of emergence, it is worth remarking that there are no properties per se, but properties possessed by objects, the same goes for processes, the object undergoes the process; ergo no process nor property without its object [165].

4.4.2 A few notes on emergent properties

The principal source, at less at philosophical level, in the study of the phenomenon of emergence, is that of the confrontation among atomism and holism.

The common attitude toward atomism reflects a misleading conception: atomicity is not a property that an object possesses, rather is a model or category good at describing objects at some particular scale.

On the other hand, holism is usually referred by the motto “the whole is more than the parts”. Holism states that in a given system, there are properties of the system that cannot be determined or explained by the system’s components alone. In this view, the properties of a system as a whole that can not be found in any of its individual components are called emergent properties. [184]

For example, the GNP of a country, is an emergent property: it is possessed by the whole ensemble of citizens and not by any single individual.

In this account, the moral can be drawn easily: holism is better placed in the biological and social sciences because can deal with the phenomenon of emergence.

It is necessary to clarify two points here. One is that emergence is an universal phenomenon, present in a number of systems, therefore, it is simply a bad idea to study it with esoteric mereologies or as a mysterious phenomena that requires even more inexplicable assumptions.

The other point is that it is possible to analyse emergence with rigour and precision using mathematical tools. More generally, when definitions or explanatory theories of complex systems are required, purely verbal approaches are in most cases unconvincing.

For example, the system composed by two atoms of Hydrogen and one atom of Oxygen have properties that lack the same single atom of Oxygen. Contrarily to properties such as the melting point, weight or density which are present in either the molecule of water and in the Hydrogen and Oxygen, there are other properties, like for example, viscosity, that only acquire sense when are used in the molecule of water and not in its individual components.

The natural numbers have properties like closing under operations -e.g: the sum of two natural number is another different natural, that are missed in one single component taken separately, for example the number 67. To call to those properties emergent or just systemic properties is from my point of view, just a matter of choice.

Once the idea of emergence as a hidden or mysterious phenomenon has been rejected, there is no objection to define that concept formally. Given a system S, a subsystem s_i of S is emergent iff s_i possesses some property that is uniquely possessed by the subsystem s_i inside the system S.

Formally,

$$S = \bigcup_{i=1}^{n} s_i, \exists p | p \in s_i \land p \in s_j j \neq i \tag{4.5}$$

Since the properties are not floating in limbo, a property is always a property of some material thing, we say that the system s_i is an emergent system.

It has been shown above, a model $m(S)$ for any system S. It ought be remarked that this model, relies on the analysis of the system S to identify its components $C(S)$, the environment $E(S)$ and the $S(S)$. The underlying idea is called ontological reductionism. In plain English, everything that exists is made of a small number of basic substances that behave in regular ways.

One important point that we can glean form the above remarks is that in order to build a model $m(S)$ of a given system S, first we have to identify the objects that compound the system, $C(s)$, then the objects that are not the system but affect or are affected by the system $E(s)$, and at last the relationships between the objects configure the structure of the system $S(s)$.

Once all its components and their relationships have been dissected, the study of the properties of the system rests just a matter of observation and measuring. But when we turn back to the subject of complex systems, that is to say, systems with emergent or systemic properties, we quickly encounter problems.

For example, if one builds the model above of a system like for example a tissue in a biological organism, it arises problems as well as questions that, I am afraid, our model has no answer to say.

First, the model requires a complete enumeration for both all the cells belonging to the tissue and all the cells that do not belong to the tissue, which act upon cells of the tissue.

Second, for every cellular component one needs to account for all its neighbors and the relationships between them.

Third, it is also required to describe all processes going on or upon all this plethora of cells. In addition to this, the tissue, of course, is a not a static object but inherently dynamic; certainly it is continuously suffering transformations. To be specific, cells undergo processes like dying, growing, moving or duplicating, so the components of the tissue may have changed in two different instants of time.

Clearly, system modelling is a primary tool for the study of the structure of a system. However, this is conditional on the requirement that the system is disclosed in its components and in the relationships between them. The structure can not be comprehended as a predefined set of pairs between components, because those components and the relationships between them, are not fixed but changing with the time.

I want to oppose this view with that of structure as the relevant organisation of a system that is invariant under transformations (homomorphism). This argument will be introduced in the next chapter 5. The consequence of the idea explained above is that we must shift the focus from the objects of a system, to the relationships between the objects that remain invariant under transformation.

This paves the way to study emergence focusing on the structure of the things, and not in their properties. The strategy drawn next falls under the rubric of epistemological reducibility.

4.5 Three Principles for Cognitive Systems modelling

In our exploration of the phenomenon of emergence, I establish three principles that may be taken into account for modeling in complex and cognitive systems.

4.5.1 Principle of Locality
4.5.2 Principle of Hierarchy
4.5.3 Principle of Multiplicity

This section attempts to relate emergence and other properties of cognitive systems like degeneracy, to three basic principles, namely locality, hierarchy and multiplicity. Establish this connection is an attempt to gain a better understanding in the design principles in cognitive systems, that may be also of help for scholars that build artificial systems equipped with cognitive abilities.

4.5.1 Principle of Locality

The Frame Problem refers to the difficulties found when trying to obtain a formal representation for a changing world. The world is inherently dynamic and in continuous change, but this does not mean that every action modifies all the possible states of the world. Indeed, there are persistent states as well as, morphisms between states. The interesting point here is to settle which morphism are homomorphisms or invariant under transformation.

It is thus required that we figure out which states of the system are persistent and which are not, that is, a set of frame axioms.

For example, if when driving my car and I change the gear, as a consequence of this action, the torque in the car's engine will be drastically modified, but of course, the color or the number of doors of the car will not(because they are a persistent state). Thus, the event of changing the gear is structure preserving.

Common sense theories have flourished in Research centers and Universities in the last 15 years, unsurprisingly the most active programs are around laboratories like the MIT, leaded by Marvin Minsky, the main referenced scholar in the vast literature of the frame problem [3]. Common sense theories are claimed as necessary because they could help in solving the frame problem [185].

Contrarily to this, the thesis defended here is that the non utilisation of the principle of locality, reveals with clarity, the weakness of the habitual tactics based on logics or in common sense theories.

Here, no solution for the frame problem, neither the development of a new logic to handle it is suggested, instead I will highlight the cause of such problem, which is locality. Without assuming locality principle, complexity is unmanageable.

The theories described in the books of physics are local in space and time, relating close spatial points and time instants, the distances between points and instants are ideally infinitesimal.

As I am proposing a general principle that must serve as a foundation for the modeling of cognitive systems, locality can not be understood exclusively in terms of physical distances.

Locality conveys direct links, also called arrows or morphisms. One link fixes a relation between two things, the links can be simple or composed, the former defines a direct relation and the composed link is a composition of two or more direct links.

For example, in the below graph, the arrows l_1 and l_2 define local relations between two nodes.

---

[3] There is some controversy about the paternity of the Frame Theory which is disputed between Minsky, Jerry Feldman and others. However the origins of the frame problem is in Hume's Problem of Induction.

$$\begin{tikzcd}\bullet \arrow[r, "l_1"] & \bullet \arrow[r, "l_2"] & \bullet\end{tikzcd}$$

The composition $l_1 \circ l2$ is depicted in the upper arrow in the graph below, which represents a global property of the graph because it encompasses the two local relations l_1 and l_2.

$$\begin{tikzcd}\bullet \arrow[r, "l_1"] \arrow[rr, bend left] & \bullet \arrow[r, "l_2"] & \bullet\end{tikzcd}$$

As it is pointed out in [186], Evolution may be alternatively thought of and analyzed as a composition of local procedures. The same metaphor could have inspired William James when defined consciousness as a continuous stream of thought. In this vein, human consciousness could be explained as the composition of local thought processes, leading to global loops of processes of processes, that result in a higher-order stream of consciousness.

Dealing with local to global properties is one of the most critical problems of today's science. It is far from obvious what algebraic structure will be useful to obtain precise local-to-global results.

Category theory provides the notion of colimit which can be used to deal with local to global properties. It might be noted that colimits do not necessarily exist in all categories. A colimit in a category can be seen as the binding agent in for example a neural architecture, the colimit captures the emergence of strictly increasing complexity.

4.5.2 Principle of Hierarchy

The Principle of Hierarchy (PH) circumvents the realist/non-realist ontological disquisitions about the structure of the world seen above, in section 4.4 (the same goes for the subjective/objective dispute).

PH establishes that nature is organized in levels, and a level is a collection of things with specific properties in common. Levels are not empirically observable things existing in nature, but mathematical abstractions, or sets of objects that share some properties.

Following the principle of hierarchy, we can add the level l to the model shown in 4.4, resulting:

$$m(S)_l =< C(S)_l, E(S)_l, S(S)_l, M(S)_l > \tag{4.6}$$

, where the level l, expresses a set of objects that share some properties.

Emergence is a systemic property, that is to say, it is a global property that can not to be found in the particular components of the system. The reason to include emergent properties is twofold:

i the property does not exist in any of the parts of the system;
ii systems have different levels of description, therefore some properties or others are "visible" depending on which level we are observing the system.

- l_i: is the level of things at level i (they share a set of properties)
- l_{i+1}: is the level of things at level $i+1$ (they share a set of properties)

We say that the level l_{i+1} supersedes or is superior to the level l_i iff all the objects in l_{i+1} are composed of objects of l_i.

For example, the level of the molecules of Hydrogen, l_i, is composed of objects belonging to the level of atoms of Hydrogen, l_{i-1}; then $l_i > l_{i-1}$ or l_i supersedes l_{i-1}.

One example of a level that supersedes more than one level, is the level of the graphs, for example l_g, which are composed of elements from the level of nodes,l_{g-1} and form the level of edges,l_{g-2}; then $l_g > l_g - 1 > l_{g-2}$.

We can now try to define emergence formally: one object,Ω, composed of objects from levels, l_1 to L_n, is emergent when has at least one emergent property, and rho is an emergent property iff ρ is presented at the i level and not in any of the precedent levels, 1 to n-1.

$\neg(\Omega_{1..i-1,\rho}) \wedge \Omega_{1..i-1,\rho}$

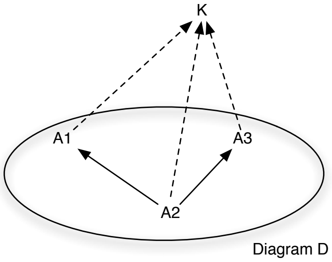

Fig. 4.2 A colimit K for the base diagram D. For the sake of clarity in the figure the diagram D has three objects $A_i i = 1..3$. The colimit shows that high level interactions .i.e: the colimit K, may arise without being directly induced from lower levels i.e: the diagram D

We can represent the configuration of a hierarchical system using the definition of hierarchical category as introduced by Ehresmann and Vanbremeersch in [12].

Definition 4.1. In a hierarchical category K, its objects can be partitioned into a finite sequence of levels, so that any object O in the level n is the colimit of at least one pattern P which is composed of objects that belong to level $n-1$ or lower.

Objects in a hierarchical category can receive and send information from objetcs at the same level, but also at lower and higher levels, that is to say, for any two objects O_1 and O_2 in the category K, there is no particular constrain in the domain and the codomain in the set of all morphisms $Mor(O_1, O_2)$.

4.5.3 Principle of Multiplicity

There is a commonplace statement that says that evolution leads to optimal systems. Such an assumption has been conveniently deflated and lay to rest in biological studies [15] and economic ones as well [187]. Accordingly, humans should not be delusional in thinking that we are optimal in our properties.

Evolution is not any guarantee for optimality and neither for progress. In a system that undergoes an evolutive process it is not true that there is always an increase of complexity, as a matter of fact, evolution can lead to simplicity if parsimony is of value.

However, complexity is a extremely relevant feature in today's science. In the words of Edward O. Wilson

> the greatest challenge today, not just in cell biology and ecology but in all of science, is the accurate and complete description of complex systems. Scientists have broken down many kinds of systems. They think they know most of the elements and forces. The next task is to reassemble them, at least in mathematical models that capture the key properties of the entire ensembles[188].

In [189], it is conjectured that a direct relationship exists between complexity and degeneracy, anyhow the authors do not provide any formal theory of such hypothetical dependency. Degeneracy is defined by Edelman and Gally in [15] as "the ability of elements that are structurally different to perform the same function or yield the same output". The capacity for the biological systems to degenerate is capital in the process of natural selection.

Degeneracy, if not an ubiquitous biological property, it is undoubtedly a property that exists at very different levels, from the genetic to the behavioral, passing by the cellular. And not only in biological systems if we follow the definition given by Edelman and Gally.

It is perfectly possible and indeed habitual, to find in nature different objects with differences in their composition that carry out the same function; in other words, objects with disparate composition have identical behavioral or functional manifestations.

It might be interesting to note that if we use the above definition, the manipulator arm shown in figure 4.3 is degenerate, because for two different configurations of the robot -the elbow above and the elbow below-, the end-effector reaches the same location A, as is depicted in the figure.

Ehresmann and Vanbremeersch in [12] introduce the multiplicity principle, which aims to set a mathematical frame to cope with the problem of degeneracy.

We can find a historical parallelism with the formulation of the principle of inertia. The principle of inertia does not impose itself a priori, but it is

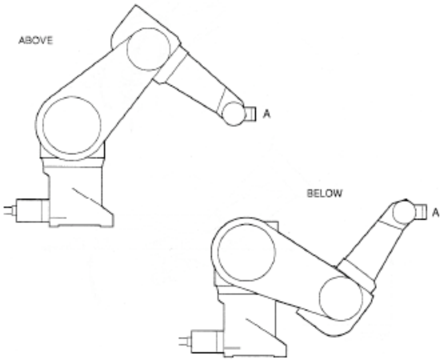

Fig. 4.3 The manipulator arm shown in the figure is an example of degenerate system if we follow the definition given by Edelman and Tononi in [15]. In robotics we say that the configuration of the robot is redundant

verified by its consequences, which is that the motion of all material objects on the universe depend on second order differential equations.

The multiplicity principle can be verified by the empirical observation: biological species show adaptability and resilience . Concepts like adaptability or resilience, in order to be usable, it is necessary to establish a way to measure them. The property of degeneracy is here that measure.

The mathematical theory of categories can help to understand degeneracy in systems. Ehresmann and Vanbremeersch define multiplicity principle as follows:

Definition 4.2. A category K satisfies the multiplicity principle if it admits at least two patterns which are homologous but not connected.

For the formal definition of homologous pattern and related concepts, the reader can consult chapter 4 of Ehresmann and Vanbremeersch book [12]. Here I will develop their ideas in a more informal way.

In a category K, for two decompositions P and P' of the same object C, such that C is the colimit of P and P', then P and P' are homologous pattern. It is important to note that this can occur without any direct link going from P to P' and vice versa, this shows that the colimit is a global property that can not be seen at the level of the components, as it was already noted above 4.5.1.

On the other hand, if C admits two decompositions that are not connected, C is a multifold object, and the passage from P to P' and vice versa is a complex switch. We are now equipped to explore the following statement: "The same function can be accomplish by two different structurally different components". This is a critical property in biological systems that can be redefined as follows: the patterns that these different components form, have the same colimit.

In brain studies the problem of selection is related to the above property. The problem of selection can be stated as follows: *1* how can we determine which areas are relevant for a cognitive process? and *2* which areas are activated incidentally when a cognitive process is engaged?

Brain image studies, detect activation of specific regions in the brain, but we must be prudent with the inference we take from those activations, they may be, indeed, incidental to the cognitive task used in the experiment. To claim that an area is involved in a task is very different that the area is specific to that task[190].

There is an important issue that was not covered above, indeed, how the patterns arise and can be formulated, has not been tackled here, but will be fully explored later in this thesis.

Chapter 5
The shift towards structure

Mathematics provides provable knowledge about the real world, and this is due to the fact that mathematical structures deal better than anything else with the structure of the world. That is to say, there is a structure preserving mapping between the mathematical structure that models the world and the world itself. The main idea is to translate the concept of algebraic structure into brain studies.

In this chapter we will fully explore the concept of structure, furnishing a formal definition of structure, and a categorical framework for representation.

5.1 Introduction

Scientists build models and theories that capture, in a more or less accurate way, the regularity of the world. This is possible because nature is structured. The Oxford Dictionary of English defines structure as *"the arrangement of and relations between the parts or elements of something complex"*.

Structure is a word used in many different contexts and situations, for example in architecture a structure is a man-made object fixed to a surface e.g. a building. In social sciences, structure is conceived in a rather abstract way as a human organisation with some common characteristic or goal, and in molecular biology, the structure of a cell is its shape or geometric appearance. In natural sciences and in social sciences as well, structure is a core concept, as a matter of fact, theories, paradigms and whole research agendas pivot on the concept of structure. There is linguistic structuralism, functional structuralism, structuralism in mathematics and even a structuralist film theory.

There is, *bien sûre*, a counter-structuralist program. The post-structuralism wave, criticizes the social imperialism of structuralism. The post-structuralism summons the structuralist project, based on the Greek ideal of rationality and in the systemic method of Descartes, as *tyranny of logos* [191]. Deconstruction is the post-structuralist tool to resist this tyranny, deconstruction would show us that there is no structure to be understood in any text. In this vein, it is key to understand Derrida´s theory of repetition. The sign is at the beginning of the text, but one does not understand its meaning independently of aspects that are not necessarily placed in the reader's expected "order". Along the repetition of the sign appearances, the sign itself is irremediably changed, therefore there is no beginning.

The post-modernism approach is not exclusive to humanities, for example *ethnomathematics*[192] is one possible case of post-modernism in the "hard sciences". However, in contrast to the social sciences, in the natural sciences and mathematics, post-modernism is rather peripheral and too localised to certain research centers and scholars.

Mathematics is relevant to science because it is able to uncover the structure of the phenomenon studied by the social or natural scientist, capturing the patterns and regularities that are hold by the system under study. Theories of mathematics, such as set theory or category theory, have the resources for describing a wealth of interrelationships between structures. Mathematical structures can be modeled themselves as a set of objects with certain distinguished relations and operations the set.

Using Elaine Landry's words "the concept of a model and so the concept of shared structure between models must be formally framed within a single unified framework, set-theoretic or other [193]".

5.2 Defining structure

In this section I flesh out the concept of structure and I explain the necessary link between structure and form. It should be emphasized here that structure is a not the kind of concept that can be easily defined. For the sake of clarity, I will begin by explaining what structure is not.

Structure, is not just a 3-dimensional physical construction, like for example a civil building, this is a layman's view of structure. By using mathematics, we can go deeper into this concept. For example, symmetry and continuity are structures that can be realised in both physical and non-physical systems. The symmetric structure in the plot narrative structure in structuralist film theory, is an example of structure in a non-physical system. Continuity is a kind of structure that can exist in time as well as space [194]. Later, in section 5.4.1 this idea of continuity as a kind of structure is exemplified using the Brouwer's theorem.

Modern physics studies symmetry by discovering the physical laws that remain unchanged when the system is viewed from different perspectives or undergone transformations[195]. In an informal and general way, **structure** can be defined as *the set of components of a system and their relationships.* Based on this definition, one system has a multitude of possible structures.

The structure is therefore contingent on the configuration of the system; to put it in a different way, the structure depends on which parts of the system are taken into account, and how the relationships between the different parts are calculated.

But we can not accept this definition as definitive. Indeed, any partition of a system could be its structure. It is therefore necessary to make a distinction between relevant and irrelevant structures. The relevant structures are those that are preserved under transformation. Consequently, the structure must reflect the invariance through transformations.

The idea is to go beyond the predominant view of structure as a geometrical shape (life sciences) or as a construction permanently fixed to Earth's surface (architecture), and translate the mathematical concept of structure into the arena of the natural sciences, in particular in the cognitive and brain science. In order to make this point clear, let us explore the habitual view of structure as a shape. The insufficiency of this approach is rooted in the fact that the shape of a system, when it has one, usually is a mere appearance; it lacks of explanatory power about how the shape came to be, or about how it will evolve in the future.

The physicists do not think in terms of the shape of the proton. Although the proton overall has a spherical shape, its exact shape dramatically depends on the momentum transferred to the proton, so if a photon couples to a quark with spin parallel to the spin of the proton, then the shape is not spherical but similar to a peanut[196]. The electrons are shapeless particles, we know it since they were discovered by J.J. Thomson. Obviously, the same goes for more complex systems like biologival organs, individuals, societies, countries etc.

The extraordinary diversity of the different organs, makes the view of organ as a purely morphological concept completely inadequate. The organs are, for the most part, determined by their forms (shapes) and also by their inner structure, which lies at the base of their outer shape. The concept of organ is a pure physiological concept, rather than a morphological one.

Only fractals, like coastlines or snowflakes, have parts, which are similar in shape to their whole. Perfect fractals, however, are very rare in nature. In conclusion, the form or external shape of systems can not be utilized as an universal feature for modeling of systems, complex or not. Unlike energy, for example, shape is not an universal property. Thus, strategies for system modeling that exploit the symmetry or similarity of shapes between the whole system and its parts can not be universal solutions.

Definition 5.1. Structure is the internal organisation of a system.

Proposition 5.1. *The operational aspect of structure is given by set of objects which are invariant under transformations (homomorphism).*

This definition of structure will lead naturally to the concept of structured-system and later, to the category-theoretically framed *cat-structure system* that will be developed later. Note that the term *cat-structure system* that will be used in chapter is equivalent to the notion of general category, that was formally defined in chapter 3 .i.e: a two-sorted system, the sorts being called objects and arrows that fulfill the axioms of associativity in composition and identity.

But prior to get into the sophisticated debates around the different interpretations of mathematical structure, I will try to be as didactic as possible and I will serve myself of the bellow allegory in order to hopefully transmit to the reader the intuition of the concept of structure, which will serve for the formal treatment of the concept that will follow.

5.2.1 The shepherd's tale

In a remote land a shepherd lived with his flock of sheep. He never went to school, so he never learned how to count or knew about the natural numbers, let alone arithmetics. Though uneducated, the shepherd was a curious man and wanted to know if he had more or less sheep than his neighbor in the valley.

At the end of the day when the sheep were back in the cave, he placed some pebbles that the shepherd collected during the day and matched each sheep with each pebble. Since then the shepherd could positively know if some sheep were missing or a new sheep had joined the group. The shepherd had arranged a mathematical structure. The sheep were now a structured system whose structure was exemplified by the amount of pebbles. This supposed a great achievement, the matching of pebbles-sheep represents the cardinality or number of members in the set of sheep. The first morphism or function between two structures had been attained.

The shepherd, as it was already said, was very curious, and after having figured out the structure of the sheep, he went further. Obviously the sheep were not all the same, our friend knew very well, who was the most lazy, the one with the darkest skin, the one that always got lost etc. As before, he put pebbles in the cave where he kept the sheep, but in this occasion each pebble matched each particular sheep, for example the biggest pebble matched onto the biggest sheep, the most grained pebble represented the least gregarious sheep and so on.

Now the shepherd could know, not only if some sheep were missing, but also which particular sheep, simply by looking at the unmatched pebble. Therefore, the structured system is now more complex than before, when it only captured the cardinality of the set of sheep. The structure exemplified by the pebbles with different aspects, exemplified different physical appearances, moods . . . of the sheep etc. Notice that this fable works as an anthropological explanation about how the mathematical structures started to be applied to physical systems. It is indeed a narration that tells us the way in which, a series of past events could happened in the past.

The same fable can have a phenomenological explanation: the cognitive agent, in our case the shepherd, describes in first person how his mind is working. Finally, a neurophysiologist will tell us what is going on in the brain of the shepherd, using tools that measure the brain activity.

From the three levels of explanation, the anthropological, the phenomenological and the neurophysiological, the last one is the only one that is positive in the sense that may empirically refute the theoretical hypothesis. The anthropological level is not falsifiable. Moreover the phenomenological level is exclusively verbal and subjective, rather than inter-subjective as modern science proceeds.

With the tale above, I hope have made my point clear: for any physical system, the relations between its components can be encoded into a mathematical structure. For example, the structure of a pack of six can of beers is 6 and the structure of the 12,349 telephone poles that connects Berkeley with San Bernardino (CA) is 12,349. Obviously 6 and 12,349 are not the only possible structures of those systems, but undeniably they are admissible abstract forms of those physical system.

That said, we can not dream that just by matching physical systems with numbers we can always get some relevant knowledge about the world. As a matter of fact, physical systems may present non empirically perceivable patterns and intricate structures that require more complex mathematical structures than the numbers.

5.3 Categorizing structured systems

As it has been showed though the above allegory, an abstraction e.g. a number, is a mental construction that makes it possible to talk about objects without point to them [197]. The abstractions function as a toy model of the world. In order to interact meaningfully with the world, we impose abstract forms on it. The abstractions are our interfaces through the world.

The cognitive agent establishes sets of equivalence, or classes between the systems (things), and the structured systems (abstractions), this is how it grasps the meaning of the world. It might be emphasized that structures can not be created from nothing, *Ex nihilo nihil fit*, there must exist a preexisting order. Michael J. Katz [198] coined the word *templet* to refer to the pre-order that makes the perception and the following abstraction of a physical system possible.

The *templets* are the necessary blueprints or natural patterns for perception. They collect the configurational information that is embodied in the physical systems. The ontogenesis of the natural patterns, or in other words, how the pattern of a system, P_S came about, are metaphysical questions.

Metaphysics, as much as other forms of constructing metaphorical worlds, can not help to represent the state of the natural environment for purposes of prediction and control; which is the aim of scientific models[199].

It is preferable to try to provide scientific explanations. This thesis sheds some light on questions like these: Given the pattern P_S, *How* can we obtain a model or representation of the system S, and *What* is a structure preserving form of S?

The inclusion of the formal representation of S in this theory, is graphically sketched in figure 5.1. Figure 5.1 resembles the schema of levels of reality proposed by Stanley Salthe {mass {form {organisation}}} in [180]. We must be prudent in the use of speculative or metaphysical disquisitions about what a pattern in the physical world is. We should ask which formalism can describe more accurately the patterns and systems existing in the physical world. In other words, how good is the mapping that relates physical systems with mathematical abstractions.

Of course, *Nihil novum sub sole*, the point I just exposed is indeed nothing but the core of the scientific explanation which consists of translating world phenomena into other terms that we know and understand independently of the phenomena being explained [162].

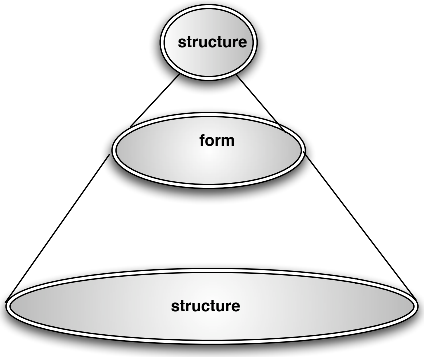

Fig. 5.1 *Structure* is a kind of *form*, while *form* is one possible kind of *mass*.

5.3.1 General systems theory

Bourbaki mathematical structuralism, assumes an *ante rem* conception of structure. The Bourbaki's interpretation of structure leads to a reification of structures as entities independently of the things themselves. Thus mathematical structures are constructed by taking an abstract set and appending to it an additional constructs [200].

We can see the systems theory (ST) approach as a kind of *in re* structuralism. In the systems theory discourse, structures are subordinated or even eliminated in favour of systems. Indeed structure is a shorthand for talking about systems of objects [201].

The models built by system theorists, rely on the measurability of the relevant properties of the system, namely the variables of the system. There are two main complications in the approach of the system theoretician.

Firstly, it is not stated what the real nature of the properties of the system are, neither which are the most relevant properties. Secondly, the definition of system in ST is too vague and general. From a philosophical perspective, there is an ontological positioning that must be seriously tackled: are the systems real objects existing in the physical world?; or are just conceptual tools useful to disentangle the components of which the things are made?

A rigorous analysis about whether the systems are *ante rem* or *in re* objects, is somehow missed in the ST literature. For example, in a purely philosophical standpoint, system means model, and objects of a system can be considered as positions in the model.

ST puts the focus on the components of the system and their properties, in other words, the system is broken down into its sub-objects, and so that its properties or attributes can be measured. Based on this idea, the general system theory states that a complex system is more than a sum of its components, because there are properties that are in the system but can not be found in its components.

Whether or not we want to admit it, the general systems theory (GST) has more of curiosity in the history of science books, than in a scientific program that progresses and makes quantifiable contributions. Sadly, GST never had a major impact in either the scientific or the philosophical context. This is not the best place to argue about the obliviousness of professional philosophers and academic scientists with regard to the GST program. For a deeper review in this topic [202]. However, it ought to be remarked, that despite its flows, the GST systemic approach and its emergent reductionism, are tenets to which, this thesis subscribes.

5.3.2 Structured systems

Here, the term system is used equivalent to the term *thing*, that is, the systems are the matter of which the physical world is made. Thus, the things of the world are systems. As the world is nomologically ordered, the systems are undoubtedly structured. The term *structured system* is introduced here as an abstraction of a system, that is to say, models of the things that exist in the physical world. The structure of a system is the abstract form of a system.

Any system, when perceived by a cognitive agent, is "naturally" structured. If the system lacks structure, the problems disappears, not because we solve it, but because we erase the statement of the problem. A system without any structure at all can not be a system. There are underlying structures which are present in the physical world, these structures may be common to mathematical structures and therefore can be formulated in mathematical language. We must start by mentioning the two main interpretations of structure in mathematics:

- in re: mathematics is about systems with structure or *structured systems*, rather than about structure itself. The shepherd's allegory is indeed in re interpretation, the numbers are a shortcut to talk about systems (sheeps) that exemplify the structure. Thus there is no structure as such but structure needs to be exemplified in a system.
- ante re: this interpretation assumes that structures exist over and above systems that exemplify them. Thus, the ante rem interpretation reifies structures because they exist independently of any given system or object, structures are prior in place, in this view, structure is reminiscent to Platonic *forms*.

Category theory provides, a *in re* interpretation of mathematical structuralism, it can be seen as a framework or tool that allows to organise what we say about structure. It is interesting to compare this categoric approach with the set-theoretic interpretation of structure, which has been the predominant view, at least in the period in which Bourbaki school was the most strong voice in mathematics and very specially in the didactic of mathematics. For Bourbaki, structure are types of set-structured systems.

Category is not just a new fancy mathematical abstraction that means the same as set or class, it provides a frame for the study of several structured systems related in some way. Contrary to set theory, in categories, the objects do not need to be elements and the arrows do not need to be functions. Categories are more general than sets and than any other mathematical structure, and using Elaine Landry's words "is able to organise our talk about both structures and structures of structures" [203].

Here, I quickly sketch why a category is a structured mathematical system, and why a set is not. Let two sets A and B in set theory, we say that A and B are equal, $A = B$, if and only if they consist of the same elements. Thus, $A = B$ stands for $x \in A$ if and only if $x \in B$. The sets A and B, $A = B$, because they have the same elements, *Anne, Tom* and their dog *Fido*. Now, imagine that Anne has developed an allergy to dogs, the sets are still the same, they still have the same components, but the relations between the components is not taken into account.

On the other hand, the category K of sets A and B, includes the mapping between the components. Contrarily to sets which have no structure, category theory makes explicit the relationships among their components. The evolution from set-structured system to category-structured systems is explained bellow in three steps. A *cat-structured system* , is a two sorted system, it has objects and morphisms as defined in the category-theoretic axioms showed in section 3.

system	the things of the world
structured system	the abstract form of a system
cat-structured system	a family of structured systems

First step, it is possible to define categories using set-structured sets and morphisms between them. For example, the category of topological spaces **Top** can be defined as a category such that its objects are topological spaces and its morphisms are continuous functions, or the category of vector spaces **Vec**, with vector spaces as objects and linear transformations as morphisms. Once categories have been defined, it naturally arises the question whether there is possible a category of categories. It is indeed possible to build a category where its objects are categories and its arrows are structure-preserving maps between categories, also called functors.

Second step, now we can start directly, not from set-structured systems but from the categoric concepts that were built in the previous step like category and functor. category-structure systems. This is an important step because it is here where the concept of category-structure system arises. We do not need to start from set-structured systems in order to build categories, the structure is unveiled in the patterns of morphisms defined in the category.

The nature of objects and morphisms in the category-structure system are practically irrelevant, what really counts is the patterns of morphisms in the category.

The last step establishes the context-principle, by which we can not ask for the meaning of any mathematical concept in isolation but in the context of a category.

The axioms of category theory: morphisms compose associativitely, with identity, provides the means to talk about structured-systems without being bothered with what the objects of the systems are made of, but focusing on how morphisms relate to each other. The relationships between the morphisms in a category or between the functors in the category of categories provides the patterns that allow to analyse the shared structured between systems.

The axioms for a category provide the context from within which we can analyze the shared structure of abstract kinds of structured systems in terms of the morphisms that exist between them [204]. The important point to be retained here is that it is possible using this perspective to talk about shared structure of structured systems in terms of types of cat-structured systems. A category acts as a schema that is used to frame the statements we can make about the shared structure of systems, in terms of types of cat-structured systems [205].

The sections above facilitate a review of the concept of structure and paves the way for its mathematical formulation. In what follows, we will try to apply it to solve real-world problems.

5.4 Linking the mathematical structure of the physical world

It is worth reminding our selves, that the central concept of this thesis is that of structure. The idea is to provide a framework that deals with objects as relational components in structured systems, by way of their shared structure, rather than as clear-cut systems that posses relevant properties to be measured.

5.4.1 An example of theoretical hypothesis in biological systems: The Brouwer theorem

In this heading, I exemplify how pure mathematical statements, may stimulate the research in areas which are apparently disconnected to the mathematical realm, like spatial representation in the brain, for example.

The Brouwer's fixed point theorem was originally created for topological spaces and continuous maps. I begin by giving a formal definition of the fixed point theorem, next an original use of this theorem is provided, showing that Brouwer's theorem may illuminate the place-cell/place-field mapping problem. The place cells are nerve cells in the brain hippocampus that encode the location.

Theorem 5.1. *Let I be a line segment $I = [a, b]$ and $f : I \rightarrow I$ a continuous endomap, then there is a point x_0 in I such that is a pixed point $f(x_0) = x_0$.*

This is the simplest version of Brouwer theorem because the continuous function f is defined in the one dimensional space I (a line). For a more general form of the theorem we need to take the mapping f as $f : K \rightarrow K$, where f is a continuous functions from a convex compact subset K of Euclidean space to itself. An example of K is the disk D in figure 5.2.

Thus, f maps topological spaces, that is, f is necessarily continuous, so a small change in the domain corresponds with a small change in the co-domain. Therefore no big leaps are permitted in the function f.

Let us see how to make use of the theorem in the hippocampal place cells. Two requirements are in order here. First, as f is an endomap, we need the same category at both sides of the function; and second, f needs to be continuous.

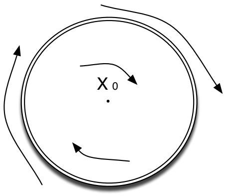

Fig. 5.2 Rotating the disk gives a continuous endomap $f : D \rightarrow D$. The center of the disk is the fixed point x_0

Let us imagine now the experimental setting. We want to explore how the place cells in the hippocampus of a real rat, represents the environment in which it has been placed. To make the experiment easier, the environment is one dimensional, that is, the rat can move in only one direction like for example a rail track. Suppose that f represents the mapping between the place cells and the physical positions in the rail track visited by the rat, as f is an endomorphism, it takes values in the same space in which it is defined, $f(x) \in I$ for all x, where x is the location of a place cell in the hippocampus.

Hence, if we find a setting in which the assumption above holds, the fixed point theorem applies, then there is a fixed point.

But in the rat's brain, such a continuous function f has not been found, as a matter of fact, close pyramidal neurons in the hippocampus do not necessarily represent close points in the environment [43].

Thus, as the hypothesis of Brouwer theorem does not hold, f is not continuous, we can not say anything about the possibility to have fixed points $f(x_0) = x_0$, which is the definition of place cells.

From the analysis of Brouwer's theorem, we can infer important consequences that are sketched in table below. Note that Brouwer does not directly apply to the experimental scenario, notwithstanding, through the logical analysis of the theorem it is possible to make progress in the research of fixed place cells in the hippocampus.

Brouwer states that

if H ($f : I \rightarrow I$ is a continuous endomap in interval I) *then* C (f must have a fixed point $f(x_0) = x_0$)

Logically, if C is true then we can not say anything about the validity of the hypothesis, H, which can be true or false.

Experimental research is focused on factual data, this is a very fine praxis as long as the theoretical aspiration to put the data together, in a sound theoretical body, is not neglected.

For example, given that we have discovered place cells, that is, C is true, we should try to incorporate this fact into a sound theoretical scheme, for example, the Brouwer theorem that allow us to obtain more general knowledge that can lead the experimental research.

On the other hand, if H is true then C is necessarily true, then provided that there is a continuous function, the fixed place cells exist.

If H is not true, then we can not endorse or reject the statement "fixed place cells exist".

These three cases are sketched in the table.

$H \Rightarrow C$	
if $\neg C$	no inference about H
if $\neg H$	C can be true or false
if H	C is true

So far, critics would say: to use Brouwer's theorem is an exotic attempt that does not reveal any deeper insight about how hippocampal place cells encode location representation.

This is true, especially if the research is driven by strict experimental analysis or computational simulation modeling. In the first case, the research's aim is to collect data in order to validate or reject hypothesis, and in the second case, the target is the reconstruction by computational simulation of biological phenomena.

Experiments with rats in a linear track show that different neurons are activated when the rat visits the same position, in this light the concept of place cell needs to be, in the best case, reformulated or even discarded.

On the other hand, computational reconstruction of place cells by neural networks able to learn physical position, can only hardly provide useful knowledge far from the particular requirements for which the simulation was planned.

The idea behind the introduction of the Brouwer theorem here is simple: rather than corroborate hypothesis or simulate phenomena, the goal of this approach is to generate new hypothesis. It might be remarked that "correct" hypothesis, that is to say, theoretical statements that are correct because they are embedded into a sound theoretical model, are not enough. The hypothesis must be empirically explored and verified. Only under these two constraints - theoretically sound and empirically verified- a hypothesis can be called "true" hypothesis.

5.4.1.1 Practical case: Measuring the activation of pyramidal neurons in the hippocampus

In this heading, I apply the formal theory of representation to cells in the hippocampus that are related to the representation of spatial features, like location, orientation or distance.

The idea is to build a theory for representation that is grounded in empirical data, rather than driven by biased experimentation, that leads to a posteriori formalisms that fit into the empirical data for which the formalism was designed.

It should be emphasized that the research in grid cells will likely be obstructed with similar problems that those found in place cells, namely the lack of a mathematical-theoretical framework for representation and the misconception between measurement and representation. This point will be conveniently explored in chapter 8.

Consider now that the observable ψ is measuring the activation of pyramidal neurons in the hippocampus. In this scenario what we want to know is whether the neurons are place cells, that is to say, neurons with a spiking/ firing rate related to the position of the animal.

In this case scenario, if ψ is not continuous, the position of two very close neurons can result in a big difference in their activations. Then we can not reduce the granularity by defining a finer topology that makes continuous ψ, because the basic unit of computation in the brain, and of course in the hippocampus, is the neuron.

At this point three important conclusions must be pointed out. First, the state space Θ composed of individual neurons used in experiments for detecting place cells, is unreliable in the mathematical topological framework. Unsurprisingly, ψ is not a continuous function.

Second, we need to make sure that Θ is a Hausdorff topological space. That is, given two different points N_1, N_2 of Θ we must guarantee that it is possible to find two neighborhoods θ_1 of N_1 and θ_2 of N_2 that do not intersect, $\theta_1 \cap \theta_2 = \emptyset$. If Θ were not Hausdorff, we would have two distinct states, N_1 and N_2, that are not distinct at all because we always can find another state as close as we want to be to both N_1 and N_2, which is at odds with the idea of distinct state. The problem with this can be epitomized with the following interrogation: How can we measure states that are not distinguishable from each other?

It may thus, be recommendable to have the Hausdorff condition explicitly stated in the formal definition of neural spaces, just as many spaces in mathematical analysis -e.g: topological manifolds, do.

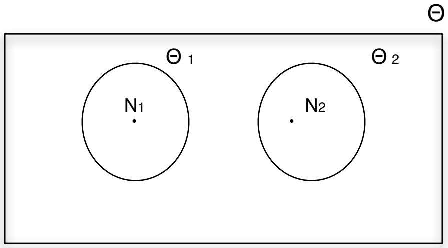

Fig. 5.3 Points N_1 and N_2 in a topological space Θ can be separated by neighbourhoods if there exists a neighbourhood Θ_1 of N_1 and a neighbourhood Θ_1 of N_2 such that $\theta_1 \cap \theta_2 = \emptyset$.

Third, the working hypothesis in the obtention of place cells is flawed in the following sense. It is assumed, that from the measurement of the firing rate of single neurons (the states of the system), it is possible to predicate the representational power of that neuron. Or, to put it in a different way, to be able to measure, for a physical position, the firing of one neuron compared to others, means that a possible representation of that physical place in the brain, is that precise neuron with larger activity than its neighbors, called place cell.

But, to measure and to represent are two completely different things. In fact, trying to discover what the neurons represent by measuring states in a topological space that could be discontinuous may lead to incongruent results.

Fourth, the mapping $f : \Theta \times t \to \Theta$ tells us in which state the system will evolve to, after some time t, $f(\theta, t)$. This assumes a complete knowledge of the system. In complex systems such as the hippocampus of a mammal, this assumption simply can not hold. We need to deal with “inconveniences” like noise and stochasticity.

To summarize, research focused on place cells, seems to have been put on hold as a result of the incongruent results, this could have been anticipated with the theoretical notions sketched above. It is now the grid cells that are attracting the attention and efforts of neuro physiologists. Without pretending to forecast anything, the research in grid cells will likely be obstructed with similar problems than those found in place cells, namely the lack of a mathematical-theoretical framework for representation and the misconception between measurement and representation. The notion sketched above of cat-structured systems or categories can provide new and powerful explanations to how hippocampal cells mediate in spatial representation, as it will be seen in chapter 8.

Chapter 6
A general framework for representation

In this chapter I present a general framework for representation based on category theory. The idea is to bring a mathematical formalism into the domain of neural representation of physical spaces, setting the basis for a theory of mental representation, able to relate empirical findings, uniting them into a sound theoretical corpus. The theory of representation introduced in this chapter, will provide a deeper insight into the process of representation. The major benefit of the application of this theory based on category theory is that, on the one hand, it may help to discard conjectures that are at odds with the formal framework. On the other hand, it will facilitate the integration of different models of representation in the hippocampus into a durable theoretical framework as I will show in chapters 8 and 9.

The chapter is structure as follows. The first two sections present two particular modes of representation: denotation and measurement. Section 6.1 investigates the idea of representation as denotation, and in doing so, provides insight into the concept of representative models. Section 6.2 describes mathematical representations, such as measurement theory and phase space. In this section, key terms like theoretical law, observables or measurement will be scrutinized. Section 6.3 provides a general theory of representation and sets the basis for the theory of representation based on category theory. This is fully described in Section 6.5, where a general theory of representation, based on category theory is introduced. Moreover, examples and applications of the theory are given in the domain of discourse of the category of metric spaces and the category of neurons.

6.1 Representation as denotation

We all know that there are symbols; for example icons, pictures or words that represent things, real or abstract. With the use of symbols we can talk about objects without having them in sight. The symbols do more than simply denote other objects. Let us see why. Representation is ordinarily assumed to be the same as denotation. For example, the red traffic light represents stop, or the word "MAN" for an English speaker, means a human male. But denotation is not the only way to represent. In fact, denotation is a weak version of representation, anything can denote anything, it is just a matter of convention, the red in the railway network in the United States can mean "clear" rather than stop and the word "MAN" can denote a brand of a truck.

I disagree with the philosopher Nelson Goodman who claims that "denotation is the core of representation and is independent of resemblance"[206]. Denotation is far from being the core of representation, indeed denotation is a non-epistemic representation. A model, m, is an epistemic representation of a system, T, when the user of th model can perform sound inferences from the model m that are applicable to the target T. This is also called *surrogative reasoning*. For example, while a map of the metro is an epistemic representation, the logo of the metro is not; the user can make surrogative inferences from the map to the real network and not with the logo.

In conclusion, denotation is colloquially taken as representation, but in truth, it is a weak form of representation because it lacks *surrogative reasoning*. In figure 6.1, the drawing is an example of epistemic representation of a man. A user, by observing the figure can infer properties of the real man represented by the figure such as the symmetry or the parity of the extremities in the human body.

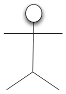

Fig. 6.1 We can infere properties -e.g: symmetry in the limbs, from this simple drawing that may represent a man

This has an important consequence: either denotation and epistemic representation define triadic relations and not dyadic, as is usually assumed. A vehicle e.g. a photograph or a map are representations of a target for a user or set of users. A symbol is not a representation in itself, but is always a model for some cognitve agent, in some context. The artist Joseph Kosuth expresses this idea of representation with his "One and Three Chairs" 6.2.

Figure 6.3 represents the process of perception, how the triadic relationship symbol-concept-referent is produced is sketched next. The referents of the external world are grasped by the agent. This is because the world is populated by material things which undergo processes which emit energy to be captured and encoded by the sensors 6.3(1). The sensory stimuli captured by the agent are objective and quantifiable. The properties of the perceived object can be measured, of course, the agent has perceptual limitations about what can and cannot perceived, based on its sensors and the way they are

chair (tshère), *n.*, chaise, *f.*; siège, *m.*; (of a professor) chaire, *f.*; (of the chairman or president of an assembly) fauteuil, *m.*; (rail) coussinet, *m.* Arm- —; *fauteuil.* Bath —; *vinaigrette, f.* Sedan- —; *chaise à porteurs, f.* Easy —; *bergère, f.* Rocking- —; *chaise berceuse, f.* Deck —; *chaise longue, f.*; *pliant, m.* To be in the —; *occuper le fauteuil.* —! —! *à l'ordre! à l'ordre!* To fill the —; *présider.* To leave the —; *lever la séance.* With . . . in the —; *sous la présidence de . . .*

Fig. 6.2 The artist Joseph Kosuth's famous work "One and Three Chairs", which a visual expression of Plato's forms. The piece features a physical chair, a photograph of that chair, and the text of a dictionary definition of the word "chair".

attuned. The patterns are instantiations of concept's properties for certain kinds of perceptions 6.3(2) that try to achieve the matching with the encoded information of the sensor's channels 6.3(3). When this computation succeeds, the referent is incorporated to the concept ontology.

In other words, the salient features or properties of the referent are identified and related to the agent's ontology of concepts. The conceptual component of a sign is depicted 6.3(4). In actual fact, it is an ontology of concepts which represent things or processes with common properties. According to this, the ontology of concepts is nomologically related by the relationship among the properties of the concepts. Due to the lawfulness of the concepts relations, learning is possible; if the brain lacked this, that is to say, the properties that belong to the concepts, the perceived item from the external world could never be classified. There it would be an agent with deficient cognition and scattered options to survive in a world ruled by laws.

Alternatively, if the agent, as is the case in humans, has a language or some other sign-denotative system of symbols, the relation between the external referent and the ontology of concepts can be by- passed by a symbol. The symbol 6.3(5) serves as a vehicle to share concepts within a community of agents [190]. However, there are others symbols that are not merely denotative, rather they permit us to infer, validate and even to create novel knowledge, we call such symbols, models. How do we know what is and what is not a model? The search for the essential features that make something a model seems a futile exercise; as a matter of fact, *we* make something a model by determining to use it as such [207].

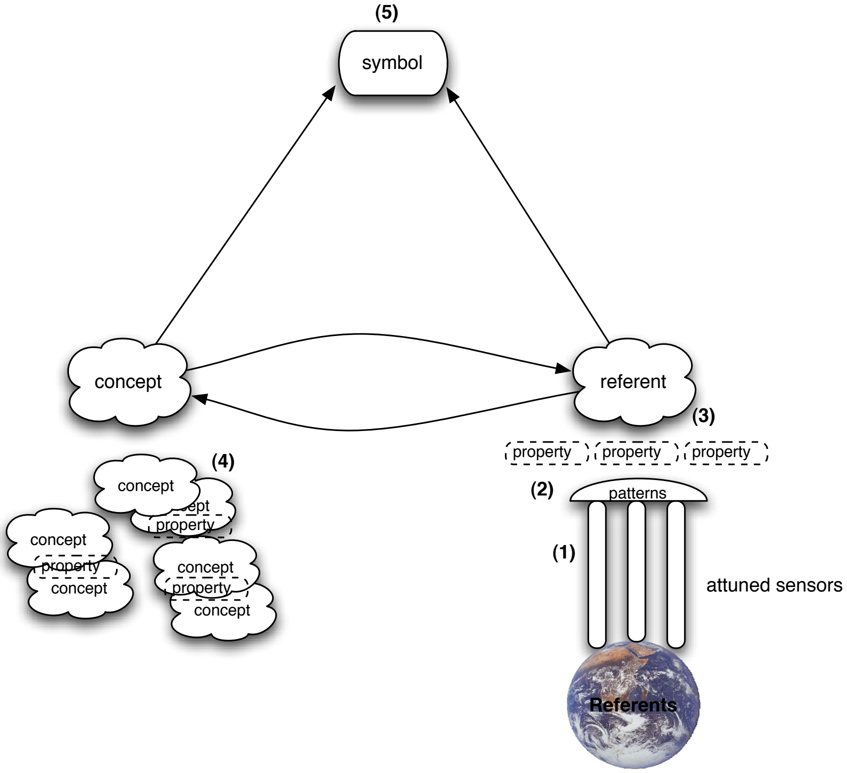

Fig. 6.3 The triadic representation conveys to us a new definition of representation as *the process of construction and relationship of signs within a system of signs.*

We may also consider, how good or "representative" a model is. The practice of science, can be equated to the practice of building models of parts of the world. Scientific models, provide statements which are empirically refutable, and able to predict and explain phenomena to a certain degree of accuracy. It goes without saying that the search of a perfect model of the world is essentially misguided because the perfect model of the world is the world itself.

The concept of model is very widely used, so a single definition that pleases everyone seems unrealistic. However, there is a general agreement in that *a model represents aspects of the world*.

In adopting this definition of a model, the problem now is to understand what representation is, that is to say, *in virtue of what a certain model represents a certain system?[208]*. The concept of representation is probably as elusive as that of a model. The reason is simple, one word can mean many different things. The term model is intended here as those more or less abstract descriptions of a system. We say that a model M is adequate for modelling a real system T, when T can be replaced by S for some purpose. For example, having a model of the Golden Gate bridge, permits us to calculate how many cars can be simultaneously crossing it, without causing any damage to the bridge. Maxwell's model of electromagnetics, Bohr's model of the atom or Watson and Crick's model of DNA are noted examples of scientific (representational) models.

But how adequate can models be? It is thought that a good model is one that predicts the future states of the real system. This argument, apparently sound, fails when we realize that Ptolemy's model of planetary motion was able to predict the path followed by the sun and the moon.

The adequacy of one model or other is linked to the purpose with which the model has been constructed. The model can be built only to show some similarity with the real system, or it can be built in order to make the model and the real system interchangeable, in this case we say that the model M is isomorphic to the system T or $M \simeq T$.

Carver [209] distinguishes between, phenomenally adequate models and explanatory models. The former are those that subsume the phenomenon under observation and the later are those that constitute an explanation because they afford the control and manipulation of the real system. The ability to control a system involves the understanding of the system, the reason is that we would be able to understand how the system behaves if the initial conditions of the system vary. This is what in [210] is referred as the modeller, which can answer more *w-questions* of the system, a *w-questions* is a "what-if-things-have-been-different" kind of question.

6.2 Mathematical representation

The main purpose of this section is to ground the concept of representation in a basis as unambiguous as possible. One of the major virtues of brain science is that it has been able to take theoretical laws with unobservable terms, and transform them into experimental ones, that is, composed of observables. For example, to ask what emotion is, what intelligence is, or what perception is, through an explanation exclusively based on experimental data and concepts, will lead us to ad-hoc explanations of those terms.

What is indeed possible and desirable, is to formulate empirical laws, based on observables in the brain, and formulate them using theoretical mathematical terms. It might be remarked that from experimental data alone, we can not directly infer theoretical laws, instead we need a set of rules that makes possible such a conversion.

Definition 6.1. An observable is a phenomenon that can be directly measured with the appropriate tool.

In physics, an observable is a quantitative magnitude that is measured on a system; an example of an observable is the torque of a motor, or the mass of a body. For example, in figure 6.4, the thermometer is the tool that assigns numbers to the observable which is the temperature of the body. It ought be remarked that for a philosopher, an observable is something perceived by our sensorial apparatus. Thus, external tools, a thermometer, for example, are excluded from this definition, which is *man-centered* and leads to puzzling problems like qualia, or the nature of consciousness that will not be addressed in this section.

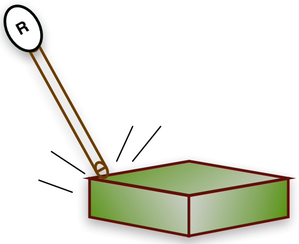

Fig. 6.4 The figure depicts a thermometer, the observable ψ, that when in contact with a system, the box, returns a real number **R** that is the state of the system.

The importance of observables [1] is that *empirical laws* are made out of them. The repetition of observables define generalities that can be expressed as empirical laws. For example, Galileo came to an understanding of uniform velocity and uniform acceleration by measuring the time it takes for bodies to move various distances [212], or the empirical Ohm's law, which is based on measurements of applied voltage and current through simple electrical circuits containing various lengths of wire.

$$I = \frac{V}{R}$$

On the other hand, *theoretical laws* do not arise from direct observation. Thus, *theoretical laws* can not always be created from the generalisation of observables, they are purely hypothetical. The utility of *theoretical laws* is to two-fold, first it allows us to unify in a common corpus, a set of empirical laws based on observables and second; it may suggest new *empirical laws*

---

[1] Observable and unobservable are not two fixed categories, indeed it is possible that as science progress an unobservable becomes an observable. For example, since Democritus atoms were taken as purely theoretical constructs; until 1905 when Einstein, in a series of three papers, explained the Brownian motion described by visible pollen grain, with the "invisible" molecules and atoms. Moreover, Einstein's paper also made predictions about the properties of atoms that could be tested [211].

(inexistent at the time in which the theoretical law was created) that need to be validated empirically. We need to turn to mathematics to find *theoretical laws.*

6.2.1 Measurement Theory

Measurement is a particular way of representation, a mathematical representation, that assigns mathematical objects, for example numbers, to empirical objects or qualitative entities. By the end of the XIX century, scientists and psychologists with an accurate knowledge of the physics and mathematics of their time, set the basis of the modern measurement theory. Measurement theory is the theory that assigns mathematical objects i.e. numbers, to empirical or qualitative entities e.g. length or mass but also for psychological concepts like pain or colour. This theory is part of Psychophysics which is the discipline that studies how physical properties are represented in the subject's mind.

First Helmoltz (1887) and later Hölder (1901), with more mathematical rigor, studied how to measure numerically "fundamental extensive" magnitudes such as mass or length [213]. Hölder demonstrated that there is a homomorphism between the represented *empirical* relational structure and the representing *mathematical* relational structure.

So, if $\mathcal{X} =< X, \lesssim, \oplus >$ is a continuous extensive structure then there is an effective procedure to find a homomorphism Φ of X into the structure $\mathcal{R} =< \mathbb{R}, \leqslant, + >$ of the real numbers $\mathbb{R}$ with the ordering relation $\leqslant$ and the addition operation $+$ such that for each x,y $\in X$, $x \lesssim y$, iff $\Phi(x) \leqslant \Phi(x)$ and $\Phi(x \oplus y) = \Phi(x) + \Phi(y)$

One main consequence that arises from this is that for any empirical system that satisfies the set of axioms described in [214], there is a homomorphism or function that preserves the quantitative relations among the objects of the empirical system, which are mirrored by the mathematical structure -e.g: the real numbers. In order to define homomorphism in measurement theory we need to first define relational structure.

Definition 6.2. $\mathcal{X} =< X, R_j >_{j \in J}$ is a relational structure iff X is a non empty set of empirical entities or numbers, J is a non empty set usually of integers called index set and R_j is a relation of finite order on X.

Thus, R_j is an element (a 0-ary relation on X) or is a subset of X (a 1-ary relation on X) os is a n-ary relation on X, $n \geqslant 2$.

By a homomorphism Φ of the relational structure $\mathcal{X} =< X, R_j >$ onto $\mathcal{Y} =< Y, S_j >$, we mean a function between the relational structures that satisfies i) ii) and iii)

i Φ is a function from X into Y
ii $\forall j, R_j$ is a n-ary relation iff S_j is a n-ary relation

iii $R_j(x_1, x_2, \ldots, x_n)$ iff $S_j((\Phi)(x_1, x_2, \ldots, x_n))$

If iii) is a one-to-one mapping from $R_j(x_1, x_2, \ldots, x_n)$ into $S_j((\Phi)(x_1, x_2, \ldots, x_n))$, then Φ is an isomorphism.

Example 6.1. Let $\mathcal{X} =< X, \lesssim, \oplus >$ and $\mathcal{R} =< \mathbb{R}, \leqslant, + >$. X is a set of physical bodies, ergo with mass, $R_1 = \lesssim$ is a relation that establishes that two bodies can weight less or the same and $R_2 = \oplus$ means physical addition of two bodies.

$\mathcal{R} =< \mathbb{R}^+, \leqslant, + >$ is a relational structure comprising the positive real numbers, the ordering relation and the addition relation defined by them.

The homomorphism $\Phi : \mathcal{X} =< X, \lesssim, \oplus > \rightarrow \mathcal{R} =< \mathbb{R}, \leqslant, + >$ assigns real numbers to bodies $\alpha, \beta \ldots \in X$ satisfying:

i $(\alpha \lesssim \beta) \Rightarrow \Phi(\alpha \lesssim \beta)$
ii $(\alpha \oplus \beta) \Rightarrow \Phi(\alpha + \beta)$

In words, *i* means that when the body α weighs less than the body β, the number associated with α e.g. 1.0, is smaller than that of β e.g. 2.2. And *ii* means that for any two bodies there is a number which represents the weight of the resulting weight of those two bodies put together.

However, Measurement Theory does not say anything at all about the neural substrate of representation or how physical properties are neurally encoded. We need more powerful ways of mathematical representation for this.

6.2.2 Phase space representation

The idea behind the mathematical representation of physical systems is extremely powerful. It lies on the assumption that a single mathematical object, is identical to the collection of measurements of the system, which are potentially infinite.

For example, a rotation matrix contains all possible rotations along one axis. For example, a rotation by an angle α about axis z is given by:

$$\mathbf{R}_z(\alpha) = \begin{pmatrix} \cos(\alpha) & -\sin(\alpha) & 0 \\ \sin(\alpha) & \cos(\alpha) & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

And just by composing them in good order we have the resulting state of the system that has undergone a series of rotations: a rotation about an angle α about z axis, followed by a rotation by an angle β about the original y axis.

$$R_y(\beta)R_z(\alpha) = \begin{pmatrix} \cos(\beta) & 0 & \sin(\beta) \\ 0 & 1 & 0 \\ -\sin(\beta) & 0 & \cos(\beta) \end{pmatrix} \begin{pmatrix} \cos(\alpha) & -\sin(\alpha) & 0 \\ \sin(\alpha) & \cos(\alpha) & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Let us look at another way of mathematical representation. Given a system S, the state of the system, Θ, is all what need to know about S, so Θ is a complete description of all the components of the system S, at any instant t. For example, the state space of a pendulum is given by position and the velocity $\Theta = (\theta, \dot{\theta})$.

If we know the equation of motion of S, then we can easily obtain the time-development of S in phase space, in this case a two-dimensional plot with position (x-axis) and velocity (y-axis) as shown in figure 6.5

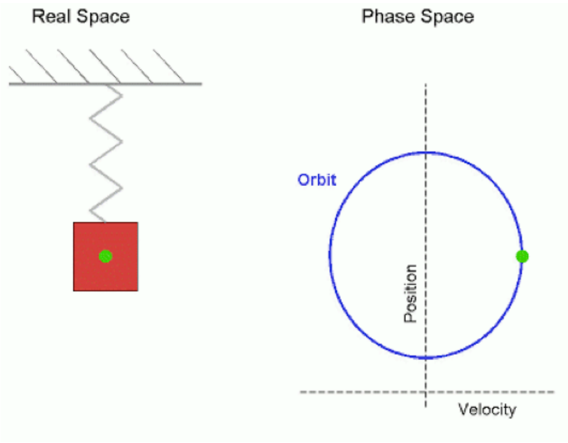

Fig. 6.5 "Real" space and Phase space of harmonic oscillator.

The representation of physical systems in phase space is extremely useful because it depicts how the system evolves over time, from one state to the other. The study of key properties of the system like stability, asymptotic stability or robustness rely on this kind of representation.

But how do we compute the different states? This is what measuring is about, to assign a real number to each state of the system. Thus, to measure it, is to build the mapping:

$$\psi : \Theta \to \mathbf{R} \tag{6.1}$$

The physical interest in using the mathematical abstraction Θ is that it is possible to exploit the concept of closeness of states. As we know, topology on a set is mainly about closeness of its subsets, therefore the set of states Θ is a topological space.

Now let us consider the mapping i.e. a measurement ψ, which can be seen as the operation to assign a number to a physical state.

If ψ is non continuous, the physical system will have a behavior which is not very realistic in the sense that small variations in the state that we are measuring, will result in huge changes in the value of the state space variable that is being measured. So we need ψ to be continuous, in order to be able to reflect the physical closeness of the states of the physical systems.

However, in the case of a discontinuous ψ -although it goes against the assumption of continuity as a necessary condition for measuring a physical system- it may still be possible, though not always, as we saw in 5.4.1, to find a continuous mapping ψ by reducing the granularity in the state space defined Θ.

6.3 A theory of representation in cognitive systems

Ronald Giere, a philosopher of science and former physicist, has worked on a unified and cognitive theory of science[215], [216]. This theory is interested in a particular type of representation, that is, established between the scientist and the (scientific) models built by him. Giere's interest in a theory of science of this kind is motivated by this fact; given that a theory of a phenomenon helps us to explain the phenomenon itself, a theory of science has to explain the phenomenon of science itself. Thus, the task of engaging in scientific inquiry is considered as just another cognitive activity, susceptible to be studied in the same way as any other mental activity.

Giere[217] succinctly defines his theory as:

$\boxed{\text{AMRWP}}$

that stands for: an **A**gent intends to use a **M**odel to **R**epresent part of the **W**orld for some **P**urpose. The expression contains at least two terms that need to be carefully considered. First, at the center of the formula, **R** or representation, which is the main component of the formula. It might be remarked that anything can be potentially a representation of anything else, but only those representations that embody relations that are preserved between the represented and the representing, are genuine *representational models.*

Definition 6.3. A *representational model* is a structure preserving description among two objects: the specification of the model(vehicle) and the represented object(target).

Second, the inclusion in the formula of the purpose, **P**, implies an intentional conception of representation, that is to say, the purpose of the agent in his task of modelling the world is taken into account. The following section fully investigates the nature of representational models.

6.4 Structural commonality in representation

There are two main arguments against a naturalistic and formal account of representation. One is that representation is a relation between two different domains that need to be interpreted by a user or observer; the second states that representation is context dependent, therefore, we can not get a formal definition of representation, because the phenomenological experience of the agent and the context, in which this agent is immersed, would be missing. Thus, in order to naturalise representation, we must find a kind of relation of representation free from the agent's purposes. A relation of representation between two domains implies an agreement in form. A structure is the abstract form of a system, it allows us to focus on the relationships among the objects, and ignore other features of the objects that do not affect how they relate each other [218], [219].

Cognitive agents do not have a total or omniscient access to the real structure or the world, rather they operate within it through the use of a continuous and dynamic endorsement between the form exhibited by the system and that idealized in the agent's model [207]. To represent something is intrinsically formal or structural, this view is opposed to the exclusively verbal idea of representation based on definitions. The thesis defended here, stands opposed to the widespread and long-standing approach that reduces representation as a linguistic or pictorial concern, instead it proposes a formal mathematical shift in representation.

If A and B share a structural commonality, then A represents B or B is represented by A. In order to say that A represents B there must exist a transfer of structure between the two different domains. But what is this structural commonality? Suppose we have answered this question (this is the subject of the next section), in that case, we have obtained a naturalised formulation of representation, that is to say, the purpose of the agent has been naturally dissolved in the structural account of representation. Thus, in the formula proposed by Giere shown above, the term **P**urpose disappears.

The task is as rewarding as it seems and therefore it deserves to be fully explored. Let us see how.

I identify three kind of relationships as possible candidates structural commonality:

1. Similarity: A represents B iff A is similar to B.
2. Isomorphism: A represents B iff there is an isomorphism between A and B, that is to say A and B are two isomorphic structures.
3. Homomorphism: A represents B iff there is a homomorphism between A and B.

The rest of the chapter is devoted to the study of the three kind of relationships: similarity in section 6.4.1, isomorphism in section 6.4.2 and homomorphism in section 6.4.3 At last, in 6.4.4 is analysed which relationship is required in a true relation of representation.

6.4.1 Representation as a relation of similarity between structured entities

Given A and B, we say that they are similar iff they share a subset of properties.

1. Similarity is a reflexive relation because A is similar to itself.
2. Similarity is a symmetric relation. Let A share the properties p_1 and p_2 with B. Then B share with A p_1 and p_2.
3. Similarity is a non transitive relation. Let A share the properties p_1 and p_2 with B. Let B share p_3 with C. The A and C could share any property rather than the property of sharing a property with B. For example, A is a set of 2 Bengal tigers, B is a set of 2 soldiers from Tanzania's army, and C sunlight reflecting in Victoria lake. A and B share the property *cardinality is 2*, B with C share the property *is located in Africa*, but A and C do not share any property.

The concept of similarity presented here is more general than that of similarity as visual resemblance. Indeed, two items can be "similar" in terms of the similarity relation shown, even if they are visually disparate.

6.4.2 Representation as a relation of isomorphism between structured entities

Given A and B, we say that they are isomorphic iff there is a one to one and onto map between all the components of A and B. Therefore, a map $f : A \to B$ is an isomorphism if exists the inverse of f, f^{-1}, such that $f \circ f^{-1} = 1_A$ and $f^{-1} \circ f = 1_B$, where 1_A and 1_B are the identity maps of A and B respectively.

An isomorphism describes a deeper level of "similarity" between objects. If there is isomorphism between two objects, there is also similarity, the contrary is not true.

6.4.3 Representation as a relation of homomorphism between structured entities

A mathematical representation of a non mathematical realm occurs when there is a homomorphism between the non mathematical system and the mathematical system. Indeed, homomorphism is less restrictive than isomorphism because homomorphic structures do not necessarily have the same number of elements. An isomorphism is a homomorphism that is one-one and onto.

Books of mathematics do not give a definition of homomorphism *tout court*, rather they define homomorphism of groups or homomorphism of Lie algebras [220]. Homomorphism, is a structure preserving map between two algebraic structures -e.g: homomorphism of groups, homomorphism of lattices, homomorphism of rings etc.

Homomorphism generalizes the notion of isomorphism.

Definition 6.4. Homomorphism is a structure preserving map between two structures. Formally, $\Phi : (A, *) \rightarrow (B, \#)$ is a homomorphism if it satisfies:

$$\Phi(a_1 * a_2) = \Phi(a_1) \# \Phi(a_2) \text{ for all } a_1, a_2 \in A$$

Example 6.2. A morphism Φ from commutative ring A into a commutative ring B, $\Phi : A \rightarrow B$, is a homomorphism iff it satisfies: for all $a, b \in A$:

$$\begin{aligned} \Phi(a + b) &= \Phi(a) + \Phi(b) \\ \Phi(ab) &= \Phi(a)\Phi(b) \end{aligned}$$

Thus, the homomorphism of commutative rings preserves the two operations defined in that structure: addition and multiplication.

6.4.4 Representation implies structural similarity: Homomorphism

Suárez in [208] answers no to this question, can scientific representation be naturalised?. The rationale is as follows: as representation is non reflexive, non symmetric and non transitive, then the similarity relation can not be representational because it is reflexive, symmetric but non transitive. The same goes for isomorphism, because contrarily to the representation relation, it is reflexive, symmetric and transitive. Therefore, representation relation in Suárez's view, is not constructed out of similarity and neither isomorphism, which leaves the project of naturalisation of representation in a distressing situation.

However, as I will justify bellow, this argument is defective, indeed the premise is false, therefore the rest does not follow.

Suárez argues that the relations between similarity and isomorphism do not constitute a relation of representation, and there could be something missed or a hidden relation between the two entities that participate in a representation. He gives a name to that missing piece: *representational force* or *essential directionality of representation.* Based on this statement, Suárez enlarges the definition of similarity and isomorphism incorporating his *representational force*, redefining similarity' and isomorphism' as follows:

1. similarity': A Represents B iff A is similar to B and the representational force of A points to B.
2. isomorphism': A Represents B iff A is isomorphic to B and the representational force of A points to B.

A major allegation is in order here. It is untenable to abandon the project of naturalising representation that would have major benefits, for example, the reduction of representation to empirical facts, or the formalisation of representation in mathematical terms. The view proposed by Suárez, as long as it uses obscure terms like *representational force*, can not be provide a solid basis for a theory of representation.

Misrepresentation by accidental similarity is one of the favorite arguments against naturalisation of representation. Suarez provides the next example. If my friend is dressed like pope Innocenzo X, I could think that Velázquez depicted my friend rather than the pope. Therefore, Suárez argues, this disqualifies similarity as a valid representation relation, because it leads to a wrong inference, the figure depicted is the pope and not my friend dressed like him.

The problem with this argument is that it limits similarity to visual resemblance. As it was already said in 6.4.1, similarity is not mere visual resemblance, rather it must be understood as structural similarity. Before we go further we need to clarify what I intend by structural similarity. One or more agents with analogous perceptual capabilities, can trivially state that two physical objects are similar by putting them together and observing if they share a number of relevant properties. Needless to say, to compare two similar objects is banal. Much more interesting is when an agent describes a physical object or phenomenon, using another object, for example a mental object. Here, for some, there is a problem because the mental object and the physical object do not share properties, the former exists in an abstract or mathematical realm and the last in the concrete or physical realm. This argument is flawed for one reason: because it is dualist, and it dissolves if we acknowledge two simple things; first, concrete objects have properties and second, the properties are parts of the abstraction of the object [207].

By this account, models do not apply to the world by a relation of layman's similarity, but by a more precise notion of similarity that I have coined **structural similarity**, and that will be mathematically defined bellow.

The scientific effort to describe nature is rooted in the obtention of a model or systems of models -i.e: a theory[2] that formally captures the structure presented in the phenomenon observed, this is possible because both the model and what is modeled are morphic structures.

In order to have two isomorphic structures, STR_1 and STR_2, it is necessary that *i.* the two structures have the same number of elements, and *ii.* the relationships between the elements of STR_1 have the same pattern as the relationships between the elements of STR_2

---

[2] The semantic view of models sees theory as a set of models

One example of two isomorphic structures is ten cans of beer and the set of integers 1..10. In this example, the invariant structure is cardinality, and distinctness is the relevant relationship capturing the invariance between the two systems that are being modeled. When we need structures that include richer relations between their elements, or at least richer than the very elementary distinctness, more complex patterns than that of cardinality arise. For example, Let A be a structure formed of three brothers, John, Paul and Tom and the relation "is older than"; and let B be the structure formed of an alligator, a python and a mouse and the relation "feeds on". Then A and B are isomorphic.

Given one cylinder and the same ten cans of beer, there is not an isomorphism because the number of elements differ. Nevertheless we can find a morphism -e.g: homomorphism, between the cylinder and the cans that preserves more relevant features than the cardinality, like the height or shape. Homomorphisms are structure preserving maps and therefore, true representations. Given a type of structure and given a map from a set A with this kind of structure, to set B with the same type of structure $f : A \to B$, f is a structure preserving map, if when we replace every element a in a structure in A we get the corresponding structure in B. For example, given a poset -i.e: a set with equipped with the structure $\leq$, we say that the map between two posets $f : P \to P'$, is a order preserving map, if $p_1 \leq p_2$ implies $f(p_1) \leq f(p_2)$. Note that for particular structures, homomorphisms may be known with specific names, for example, homomorphisms between posets are called order preserving maps or between vector spaces, linear maps.

Once we have defined structure and homomorphism we can build categories. For example, monoids and their homomorphisms define the category of monoids **Mon**, vector spaces and their homomorphisms define the category of vector spaces, **Vec**. This is what Arbib and Manes call the naive approach to categories [133].

6.5 Theory of representation based on category theory

In this section I provide a theory of representation based on the mathematical the theory of categories. Let us go now directly to the problem.

Definition 6.5. Given an object A in the category **C**, a representation of A in the category **C'**, consists of an object N' in **C'**, together with a morphism γ in the category **C** such that it associates each element a in A with a morphism from N' to N'.

$$A \xrightarrow{\gamma} Mor(N', N') \tag{6.2}$$

Thus, the object A in category **C**, is represented in the category **C'** as a bunch of morphisms, and the representation reflects the structure within the object A, which is given by those morphisms between elements in **C'**. This definition exemplified categorically the statement described previously, representation is a structure preserving map.

Let us see this with an example. A representation of a metric space, M in the category of neurons **Neur**, consists of an object N' in category **Neur**, together with a rule that assigns to each element n of N' in **Neur**, a homomorphism γ_a, from neural assembly N' to neural assembly N' such that the structure given by homomorphisms between metric spaces is preserved in the category of neurons.

Of course, a theory of representation may not regard representations as static or fixed items of knowledge. In order to represent objects of increasing complexity, the theory needs to be able to manipulate and create new representations from old ones.

At this point I will present examples of the genetic power of the theory presented. The capacity to create new representation cases is depicted through the propositions described bellow.

Proposition 6.1. *Given the representations of* A *on* N' *and* P', $A \xrightarrow{\alpha} Mor(N', N')$, $A \xrightarrow{\beta} Mor(P', P')$, *it exists a morphism from to the product of* N' *and* P' *to itself,* $N' \times P' \xrightarrow{pi_a} N' \times P'$, *that makes the following diagram commutes.*

$$\begin{tikzcd}
N' \arrow[d, "\alpha_a"] & N' \times P' \arrow[l, "n"'] \arrow[r, "p"] \arrow[d, "\pi_a"'] & P' \arrow[d, "\beta_a"'] \\
N' & N' \times P' \arrow[l, "n"'] \arrow[r, "p"] & P'
\end{tikzcd}$$

Thus, a new representation, that of A on the product $A \times B$ is created:

$$A \xrightarrow{\pi} Mor(P' \times N', P \times N') \tag{6.3}$$

Note that a representation of A on the coproduct $A+B$, $A \xrightarrow{\pi} Mor(P' + N', P' + N')$ is directly created by dualizing the above diagram, resulting the following diagram, in which the arrows have been inverted.

Example 6.3. Coming back to the example above we can extract the following heuristics: for an object , A, in the category of metric space, when it is represented on n different objects N'_i of $Neur$, it is possible to get the representation A as the product and as the coproduct of n cells: $A \rightarrow Mor(\prod N'_i)$ and $A \rightarrow Mor(\coprod N'_i)$ respectively.

$$\begin{tikzcd}
N' \arrow[r, "n"] & N' \times P' & P' \arrow[l, "p"'] \\
N' \arrow[r, "n"] \arrow[u, "\alpha_a"] & N' \times P' \arrow[u, "\pi_a"'] & P' \arrow[l, "p"] \arrow[u, "\beta_a"']
\end{tikzcd}$$

Proposition 6.2. *Let A be an object in category **C**, N' an object in category **C'**, and $A \xrightarrow{\gamma} Mor(N',N')$ a representation of A. Now, given a subobject K' of N', K' is a subobject of P' if $K' \xrightarrow{\omega} N'$ is a monomorphism.*

The theory should be capable to tells us whether A is still a representation of K'. Let us see how. By definition of representation, for every a in A, there is a morphism that takes elements of K' to K', $K' \xrightarrow{\gamma_a} K'$; the problem here is that we can not take for granted the morphism γ_a because it may be possible that for some element a in A, γ_a can take elements of K' out of K'. Thus, in order to predicate the representation of A on a subobject K', we need first to prove that K' is an invariant subobject.

Definition 6.6. A subobject K' is an invariant subobject of A, when for every a in N', there is a morphism μ_a that makes the following diagram commutes.

$$\begin{tikzcd}
K' \arrow[r, "\omega"] \arrow[d, "\mu_a"'] & P' \arrow[d, "\gamma_a"'] \\
K' \arrow[r, "\omega"] & P'
\end{tikzcd}$$

The invariance is an aspect of major importance that deserves closer attention. Topology is indeed, the mathematical discipline that studies the invariant qualities of structures under deformation and transformation. Lie algebras deal principally with the invariance in geometry and partial differential equations [221]. It might be noted that Lie algebras use the concept of cyclic permutations which is at the core of Groups. Helmholtz was the first in pointing out that cyclic permutations of a group may explain psychological questions like perceptual invariance.

Exploiting properties like symmetry, it is possible to map cognitive invariances into mathematical topological ones, resulting in a class of equivalent structures in the mathematical realm. Category theory arose in the minds of Eilenberg and Mac Lane as a general theory that deals with general forms of equivalence.

Remark 6.1. From a representation $A \xrightarrow{\gamma} Mor(N',N')$ and an invariant object K', it is possible to get a new representation of A on K':

$$A \xrightarrow{\mu} Mor(K', K').$$

Example 6.4. In the case of spatial modulated cells in the hippocampus - place cells and grid cells- the representation described in 6.2 is not relevant. If we assume that neurons are the basic computational units in the brain, we simply can not find a subobject K' of N', being N' a single neuron, -e.g: place cell, grid cell. But when the object N' in the category of neurons **Neur**, is an assembly of neurons rather than an unique cell -a singleton or category with only one object- the representation described above 6.2 will be relevant. This idea will be fully explored in chapter 8.

Proposition 6.3. *Given the representation* $A \xrightarrow{\gamma} Mor(N', N')$*, and the morphism,* $B \xrightarrow{\delta}$ *. By composition of morphisms it follows that*

$$B \xrightarrow{\gamma \circ \delta} Mor(N', N')$$

is a representation.

Therefore a new representation is generated just by defining a morphism from B *to* A *and a representation of* A*.*

First, let us explore the theoretical statement just shown with an example using the categories that are being used in this section, the category of metric space Met and the category of neurons $Neur$. In the category of metric spaces **Met**, given an object A, composed of elements forming an array of equilateral triangles, an object B, and a morphism $B \xrightarrow{\delta} A$, then by composition of morphisms B is represented as the bunch of morphism from the neural assembly N' to N'. The structure of B is reflected within $Mor(N', N')$.

Now, there is a mapping between two metric spaces, $f : (A, d) \to (B, e)$, if f preserves distances, for all $a1, a2 \in A$, $e(fa1fa2) = d(a1, a2)$. Thus, if we define the same distance, $d = e$, it is evident that for any subobject B of the metric space A, $B \in B$, $f : (A, d) \to (B, d)$, therefore any subject of the grid space defined in A would be a representation of the neural assembly N'. This is empirically false, of course not any subobject B of A can be represented as the same set of morphisms between a neural assemblies, $Mor(N', N')$ that represented A. Stated differently, a mental object and one of its subobjects, do not arise from identical neural assemblies in the brain.

Therefore, from the premises: $A \xrightarrow{\gamma} Mor(N', N')$ is a representation of a mental object A on neural assembly N' and $B \xrightarrow{\delta} A$ is a morphism from mental object B to A; it does not follow $B \xrightarrow{\gamma \circ \delta} Mor(N', N')$, B is a representation on neural assembly N'.

Having defined colimit in 3.

Proposition 6.4. *Given two representations of* A *and* B *on* N'*:* $A \xrightarrow{\gamma} Mor(N', N')$ *and* $B \xrightarrow{\delta} Mor(N', N')$*, the representation of the coproduct of* A *and* B*,*

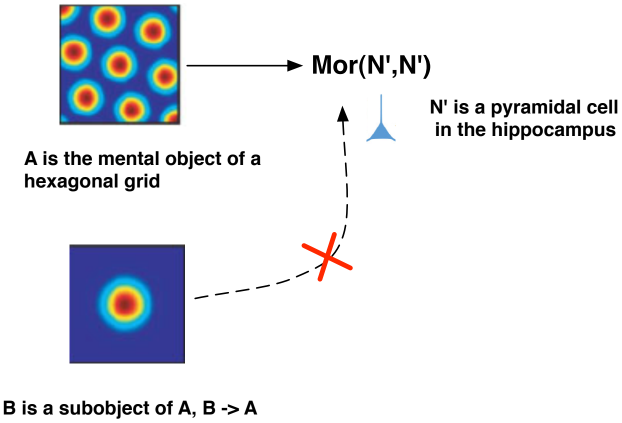

Fig. 6.6 Under the case explained above, to state that a mental object is the representation of a neural assembly is ill-founded, any subobject would be a representation of the same neural assembly.

$A + B \xrightarrow{\pi} Mor(N', N')$, *is immediately generated. (Note* π *is the unique morphism that makes the above diagram commutes).*

This proposition is an example of what I will examine in Chapter 8, which investigates place cell formation in brain hippocampus using concepts borrowed form category theory like colimit and coproduct.

Chapter 7
Towards a theory of brain structure, function and dynamics

While a lot is known about the properties and functions of neurons, the models for translating the data from the micro level to the macro level and vice versa, are still uncertain. This chapter has two main issues. The first concern is an attempt to point out the flawed use of mathematical and philosophical concepts, in our attempts to understand how the brain functions. More specifically, I draw attention to the restricted, yet inconclusive quarrel between the different schools of brain organisation and functioning. Modular, distributed or sparse conceptions of brain connectivity may be verifiable or not, depending on the measurement technique, and the scale that is being used.

The second concern is the introduction of mathematical descriptions of key concepts, such as pattern or structure, which are at the core of the natural sciences, and should occupy a preeminent position in the neural sciences. Category theory is presented as a valid foundational framework for modeling biological networks such as neural networks. Brain structure modeling is sketched under this *categorical* outlook. Concepts from Category Theory like colimit, functor or adjoint are proposed as universal tools that should be applied to the different levels of organisation of the brain. The scale-free dynamics of brain system and its subsystems need a powerful descriptive device, able to capture the creation and transformation of patterns of connectivity and activity in the brain at all the relevant levels. The categorical framework that is introduced here supersedes the K-sets theory of Walter Freeman and aims to formally describe the KV set that deals with the highest level functions in mammalian cognition, which have yet to be formally assessed [18].

7.1 Introduction

The eminent mathematician Henri Poincaré wrote *The aim of science is not things in themselves, as the dogmatists in their simplicity imagine, but the relations between things; outside those relations there is no reality knowable* [222]

The study of the relation between the external world and the central nervous system, and in particular the brain, has been approached from numerous disciplines and subdisciplines that are recognizable under the term *cognitive.* In the past, psychologists studied brain functioning, yet did not include an examination of its internal structure. The behaviorist school, that was preponderant in the departments of psychology during a good part of the last century, accorded to the biological organism, a predictable behavior in terms of input and output. The organism reacted with mechanical precision to the stimuli and what happened inside the organism was considered irrelevant because it was already prefixed by the stimulus/response pair.

Paradoxically, some time before Watson and Skinner gathered their behaviorist theories that disregarded the biological issues involved in behavior, the biological understanding of the brain had gained an extraordinary impulse. Ramón y Cajal using Golgi's stain technique provided definitive evidence that neurons are discrete cells. Cajal is the champion of the Neuron Doctrine which was established based on anatomical studies done by Cajal. Neuron doctrine states that the neuron is an anatomical unit, that through the laws of dynamic polarization, is unidirectional. The McCulloch and Pitts [63] notion of the neuron as a basic information processing unit in the brain can be seen as a newcomer in the Neuron doctrine [1]. The neuron doctrine states that neuron is the fundamental structural and functional element of the brain, both the basic building block and the elementary signaling unit of the brain [16].

At the neuronal level almost everything that needs to be known is already known. Although there still remains some technical limitations in the measurement of the brain activity, those technical barriers will efface, as the resolution in the brain imaging devices is progressively enhanced and new and more powerful techniques deployed. Today, the Neurobiology of Behavior is a reality that is continuously providing a deeper insight into the brain functioning and the neuronal correlation of behavior and cognition. Indeed, Neurobiology is vindicating Hippocrates' hypothesis that the proper study of the mind starts with the study of the brain. [223]

However, the organisation principles of the brain, able to to explain and predict the ontogenesis of meaningful patterns of connections between brain areas and their evolution along the time, is still missing. In the human brain there are approximately 10^{12} neurons and 10^{15} synapses, these are obviously

---

[1] Neuron Theory is an example of consilience where low level theories are absorbed into higher level theories that explain the base data as part of higher order structure

very big numbers. The daunting dimensionality of the problem is drastically reduced by the phrenologists and by some cognitive psychologists of modularist vein. They state that both the domain and the co-domain in the mapping from cognitive functions to brain organisation, $f : C \rightarrow B$, are both discrete. Although such theoretical positioning can be seen as justifiable for practical reasons, because it makes the problem manageable by reducing its space state, it is empirically false.

7.2 The crisis of a paradigm: Historical account

As early as the beginning of the nineteenth century, Francis Gall, developed what he thought was the mapping of the cognitive and emotional capacities into the skull and the brain areas underneath. Pierre Flourens, tested Gall's theory by trying to isolate different brain regions in order to establish the different roles played by each different brain region, as the phrenologists advocated. He reached to the conclusion that no specific region was exclusively responsible for any particular mental operation. Rather, the brain as a whole, participated in all mental operations. This holistic view of the brain as a decomposable organ, was later called the Aggregate-field view. Later on, the Aggregate-field view was refuted by J.H. Jackson who, through his studies in epilepsy, showed that it is possible to locate some motor and sensory functions, in particular parts on the cerebral cortex. However, the Aggregate-field view continued to be championed by psychologists like Ivan Pavlov and Karl Lashley.

By 1950, Lashley had reformulated the Aggregate-field view, based on two principles, the principle of *mass action* and the principle of *equipotentiality*. The former stated that the cerebral cortex acts as a whole in many types of learning, and the later established that if certain parts of the brain are damaged, other parts of the brain may take on the role of the damaged portion. Thus, in the new Aggregate-Field theory, there was not any particular singular locus to look for in the cortex that served cognitive functions. Moreover, the cytoarchitectonic approach that lead to the exclusion of a map of the functional parcellation of the cortex was considered nearly worthless [224].

However, facts are stubborn things and the empirical evidence for the anatomic localisation of the basic functions was irrefutable.

Karl Wernicke was the first proponent of the now prevailing notion of the brain as a distributed system processing, simple cognitive functions -e.g: simple sensory and motor activities, are localized in particular areas of the cortex, and complex functions can be traced as the result of the interactions between simple functional sites [225]. In the 1930s, experiments with cats showed that the application of a tactile stimulus in different parts of cat's body, triggered an electrical activity in particular regions of the cortex. The basis was used for the mapping of the cortex as in the depiction of anatomical regions described by Broadmann.

7.2.1 The old paradigm: Modularism-homuncularism

There is a conceptual and a methodological problem in finding out which brain area a is correlated with which function f_a. When one states that area a is responsible for the function f_a, what is really meant, is that the activity measured (electrical, magnetic ...) in that area a, has overcome the threshold that the experimenter is using for his experiment, while other areas b, c or h have not. As any first year student in logic knows; an event a occurs simultaneously with an event f_a, is not the same as f_a entails a, $f_a \rightarrow a$, and of course differing from a follows f_a, $a \rightarrow f_a$. However, for educational and historical reasons, in psychology, the cognitive functions have been classified in main cognitive faculties e.g. perception, memory or attention, and therefore, treated as discrete items. In actual fact, those functions are not islands, but interdependent and highly coupled. For example, perception is a function of memory and attention. If we admit that cognitive functions can overlap, their neural basis may overlap as well. It has been demonstrated that in the brain distinct functional regions exist. For example, the left hemisphere is primarily concerned with sensory and motor processes coming from the right side of the body, or the frontal lobe is related with tasks that involve planning, but that does not justify the functional parcellation of the cerebral cortex as posited by Gall. In short, and using Fuster's words, to *constrain a cognitive function within the scope of the method used to test it can not be a far reaching methodology* [226].

Historically, the attention has been put on the anatomically differentiable areas (neurobiologists approach) or in the sensorimotor or complex cognitive faculties (cognitive psychologists approach), both belong to what I call the old paradigm. From two divergent approaches (from the neuron up to the psyche and from the cognitive function down to the neuron); there is a resulting substantial gap between the domain of the anatomy and physiology of the brain, on the one hand, and the domain of behavior and cognitive abilities, on the other.

7.2.2 The brain is not an aggregate system

Wheeler, in [227] defines the concept of *aggregate system* as one that holds two conditions;

i It is possible to identify the components of the system by their explanatory function without taking into consideration the rest of the system.

ii Non-trivial cases of system-wide behavior can be explained by reference to the operation of comparatively few parts.

According to this definition, the brain, contrary to the practitioners of the functionally modular vision, is not an aggregate system. As claimed by Wimsatt [228], modularity is typified by aggregative systems, therefore the brain can not be modular in the way that Wheeler's definition of aggregative systems states. However, the choice to categorize an intelligent system -e.g: the brain, as fully aggregate is too simplistic here. The concept of Continuous Reciprocal Causation (CRC) posits that in "the presence of continuous, mutually modulatory influences linking brain, body and world." [229] we can't use componential analysis because we can't isolate the components.

CRC would make the system under study less and less aggregative over time and thus, the reciprocal causation between the components of the system would decrease its modularity and it would become more holistic. In a complex system, components are coupled and the dependencies and interactions between them are inherently tangled. The system's behavior, is not just a matter of an analysis of the components, but it is the interactions between the components that primarily must be discovered.

Despite the evolutionary process of continuous complexification of the structure and its progressive specialisation in the brain, the different subsystems e.g. cortical areas, limbic structures etc., never act independently as isolated systems. Indeed, almost any cortical neuron or neuronal assembly can be part of many networks and thus, instantiates many different cognitive functions. Accordingly, a modularist or homuncularism [230] analysis must be declined as a valid theoretical framework for the brain functioning. As a matter of fact, biological systems are typically not composed from previously autonomous systems incorporated into a common system. The biological systems are commonly broken down into several parts, which are subsequently further differentiated. This is probably how the different cortical areas evolved.

One possible exception to this general rule is the mitocondria, which is a membrane-enclosed organelle found in most eukaryotic cells, composed of sections that carry out clearly differentiated, specialized functions where the mitocondria resides. The mitochondria is a case of a big sudden change in the evolutive process.

7.2.3 The brain mapping is of form and meaning

Cellular connectionism states that individual neurons are the signaling units of the brain, arranged in functional groups, that are connected to one another in a precise way. Although this is the established view in modern brain physiology, it still poses one important issue; exactly what is a functional group and in what way they are connected to each other?

In basic theory of sets, a relation between two sets is modelled through a function f, that has a domain and a target as two related sets. Thus, a function $f : A \to B$, relates each element a in A to another element b in B. Likewise, when a neuroscientist investigates the neural substrate of the cognitive functions, he is indeed carrying out the function $f : C \to B$, that relates the domain C of cognitive functions and the co-domain B of the neuronal organization in the brain.

In order to build the function f that relates cognition with brain areas, $f : C \to B$, we are confronted with the problem of granularity that we can state with this question: what are the unit elements in the brain, B, to which the cognitive functions, C, are mapped onto?. We must note that if the neurons are those basic units, then the mapping becomes intractable and meaningless. Conceptually the problem is not straightforward. How may we get out of this conceptual conundrum?. If we assume that the space B is discrete, then the form associated with some cognitive faculty is vague. In other words, the anatomical and physiological borders of the area under study for example, the right hemisphere, or frontal lobe, are too imprecise and general. Nevertheless, if B is thought to be continuous, then there are no modules to which we may ascribe any particular function, but instead, a daunting circuitry of neurons and synapses. In this embarrassment of riches, the *form*, that is the anatomical parts where the cognitive abilities rest, can be precisely determined, but the *meaning* is missed.

The nature of brain mapping must encompass form and meaning, and the best way to capture such mapping is to focus on the links between form and meaning. Only those forms with representational power have meaning.

It is possible, in the appropriate laboratory conditions, to trace the course from the stimulus to the final action, bringing to light the neural pattern of activation triggered. I will explain this with two examples of increasing complexity.

The patterns are the basic units that we are looking for in the organisation of the brain. A pattern is a meaningful structure made up of an ensemble of neurons or a group of neurons, which are functionally connected. Pattern is a word with many, surely too many, possible uses and interpretations. Therefore, rather than provide yet another possible definition, I will outline through two examples what I intend by the word pattern.

7.2.3.1 The Sensorimotor Topographic Pattern

First, a simple case, the sensorimotor pattern. As it is well known, Cajal was the first who comprehensively traced the sequence of transformations from the sensory neurons of the skin to the motor neurons of the limb to be moved. The signal was initially perceived by the sensory neuron. The sensory neurons are attuned to the outside world, thus the visual sensory neurons respond to light, auditory to sound waves, and so on.

Features of the stimulus, for example the intensity of light or the frequency of sound, are captured by sensory neurons attuned with those events, in the form of receptor potential. When the receptor potential is bigger than the threshold for that cell, an action potential is transmitted along the axon. The greater the overpass of the threshold is, the greater is the frequency of the action potentials in the axon. Similarly, the longer the stimulus duration, greater is the number of action potentials that are transmitted. The neurotransmitters released by the sensory neurons interact with their counterparts in the motor neurons and initiate a graded synaptic potential. If the membrane potential in the motor neuron is greater than its threshold, an action potential will be created which will again, cause the release of neurotransmitters that produce the synaptic potential, able to create the final the action potential in the limb's muscle that results in the muscle contraction.

Intentionally, the above description is a simplification, for more detailed descriptions the reader may consult any good manual of neurophysiology. [223], [231] So apart from technical details, there is one important thing to be said. The transformations of the stimulus into neural signals that convey information are in fact, one-to-one maps, also called isomorphisms. The pattern is hardwired in the path between sensory and motor neurons.

As a result of the isomorphism, the neural circuitry or pattern that will be triggered by one concrete stimulus like a stretch reflex, is quite predictable, likewise, the same stimulus will produce the same neural pattern. It should be noted that the election of the above example, the signaling network between sensory and motor neurons, is not casual. The information conveyed by the sensory neuron acts directly on the motor neuron, producing the muscle contraction. Thus, an overt behavior can be registered in the animal. The study of the neural circuitry of sensory neurons connected directly to the motor neurons other than those in the skin, though for example visual or auditory neurons, is much more difficult.

However, there is a more serious obstacle here. We live in complex environment with myriad of different stimuli to be processed by the sensory neurons. The brain map at the simple sensorimotor interaction described above is topographic, but the further we move towards associational areas of the brain, such as the prefrontal cortex and parietal cortex, the less evidence is found for the existence of topographic maps. The perceptual brain is indeed topographic, but only up to V1 - beyond this, everything gets much more complex. This is only one of the many visual areas in the brain.

7.2.3.2 The Meaningful Patterns in the Olfactory bulb

Walter Freeman, a Berkeley neuroscientist, remarkably gifted for electrical engineering and mathematics, has investigated the general principles of brain organisation during the last forty years. Freeman's studies have focused on the olfactory system of cats and rabbits. Simplifying, again this is how it

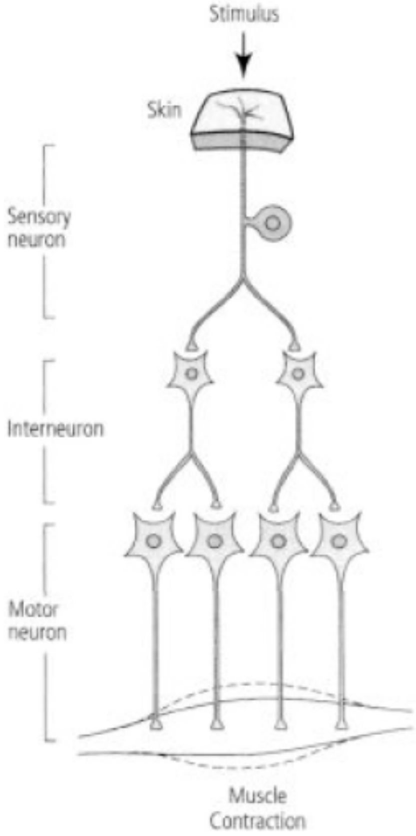

Fig. 7.1 Connection specificity -a neuron is connected with some particular neurons- and neural polarisation -within a neuron the signal is unidirectional- was discovered by Cajal. Drawing of [16]

happens. The stimulus; an odor, goes into the receptor layer, the olfactory sensory neurons in the nose. These neurons project axons to the olfactory bulb (OB) of the brain. There, it is transduced into a pattern of action potentials. Obviously, in order to be discriminated, the odorants must cause different patterns. The bulbar "code" for olfaction is spatial, this means that the odorant activity is represented in the OB by the activation of odorant receptors with different spatial locations. With each inhalation there are fundamentally two sorts of spatial pattern neural activities that coexist in the OB, the microscopic and the mesoscopic. The procedure of obtention of the two patterns is very different, the microscopic is evoked by action potentials in single cells, and the mesoscopic is by ensemble averaging.

The microscopic pattern of single neurons, as it occurs in the sensorimotor map, is topographic and 1 to 1; in other words, there is an isomorphism between the stimulus and the neuron. On the other hand, the mesoscopic pattern reflects how local masses of neurons organise their activity when they are destabilized by a sensory input (odorant). The bulbar patterns do not relate directly to the stimulus because each bulbar axon branches and distributes its output over the cortex.

Both the micro and the mesoscopic pattern express properties of the stimulus applied. But there is a main difference, the mesoscopic patterns capture the meaning of the stimulus. But, how can we test the meaning? To that end, Freeman elaborated the following experiment. A rat inhales an odorant O_1 and is rewarded. After the rat inhales a second odorant O_2 and no reward is given to the rat. The EEG patterns of O_1 and O_2 are recorded. Later on, the rat inhales again the same odorant but the stimulus is changed. Thus, the rat smells the odorant O_1 with no reward and then the odorant O_2 is presented followed by a reward. The EEG patterns of the odorants O_1 and O_2 are recorded. The pattern of the odorant O_1, when is presented prior to a reward and when is given with no reward at all are different, though the odorant is the same. Likewise for the odorant O_2.

This showed that, as expected, different odorants give rise to different patterns, but unexpectedly, given the same odorant again, the original pattern does not reappear, rather a new one is created.

The conclusion is that the bulbar patterns do not relate directly to the stimulus but to the meaning of the stimulus. The most important consequence is that patterns are not directly shaped by the stimuli, rather the formation of the patterns is endogenous. In other words, previous experience of the animal with those stimuli, together with its own goals and internal motivations, are triggered by the stimuli and as a result, a new pattern emerges. To sum up, an odorant stimulus and the following pattern in the olfactory bulb are not isomorphic, rather the pattern expresses the meaning of the stimulus for the subject.

7.2.3.3 Meaning versus Information

The technical use of the term information comes from Claude Shannon's 1948 classic "A Mathematical Theory of Communication". In Shannon's theory only the syntactic aspect of information is considered, ignoring the semantics or pragmatics. For some[232], the semantic omission in Shannon's Theory made possible its mathematical formulation. The direct application of Shannon's Information Theory in biology is ill-founded because semantics is an essential in biological systems [233], [234], [235].

In order to treat information as a reduction of uncertainty, Shannon's Theory needs two main assumptions i) the probability of occurrence, if each symbol is known a priori and ii) an external to make sense (semantically) of

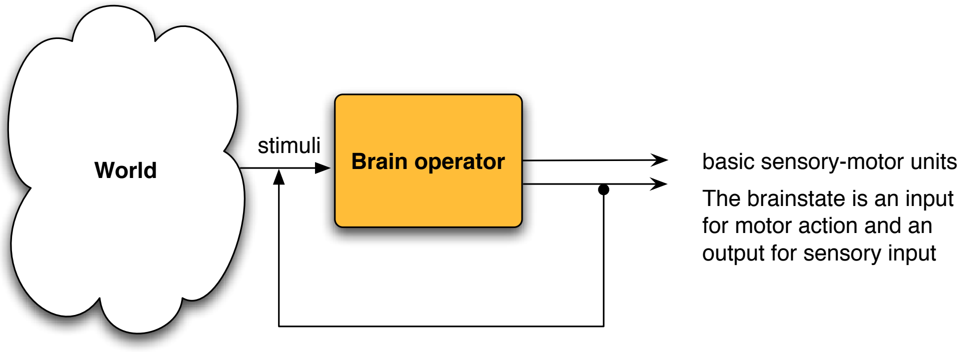

Fig. 7.2 Linear causality can tell us how a reflex response is caused by a conditioned stimulus -basic sensorimotoric loop. Thus, a stimulus excites a sensory neuron, that excites the sensory cortex, that transmits information to the motor neuron, that finally produces the contraction of the muscle. The mechanisms for perception are not strictly linear but circular. Sensory input entails motoric consequences and likely, motor output affects sensory input. In this light, the brain can be seen as an operator between the world and the body.

the symbolic exchange. Another conception of information useful in biological systems is Bateson's definition of information as "the difference that makes a difference". This recurrent definition, though apparently confusing because the *explanans* and the *explanandum* are circular, is an attempt to exploit the recurrent organisational structure in the interaction biological system-ecosystem.

It ought to be remarked that information is used here in the usual sense of the word, and not in the technical definitions utilised by Shannon or Bateson. According to this view, we can state that information, for example, the precise location and the types of the receptors that are triggered, is irrelevant, to use Freeman's words. It is the class of equivalence to which the activated receptors belong, the kind of *information* that is relevant here. In order to identify the equivalent class or category where the odorant inhaled belongs, the olfactory bulb must generalize over spatial patterns of action potentials that are transmitted to it.

This is contrary to the Information Processing Hypothesis (IPH) which states that for each distinctive odorant, there must exist a spatial pattern of bulbar activity. The IPH conjectures that the stimulus is represented directly in the pattern, as it was previously described in the stretch reflex.

The thesis I defend here, is empirically rooted in Freeman's studies in mesoscopic spatial patterns in the olfactory system, that envisages the brain as a processor of meaning rather than information. The odorant inhaled, occasions a spatial pattern, but it is not until the brain builds the representative class of the inhaled odour that the item of knowledge is created. The meaning is created by the interactions of intentional beings with each other in the world.

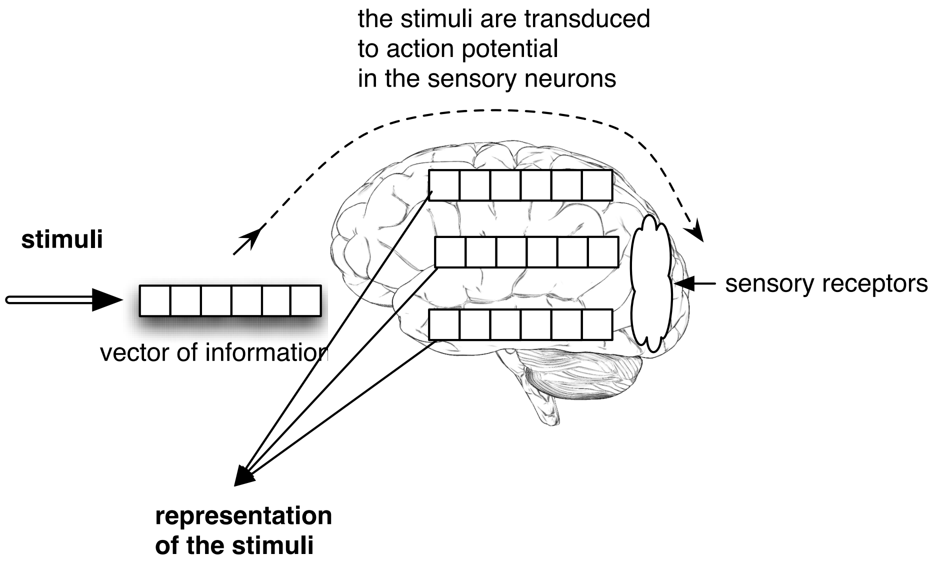

Fig. 7.3 The Information Processing Hypothesis IPH conjectures that the stimulus is directly represented in the spatial patterns in the brain. Thus, stimulus and representation of stimulus are isomorphic.

7.3 Computational neuroscience and mathematical neuroscience

Neuroscience owes a great deal to mathematics. The interpretation of how neural activity relates to behavior would be impossible without the capacity provided by mathematical/physical methodologies to quantify, generalize and convey rigorous and clear ideas. For the sake of clarity I will distinguish two main approaches; one is computational neuroscience and the other is theoretical-mathematical neuroscience (theoretical and mathematical neuroscience will be indistinctly used here). Computational neuroscience's aim is to build computational models that simulate phenomena occurring in the brain. The variables used in these models enable us to quantify relevant aspects in brain physiology like neural membrane potential, neural action potential spikes or neural firing thresholds.

Thus, computational neuroscience builds algorithms running numerical simulations of neural phenomena. The subject is vast and tremendously prolific, some basic introductory text are [236], Schutter:2001, Trappenberg:2002. For more advanced research, there are a number of journal devoted to this topic, such as, *Journal of Computational Neuroscience*, *Frontiers in Computational Neuroscience* or *PLOS computational biology* are some of the most well respected in the scientific and academic spheres.

We are interested here in the theoretical models of neural activity that help to elucidate the mechanisms leading to spatial and temporal patterns that ultimately, mediates in cognitive functions like perception or memory. [90]

The neural models that address perception, cognition and evolution may be grouped into three major classes:

i set of differential equations whose variables are key neurophysiological quantities
ii architecture that tells us how the neural structures are coupled together to form ensembles (meaningful populations, cognits. . .)
iii formalism that allows the understanding of cognitive processes and neural complexity

"In the future, when we describe mental disorders as circuit disorders, we can expect the same precision with which we distinguish conduction defects or arrhythmias in the heart"[237], [17]. Quantitative must not be confused with realistic. As a consequence of this habitual misunderstanding, computational neuroscience is, alas, equated with theoretical neuroscience. The reason lies in the wrong assumption that the principles of neurosciences are known. It is the fundamental principles which must be discovered through mathematics.

Theoretical neuroscience involves the creation of frameworks that integrate a variety of phenomena and can provide insight into the fundamental principles of brain function [80].

On the other side of the mathematical-like spectrum, we find theoretical neuroscience, whose aim is to create frameworks that integrate different phenomena, providing insight into fundamental principles of brain functioning[80]. In the brain the space that goes from molecular and cellular physiology to cell assemblies, up to behavior has not been mathematically *matched* yet. The matching that theoretical neuroscience is committed to is structural rather than merely quantitative.matching.

It deserves emphasis that theoretical and computational neuroscience are not the same thing, and neither one is included in the other. This misconception can be put this way; as the principles in the nervous system are known (just as they are known in, for example, robot dynamics) we do not need to introduce new mathematical concepts or theories, but to running simulations in order to attain a numerical matching between the data tossed by the simulation and real data. Let us use the example of a simulation of a manipulator arm doing some arc-welding. This allows the engineer to define which are the

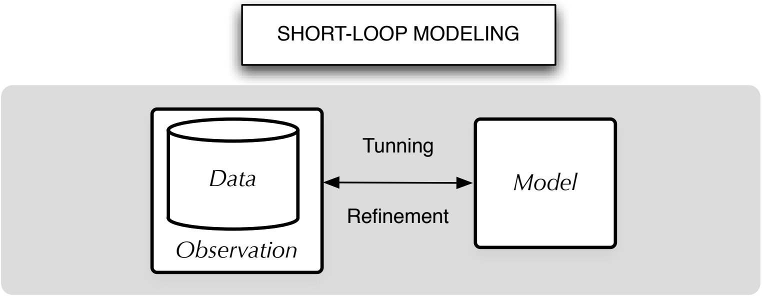

Fig. 7.4 The diagram tries to reflect the focus in the numerical matching without structural accuracy. The wide area between the cell physiology until the organism behavior requires theories based on mathematics that matches structurally the phenomenon and this must be done prior to do any computations [17]

trajectories that the robot can achieve and which are not. For example, the singular points must be avoided. Thus, the computational model of a yet unbuilt robot, makes it possible to understand the functioning and performance of the real robot. This is because the dynamics of a manipulator is known, and written in mathematical terms in 7.1.

$$\tau = M(\Theta)\ddot{\Theta} + V(\Theta, \dot{\Theta}) + G(\Theta) \tag{7.1}$$

where $M(\Theta)$ is the mass matrix of the manipulator, $V(\Theta, \dot{\Theta})$ is the centrifugal and Coriolis forces and $G(\Theta)$ represents the gravity force.

Recapitulating briefly, the laws are already known, and they are of course general, the dynamic equations are the same for a 1 degree of freedom as for a n degrees of freedom robot. A roboticist designs computational simulations in order to evaluate the performance of the robot. This is at odds with neuroscience, where the principles, in the mesoscopic and macroscopic levels, are still missing. To reduce theoretical neuroscience to computational simulation is putting the cart before the horse.

7.4 The levels of organisation in the brain

The brain systems operate at many different levels of organisation, each with its own scales of space and time. The brain, as any other physical system, is hierarchically organised. In a strictly compartmental view, the modeling of a system relies on a prior acknowledge of the different levels, later on the components at each level are separated and the intra level relationships identified. In this epistemology, the system is modeled as an ensemble of drawers, independent of each other. Every level is macro to those bellow it and micro to those above it.

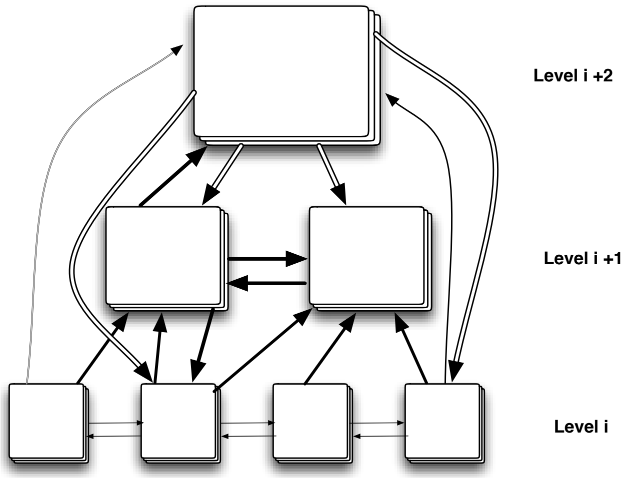

Fig. 7.5 The exchange of information between levels in complex systems, like for example the brain cortex, is not linear, rather the interaction is within the level's components, but also upwards and downwards.

But this picture falls very short of an explanation which deals with systems that are not modular. In a complex system such as the brain, the different levels interact and exchange information. As a result of the inter level interactions, the global activity is characterised by complex dynamical patterns, that may be chaotic as they occur in the brain of mammals.

However, precision is very important here. When we talk about the different levels of the brain e.g. microscopic level or macroscopic level, this is just a *façon de parler*. The levels are not real entities like a neuron or the waves generated by the brain in a particular moment. Rather, the measurement technique that is being used, imposes its own scale, and restrains the ontological reality that is being measured.

The main difficulty in modelling a complex system is that the coupling between their components can produce emergent properties, that is, properties that are not shared by the components. An emergent property is a non-local phenomenon that does not necessarily entail an insurmountable obstacle for its modelling. Admitting that a genuinely complex system, using Gerhard Mack's words [238], cannot be understood as a whole by looking at it locally, a reduction of its complexity is often possible only through local considerations.

7.4.1 Brain activity and measurement techniques

The spatiotemporal restrictions vary greatly between the different existing techniques of brain activity measuring. So magnetic resonance techniques (MRI), have a much better spatial resolution than electroencephalography (EEG), while EEG temporal resolution in milliseconds, prevails over MRI techniques which can take snapshots in the range of seconds. It should be noted that this is not just a technical issue that will be resolve with time. As soon as some technological improvement allows, the MRI scanners will be able to take faster snapshots. Indeed, the signal that is measured by the most common MRI technique, fMRI, peaks seconds after the neural firing begins in an area. So a better temporal resolution in fMRI faster, than the metabolic and blood flow, which is the phenomena that fMRI measures, won't make things any better.

The overproduction of fMRI studies that explore functional localisation, has led to an oversimplification and ad hoc explanations based on a new breed of phrenology. What fMRI really measures, is differences in the amount of oxygen in the blood in different parts of the brain. The nerve cells, when particularly active, consume more oxygen so there is a correlation between neural activity and the oxygen amount. There is no a general agreement about the real relationship between the neural activity and the metabolic signal measured in fMRI [239], [240]. Brain imaging techniques like fMRI, PET or SPECT, show metabolic patterns of blood flow which are subsidiary to the actual neural activity.

The habitual approach used in brain studies for the localization of the module of a particular behavior was:

- Focal ablation of the part where the functional aspect to be tested resides
- Focal stimulation to reproduce it and
- Electrical recording of the neural activity

This same strategy is followed in brain imaging studies (fMRI, PET, SPECT). The extreme simplicity of the phrenological conception, a concrete area local owns a particular behavior or cognitive ability, is at the core of its massive success in terms of the number of practitioners, experiments and results in the *fMRI discipline.*

Other techniques like EEG, ECoG or magnetoencephalography (MEG), which focus on the electrical and magnetic fields, provide data closer to brain dynamics than the metabolic data obtained from studies of exclusive brain imaging. EEG represents synchronized activity over a network of neurons and it is indeed the summation of synchronous activities of thousand or even millions of neurons that have a similar orientation. The potential fields measured with EEG are more accurate near the scalp than in deep areas of the brain. Indeed, the signal dramatically falls off with distances of the order, $radio_{w}aves^{4}$ [92].

The real challenge is to understand the mechanisms, from the micro level of neurons all the way up to the macro level of the machinery of assemblies of networks of neuron-systems, that underpins cognitive actions like perception or memory. The main difficulty arises in the dissimilarity of levels. On one hand, at the micro level of single neurons, EEG can provide all we need for building the dynamics of single neurons. On the other hand, at macro level, the MRI or PET spatial patterns of blood flow, provide localisation of the cognitive action, but say little of the dynamics of the brain area involved.

Another problem is that causal inference between levels is more ambiguous than within the same levels. An example of inference at the micro level, is that the stimulation of one neuron induces an evoked potential in that neuron. Par contre, at the macro level, what EEG shows is oscillations that represent synchronized activity over a network of neurons. Based on the special characteristics of the EEG signals, like frequency or amplitude, the neuroscientists infer particular states in the brain functioning like during walking or sleeping. For example, theta wave pattern (frequency range 4-7 Hz) is predominant in relaxed and meditative states and gamma waves (frequency range 30-100 Hz) in carrying out motor functions.

7.4.2 Linear vs Circular Causality

Conceptually, linear causality operates as follows. A stimulus S initiates a chain of events in a linear temporal sequencing. No effect can precede or occur simultaneously with its cause, and at some instant in the chain each effect becomes a cause. Linear causality can tell us how a reflex response is caused by a conditioned stimulus. But the reflex response is nowhere near being able to explain the causes of perception or any other fairly complex behavior. Linear causality is strictly determinist, A causes B causes, B causes C and so on. For example, a stimulus excites a sensory neuron; that excites the sensory cortex, that transmits to the motor neurons, that finally contracts the muscle.

Circular causality was initially suggested by Marleau-Ponty [241]. The action does not follow the sensation, rather both action and sensation are both cause and effect of the “intentional arc”. Circular causality expresses the interactions levels in a hierarchy. A hierarchical organisation like that of the brain has, top down constraints that apply to lower levels of the organisation, which in turn, constrain the higher level functions. In biology, it is very rare to find one function that is located in a single component. Rather the function arises from the interaction between several components. A possible exception that confirms the rule, is the haemoglobin that transports oxygen in the red blood cells of vertebrates. The interaction between components and levels in a system is at the core of the bidirectional causation that circular causality describes.

Circular causality in brain network is created by virtue of synaptic interactions between neurons. In linear causality, the stimuli are always injected into a passive brain which has attuned sensor to the stimuli. Contrary to this view, the concept of circular causality is necessary to cope with the self-organising chaotic dynamics within the brain. In section 7.7.2 which is devoted to neurodynamics, I explain how self-organised patterns of neural activities are produced.

For a critique of the idea of Circular causality see [242].

5Technique name Spatial resolution Time resolution Pros&Cons

7.4.3 Granger Causality

Granger causality is a term for a specific notion of causality in time-series analysis. Originally, the term was coined by Norbert Wiener and the Nobel laureate economist Clive Granger, who applied the concept to economic realm. In neuroscience, Granger causality analysis is used to understand the relationship between network structure and network dynamics. It might be remarked that the term network dynamics, in the context of graph theory applied to neural systems, means network function. Data regarding local field potential and recordings using EEG, MEG or fMRI have received a G-causal analysis. Causal network analysis provides a radically new insight into the mechanisms underlying higher cognitive functions, including consciousness. The neural correlations are not explanatory and there is an urgent need for theoretical concepts that connect the phenomenological aspects with neural dynamics. Causal network seems to be useful in revealing the general principles governing causal interactions between long distance neural populations and also as a framework to test the functional connectivity.

“Causality analysis makes no assumption about information processing or neural code.[243]” However Granger, causality has been more successful in generating heuristics for the analysis of biological data in bio inspired robotics, for example in the robot Darwin-X, than in building theoretical framework for the understanding the global functioning of the brain.

Some drawbacks of G-causality are: linearity, stationarity and dependence on observed variables. The causal core analysis requires exhaustive knowledge of structural and funtional connectivity which, currently can not be supplied with real data, except for simulated environments.

On the positive side, the “causal approach” will strengthen the mesoscopic analysis in brain connectivity and dynamics. This is a necessary step because it has historically received minor attention if we compare it to the microlevel of single neurons, or with the macroscopic level of brain systems.

7.5 A Theoretical Framework for Connectivity in Biological Networks

During the 40's and 50's the study of biological systems acquired a totally new dimension. The pioneering work of cyberneticians like N. Wiener or D. Cannon started a conceptual revolution in the study of life systems. The introduction of mathematics and in particular control theory in the modeling of biological process, made a rigorous understanding of principles like homeostasis possible. Homeostasis is defined as the property of a system that regulates its internal environment and tends to maintain a stable condition. This is now accepted as a principle in physiology[2]. Living systems use control methods, hence we can use control theory to tease out the underlying principles of such systems [9].

Theoricists like N. Rashevsky, in his effort towards a quantitative biology, steer a new way of thinking in biology. The search for general principles in biology gives rise to a whole new discipline, mathematical biophysics. Examples of the principles stated by Rashevsky are The Principle of Optimal Design (POD) and the General Principle of Relation Forces (GPRF). POD states that an organism, for a set of prescribed functions, has the optimal design with respect to the material used and the energy expended. It should be noted here that the optimal design of Rashevsky, is not global but local, that is to say, the existing structure is optimally modified by a function. A completely new structure is unlikely to appear, for example a thumb in bear[5]. In GPRF, the development of an organism tends to increase the number of relations between the organism's components, as well as the diversity in those relations.

In this section I argue that the interactions, rather than the components, constitute the very essence of the biological systems. Biological systems are defined by the interactions between their components [244]. It is relevant, for instance, that the biological components has an expiration date. Thus the components of a cell at one instant t, represents only a tiny fraction of the cell later on, at some instant $t + k$. As biological systems are hierarchically organised, the same goes for cells within the tissues or for the individuals that form a society.

In order to determine the functional state of a system we need to identify and properly described all the links in the network. By network I mean an abstraction that represents the relationships between objects. In biological systems, the networks represent interactions within and between biological systems. The networks are typically depicted as directed graphs and capture

---

[2] The paternity of the concept of homeostasis is disputed between Claude Bernard and Walter Cannon. However the concept comes from Bernard's *milieau interior* which states that the constancy of the internal environment is the condition for a free and independent life

the result of evolutionary pressure upon the biological system. It is interesting to note that Darwin himself used this kind of formalism[124].

Every cell contains networks of thousands of interactions. More precisely, within a cell, the links between molecular components are chemical reactions or associations between chemical components. In the brain, the links are synapses. Despite the gloomy complexity that biological networks can present i.e. non linear interactions of highly coupled components, which have undergone continuous structural changes driven by external forces; we can formulate general laws of connectivity that apply to biological networks. Thus, the details of the components may not be needed to understand the functioning of the system and its evolution. According to Polsson [244], the interactions are the dominant variables in the state of the system. However, even if we could characterise all the links in a biological network, the phenotype of the biological system would remain uncertain. This is because the possible functional states in the network, grow much faster than the number of its components or nodes.

For example, in the human brain, a network would contain 10^{10} nodes or neurons and 10^{14} links or synapses. It goes without saying that apart from the illegibility of such diagram, practical issues like knowing which neuron is connected with which, or what is the shortest path between two distant neurons, pose problems of non computability. In order to be operational, the numbers of degree of freedom need to be drastically reduced. Let me emphasise this point. However, in order to understand brain structure and function, we need to search for its patterns. This search will be futile if we do not constrain the search. What it is necessary is to surmise general principles of the system, for example: stability, robustness, adaptability. The constraints are embedded in those principles.

So far, I have aimed to make clear, that in essence, a biological system is a network of interrelated elements and that the relationships between the components play a major role. However, we must also remember that the components themselves may not be neglected. In the neuroanatomy of the cortex, the problem of how the cortical circuits work is still ontological. Stated simply, the different types of neurons, their morphologies etc. must be precised. It is worth reminding ourselves that a complete synaptic map of the human cortex has not been accomplished yet.

As an example of large variety of nerve cell types, the worm C.elegans has only 302 neurons, despite this short number of nerve cells, White et al. [245] have identified, nothing less than, 118 different classes of neurons. The largest class has 13 neurons which have identical morphology and connectivity patterns, but many classes contain one single neuron.

The current technologies for the morphological exploration of the brain has lead to a embarrassment of riches. The few types of neurons identified by Cajal and his pupil Lorente de Nó, provided a simplified taxonomy of neurons. Today, as parameters that can be measured with new technologies increase in number and precision, the taxonomy becomes unmanageable. Attending to the different neural properties that can be measured, it is possible to characterise many classes of neurons, probably many more than what are really necessary.

The criterion behind the emphasis in the interactions, is that the gulf between structure and function, can be bridged exploring the interplay between connectivity and dynamics. The problem with delineating connectivity is that in forming synapses, cortical neurons are promiscuous. Given that a neuron is polineurally innervated what we need most is to identify the presynaptic neuron of any particular synapse. Botzenger [246] showed that for some layers is at least possible to account for most of the synapses, assigning a source to every synapse.

This leads us at once to the question: how spatiotemporal distribution of neural activity is shaped by the underlying network connectivity? In [247], is suggested that brain changes of connectivity is a major mechanism for learning.

7.5.1 Defining Brain Connectivity

Kids in the school learn that we have five senses: taste, smell, sight, sound and touch. However, such classification, though didactically useful, is in the words of Patricia Churchland "notoriously inept [248]". Indeed, the nervous system has receptors not only for the *five senses* but also for different things, like for example, detecting changes in the position of the head, detecting changes in the blood pressure, or in the oxygen level in the arteries. Besides, the classification is careless with the senses of other animals. For instance, bees can detect ultraviolet lights, and a magnetic compass for positioning has been demonstrated in 18 species of migrating birds [249].

The senses are located in particular brain regions where specific type information is processed. The sensory neurons in those areas typically form specific maps of the receptor sheet of cells in the proper organ -e.g. maps of the body surface (touch), the retina (sight), the olfactory bulb (smell) or the basilar membrane of the cochlea (hear).

The understanding of these maps is the first stage in the process of representation in the brain, of the world in which we live. The neural maps for sensory perception seen above, and also those representing the limbs' musculature and specific movements are more or less are well-known. What is still not fully understood is the ways in which these maps are interconnected. If we know which are the preferred stimulus features for single neurons, we can infer the function of a local assembly of neurons form its constituent neurons. That is to say, only when we have a 1 to 1 mapping stimulus/sensory-motor neuron, the anatomical structure of the neural set will be enough to determine its function.

Of course, this fortunate coincidence of the anatomical and functional connection is marginal. A distinction between both is mandatory. However, before getting into the difference between anatomical and functional connectivity, it should be noted that this demarcation is not exhaustive and other approaches are possible. In [250] a distinction is made between *effective connectivity* and *functional connectivity*. The former is akin to the neuroanatomist's effort in finding out how neurons connect each other and form coherent networks and the last is associated with neurophysiology in designating *units* of functional specialisation. More precisely, *functional connectivity* means the temporal coherence among the activity of different neurons, measured by correlating their train spikes and *effective connectivity* is the simplest neuron-like circuit that produces some temporal relationship as observed experimentally between two neurons. One possible example of *effective connectivity* is the axon and the muscle fibers it enervates, which in [223] is referred as *motor unit*. Each muscle fiber is one single motor axon that innervates several muscle fibers. For example, in the leg we find circa 1000 muscle fibers, while in the eye an axon is connected to fewer than 100 muscle fibers.

The *effective connectivity* can be traced by studying the sequence of signals that produces a reflex action and is transmitted all the way down the axon to the muscle fibers. Let us see how. When a muscle is stretched, the amplitude and duration of the stimulus are reflected in the amplitude and duration of the receptor potential in the sensory neurons. The graded signal is transformed an all-or-none signal. This happens only in case the receptor potential exceeds the threshold for action potentials in that cell. The more the receptor potential exceeds threshold, the greater the depolarization and consequently the greater the frequency of action potentials in the axon; likewise, the duration of the input signal determines the number of action potentials. The frequency and number of action potentials, is the actual information conveyed along the axon's length to its terminals and determine how much transmitter is released from the sensory neurons to the motor neurons. The transmitter interacts with receptors on the motor neuron to initiate a graded synaptic potential.

If the membrane potential of the motor neuron reaches its threshold, an action potential will be generated and propagate to the motor cell's presynaptic terminals where the action potential causes transmitter release, which triggers a synaptic potential in the muscle. That in turn, produces an action potential in the leg muscle, which leads to the final muscle contraction and an overt behavior. This sequence of transformation constitute an example *effective connectivity* with one sensory neuron and one motor neuron.

7.5.1.1 Anatomical Connectivity and Functional Connectivity

Anatomical connectivity is a structural linkage such as synapses or diffusion channels between two neurons or groups of them.[20] The anatomical connectivity is spatially structured. The combined activation, through synapses, of discrete and local assemblies of neurons give rise to spatial connections. One anatomical local area connection may admit several functional implementations. This means that connectivity is not solely spatial but temporal as well. The arrangement of neural networks i.e. the synapses, vary during the realisation of cognitive operations. Thus, there is not a clear cut border between structural and functional connectivity maps.

To exist, functional connectivity, needs two prerequisites; one is anatomical connectivity which must be already in place; and the second is a function that sets up co-active states between the neurons. The more complex the cognitive functions are, the more predominant the functional connectivity is over the anatomical connectivity. Thus, even with the astoundingly detailed and precise knowledge of the brain anatomy that technology is now making possible [62], this won't lead us to the understanding of the principles underlying all cognitive function, on its own.

One habitual approach to get the neural substrate for the implementation of cognitive function in the brain, consists of identifying specific local networks that participate in cognitive operations. Functional imaging techniques have led to important advances to this respect [251],[252],[253],[254]. However, due to the poor temporal resolution of this type of technology, it is easier to establish associations between large scale areas and highly integrated cognitive functions, than to identify the local brain areas specialised in medium-low level cognitive functions.

We know that a local area can be related to several functions. Thus, any attempt to establish the function of local area is bound to be rather arbitrary. The function of a local area network, depends on its pattern of synchronous interconnectivity with other local networks. The patterns of the brain and its dynamics are the subject of section 7.6.

7.5.1.2 Reformulating Hebbian rule for learning

Francis Crick, in [255] claims that in the neocortex the correlation of the firing of neurons is stated differently, neighbor neurons seldom fire precisely at the same time. But by the time Crick wrote this, no in vivo synaptic mapping had been made. Thus, assuming Hebbian principles, Crick derived that if neurons fire together, the also must necessarily wire together.

However, for the principle of economy, it is more likely that there is a connection between two neighbor neurons, than between two neurons that are far away. Obviously, if they are far, their exchange of information is more energy consuming and slow than if they are close. The real significance of in vivo imaging of neurons, is that it dramatically challenges the "neuron-point" notion. The cytoarchitecture and dendritic arbors become relevant for modelling the physiological action of neurons.

A neuron is a cell body (soma) and a set of projections; axons and dendrites, which extend out of the soma. This picture is a rough generalisation, however. For example, there are neurons with axons and no dendrites, and in the olfactory bulb we find neurons without axons and dendrites [248]. In [256], Bullmore and Spoorns, using diverse experimental modalities in humans (structural and functional MRI, diffusion tensor imaging, magnetoencephalography and electroencephalography) show that neurons and brain regions that are spatially close, have a relatively high probability of being connected; whereas connections between spatially remote neurons or brain regions, are less likely linked. More evidence of this functioning principle can be found in [257], [258], [259].

Cortical neurons, except when they are driven by stimulation, give very low correlation coefficient of their time firings with those of their neighbors (0.01 aprox.) [260]. This means that the fraction of the variance of the activity of a single neuron that is covariant with the neighborhood is 1 in 10,000.

The basic biophysical process at the root of memory formation is the modulation of transmission of information across synapses. The cortical representation of our internal and external world are built by modulation of contacts between neurons. It is through the comprehension of the connections between the neurons that we can predict and control the behavior of the single cells and the networks formed by them.

If there is one thing everybody agrees upon, and in neurosciences we do not have that much parsimony, it is the central role of synapses in the making of cognitive networks. However, the mechanism is not yet known by which a synapse becomes a link, or a part of a link, in a cognit representing a memory or a feature in the environment. Indeed the Hebbian rules do not exhaust the possible explanations for synaptic linkage of networks of neurons with *representational content*. Besides, in the formation of networks, the neurochemical modulation, neurotransmitters like GABA (mostly inhibitory neurotransmitter but not exclusively like Crick wrongly assumed) or NMDA, which induce long synaptic currents facilitating the integration of non synchronic inputs. are involved

To sum up, Hebbian principles are a still valid in guiding the understanding of the formation of neural networks with cognitive significance, but as likely as they are, they are not exclusive nor universal. Hypothesis from inspired theoricists are now more pertinent than ever in neuroscience.

7.5.2 Characterisation of Network Topology with Graph Theory

The study of complex networks focusing on network topologies and using graph theory concepts, has provided a fresh insight into the problem of neural organisation. Since Watts and Strogatz' article in Nature, in 1998 about graph theory and its application to complex biological networks[261], biologists began to see network theory as a promising theoretical framework for quantifying the structural and functional properties of biological networks. The network theory provides a quantifiable framework, making explicit terms like connection density, path length or node degree distribution, that may be used to explore the structural and functional connection properties in the brain.

According to Ermentrout [262], it is possible to write the equation of the network dynamics for point-like neurons (dendrites are ignored). Let $v_i(t)$ the membrane potential of neuron i at time t and $u_i(t)$ the firing rate of neuron i at time t. It is normally assumed that the firing rate $u_i(t)$, is a function of the membrane potential $v_i(t)$. Thus $u_i(t) = F_i v_i(t)$.

The dynamics of $v_i(t)$ is given by

$$v_i(t) = \int\limits_{t_0}^{t} G_i(t-s) \sum_j w_{ij} F_j(v_j(s-\delta_{ij}))ds \tag{7.2}$$

where w_{ij} is the scalar that represents the synaptic weight. The connectivity matrix w defines an effective geometry in which the dynamics of the system evolves. Here, w shows the network topology but does not say anything about the neural mechanisms that qualitatively affect the dynamics of the network.

Let us give now some introductory notes about graph theory that we will need next to identify some relevant network topologies. A graph is a collection of nodes N and a collection of edges E, $G = (N, E)$ that connect pairs of nodes. A path is a sequence of edges that connect nodes. The number of connections a node i has is called degree, D_i. The degree distribution P_k is the probability that a node has k connections. The average path ap is the average of the shortest path between of every two nodes in the graph. The complete description of a particular graph is provided by its adjacency matrix w. Each of its elements w_{ij} characterises the connection between the node i and j, for example the number od edges among these two nodes.

Taking this into account, I introduce three kind of network topologies: random, scale free and small world.

7.5.2.1 Random network

The main characteristics of a random network are: every two nodes have the same probability to be connected, most of the nodes have a degree close to the average degree and the degree of distribution P is Poisson. A random graph, also called Erdos graph, is a statistical ensemble of all possible graphs of N nodes and E edges, where each member of the ensemble has identical probability of realisation. A random graph is obtained by starting with a set of nodes and adding edges between them at random, as a result an interesting property is generated: every possible edge occurs independently with probability p. It ought to be said that "random" refers to the assignment of a probability value to a link between each pair of nodes, not to lack of structure.

7.5.2.2 Small world network

Small world networks have as main features high clustering and short paths. Put another way, most nodes are not neighbors of one another, rather they can can be reached from every other by a small number of steps. In small-world graphs a low percentage of the local connections is replaced with uniformly distributed long connections [263]. Small world graphs characteristics, high clustering and short path, can be seen in cortical connectivity. Indeed axons projecting long distances are few in number than local axons forming clusters of neurons. An implication of the key features in small world networks -high clustering and short path- is that they provide high speed communication channels between distant parts of the brain, facilitating dynamic processes that require global cooperation between neurons, like for example occurs in neural synchronisation [264].

Small world networks are more clustered than Random networks in the sense that if for example A is connected to B, and B to C, then it is very likely that A is connected to C. Thus, transitivity ($A \rightarrow B$ and $B \rightarrow C$ then $A \rightarrow C$) is a common property in small world neworks

According to [265], [266], [267], [268], the brain's structural and functional systems have a small-world topology, this means highly connected hubs and modularity, both at the whole-brain and at cellular scales. Here I would caution that presently, the only complete connection matrix that has been accomplished is that of the *C.elegans*, a nematode worm of only 302 neurons linked by some 7,000 synapses. The *C.elegans* has a special property that, it goes without saying, animals more complex than this worm, are lacking. This is the one-to-one map between the cells of the central nervous system in different members of the *C.elegans* species.

7.5.2.3 Scale free networks

Scale-free networks is a very interesting example of complex network topology. The degree of distribution, p_k, means the fraction of nodes that have k links. Thus, p_k quantifies how much connected are a node compared to the other nodes. Differently from random networks seen above that have a Poisson distribution, the scale free networks have a heavy tailed distribution, p_k decays slower than in a Poisson distribution.

$$p_k = k^{-\gamma} \tag{7.3}$$

where $\gamma = [2.1, 2.4]$

Scale free networks are fractal in the sense that no single characteristic scale can be defined [269]. They have structural self-similarity, or borrowing the expression from modern physics, scale free networks present symmetry under scale change.

In the cortex, scale-free dynamics are characterised by hierarchical self-similarities of patterns of synaptic connectivity and spatiotemporal neural activity [18]. One functional advantage of this kind of network is its resistance to random failure, only crashes in hubs are critical to the global network functioning. Of course, this same advantage turns out to be a weak spot when one hub fails.

7.5.3 Connectivity and Network Dynamics of the brain

At this point, it is crucial to realise two things. First, the literature of brain networks analysis using graph theory [256], [261], [270] assumes that network dynamics is determined by its connection topology. Thus, the dynamics changes as the connectivity matrix changes. So, in the network dynamics equation 7.2, the membrane potential of a neuron depends on its synaptic weights which are quantified in the matrix w_{ij}. But the only way of generating such association matrix w_{ij}, is by compiling all pairwise associations between nodes and apply a threshold to each element of this matrix. Hence, depending on the threshold that one chooses, different network topologies arise. And the problem is that the threshold is arbitrarily selected.

Second, and more importantly, there is caution word to say about the use of the term dynamics. In the “small world” literature, dynamics is used as the algorithm that puts in order the network structure, that is, how the connections between the nodes are modified over time. On the contrary, this paper defends a more fundamental interpretation of dynamics. Here, the dynamic nature of biological systems is understood, not in the customarily sense as the organisms’ evolution over time, but as the capacity to organise invari-

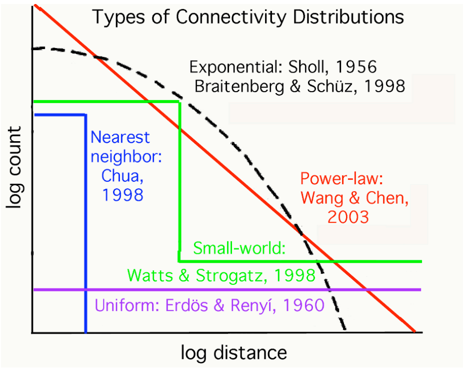

Fig. 7.6 The drawing is from [18] and shows five types of connectivity distributions. Uniform: the distribution is constant, nearest neighbor: nodes have the same degree of connectivity or 0, Small world: high clustering and short path, Power law: Poisson distribution and Exponential

ant patterns. In chapter 7.7, dynamics is explored as the mechanism that underlies the organisation of patterns in the brain.

7.5.4 Regular vs Complex network topologies

Lets remember the objective that we are trying to achieve here, is to characterise the neural operations by which patterns of activity are created. As stated in chapter 7.7, these patterns are dynamic structures evolving over time. The difficulties we are facing are those that we find in the complex systems. Suppose we have identified and modeled different subsystems, so we are able to predict the future behavior of each subsystem treated separately. Now if we plug them together, we no longer have the accurate control we had when we dealt with the components separately. The coupling between the different components makes the behavior of the resulting system not directly reducible to their simple components. As the layman would say: the total is not just the sum of the parts. Circular causality rather than linear cause-effect; and synthesis, rather that mere analysis, need to be in place before dealing with complex systems.

However, when facing a complex system, a successful strategy is to suppress certain complications while highlighting others. More precisely, in the study of large scale neurocognitive networks in the brain, we can identify two main approaches. One is focused on the dynamics of the system and the other concentrates more in the network architecture. Both approaches are sketched in the next two sections.

7.5.4.1 Dynamics in regular network topology

In the study of dynamics in regular network topology, we assume that the network architecture is regular and static. This facilitates the study of nonlinear dynamics because we suppose nearly identical dynamical systems coupled together in simple geometrically regular ways. Bellow there are four good reasons for omitting the network complexity and focusing on local dynamics [264].

1. Structural complexity: the anatomical connections can be extremely tangled. This is problem in neural network because, in forming synapses, cortical neurons are very promiscuous.
2. The plasticity of the network: the wiring is not static, it changes over time, for example in the learning process.
3. Diversity in nodes and links: There are several kind of neurons and synapses. There are several ways to classify the neurons:
   a. Based on the morphology: pyramidal and non pyramidal.
   b. Another possible classification is according to the spines on its dendrites: approximately the 80% of neurons are spiny cells and the 20% are non spiny.
   c. The neurons can also be classified functionally in sensory, motor and interneurons.
      i. Sensory neurons: transduce physical signals into electrical signals
      ii. Motor neurons: innervate the muscles to produce contractions. They have large cell bodies and complex dendritic trees.
      iii. Interneurons: connect sensory neurons (afferent neurons) and motor neurons (efferent neurons) in neural pathways

   The synapses are classified morphologically into two types:

   a. Type I: The membrane is asymmetrical, thicker in postsynaptic cell than in the presynaptic one
   b. Type II: These synapses have symmetrical membranes

This morphological classification has a major interest in physiological terms. Remarkably, it seems that most type I synapses excite, while most type II synapses inhibit [255].

4. Dynamical complexity: the nodes could be nonlinear dynamical systems.

In addition, these four complications do not act isolated from each other, but rather they are connected. For example *ii.* affects *i.* because the anatomical connectivity varies with changes as the network evolve over time. Neurons are created and die (apoptosis). To sum up, we can always assume that the network is regular and static and focus on the dynamics of systems coupled in uniform ways. The dynamics of local collections of non interactive neurons can be represented with ODEs, recurrent feedback relations model the relations between such local collections. The interactions between excitatory and inhibitory groups of neurons create a mesoscopic order [271], [272] that is typically tackled with Chaos theory in terms of state transitions between wings of global attractors.

7.5.4.2 Complex network topology

Now that the dynamics of systems has been addressed, we put them together in order to incorporate the network complexity aspects. But here lies the problem, using Strogatz's words

> Unfortunately they lie beyond our mathematical reach - we do not even know how to characterize their wiring diagrams. So we have to begin with network topology.

It is needed a different approach, complementary to the one seen above, that copes with in the dynamics assuming that the network is regular. Now, what is at stake is the network architecture. Thus, we sidestep the dynamics in order to deal with complex architectures. I am using here architecture as identical to topology. In the subheading 7.5.2, I sketched complex network topologies using graph theory concepts and methods. Namely, random networks, small world networks and scale free networks.

I posit the idea that a mathematical breakthrough is necessary. The way in which non linear dynamics of biological systems are coupled, according to small networks or scale free networks, is a path worthy to be followed in order to explain and predict how meaningful patterns -population of neurons subserving cognitive functions- are formed and evolve. In order to encompass the two complementary approaches above, we need to discover the organisational principles. Because structure necessarily affects function, the generic properties, rather than properties of particular nodes in the network topology, need to be worked out. I claim that principles like stability, robustness and scale free, which are mathematical abstractions, as well as features that can be found in the biological systems, must guide the move towards the mathematical progress in brain sciences.

7.6 The search for patterns in the brain

The term pattern is one of the most prominent in science today. For example, crystallography studies the symmetry properties of crystals based on a particular arrangement of atoms or pattern; in psychology and ethology the behavioral pattern of an animal is common jargon; and in computer sciences pattern recognition is one of the most prolific research areas.

The use of pattern in this section is restricted to the synaptic connectivity in the brain cortex. As it is well known, the brain cortex is a most important structure for human cognition.

The patterns of connectivity in the forebrain are not merely the effect of processes such as memory or learning, which are possible due to the plasticity and developmental characteristics of the brain. Additionally, to the experiential and ontogenic aspects just mentioned, there is also a previous phylogenetic pattern formation. It is evident that animals are not get born with a tabula-rasa, rather we have, from the moment of our birth, basic patterns of anatomical connectivity that represent basic actions and percepts. The specificity of these patterns of anatomical connectivity is due the phylogenetic modulation, common within the same specie [273]. Morphologically, the entire anatomical architecture of the cortex of a mammal, though not complete yet, will be eventually achieved thanks to the improvement in the current visualisation techniques. This claim is not a manifestation of positive thinking or a religious wish, rather it is rooted in the fact that brain connectivity is highly generic and redundant [274]. But, as mentioned before, a complete map of the intricate anatomical connections, is not the same as a functional map. The anatomical patterns undergo continuous changes over the life span of the animal. Contrary to the innateness of the phylogenetic patterns, the ontogenetic patterns are created through synaptic modulations, elicited by cognitive factors such as emotion, memory or attention.

According to [275], [276], scale free and small world properties have been found in cortical connectivity. The *Caenorhabditis elegans* is a worm which is particularly relevant in brain studies because it is the one species with a complete neural map available[277]. In the seminal article of Watts and Strogatz[261], they demonstrated the existence of small-world network properties in the *C.elegans*. One decisive value of this paper, is that establishes a quantitative analysis of complex networks and translates it to brain network, concluding that at anatomical level, the large-scale connectivity of *C.elegans* brain is not random or complete but small-world.

An attempt of this kind is powerless when dealing, not with anatomical level of fixed connections at in the *C.elegans*, but with the functional connectivity and its dynamics. Now one remark is in order, practically any cortical neuron or neuronal assembly can be part of many larger different assemblies of neurons of functional relevance. Thus, at first glance, while the anatomical arrangement is regular and fixed, the functional one manifests more as a widespread constellation of activated areas all over the brain than as an ordered arrangement. This seems to suggest a strong contrast between anatomical and functional connectivity. The way to conciliate this tension is to acknowledge that the understanding of neural organisation requires the integration of the anatomical, functional and dynamic approaches [278].

7.6.1 A first approach to meaningful patterns in the brain

The cognitive operations are represented in functional maps which are implemented in neural structures that evolve over time. The challenge is to identify those maps of neurons that represent the cognitive actions. On this basis, the term meaningful pattern is defined as a neuronal network with cognitive representational power. Stated another way, meaningful patterns are dynamic structures i.e. neuronal networks, that express cognitive operations. In the brain, these patterns are meaningful neural networks that transcend the anatomically observed modules of the brain. They are meaningful in the sense that they are functionally relevant.

The concept of the functional organisation of the brain has undergone a remarkable change in the last decades. The concepts of representation, pattern and dynamics are pivotal in the understanding of the neural mechanism of the functions of the brain.

It is very important to remark that I am using here the term cognitive not in the usual way it is used as mental ability, but in a more radical sense, keeping away from arguable terms like mind or intelligence. Here, a cognitive state simply means an internal representation of a relevant feature for the survival of the organism. Thus, in this framework, a cell cognizes. More precisely, a cell represents environmental states through a set of proteins called transcriptional factors, which represents internal states of the cell that are important for its survival in a complex and changing environment.

Here, I defend the view that it is at the level of internal representations of the external reality, where the gulf between biological basis of for example, perception or learning, and the mental constructs associated with those mental aspects can be bridged. These internal transformations are driven by dynamic neural processing mechanisms, resulting in patterns of neural activity.

The study of perception has held as a central position in philosophy, psychology and of course brain sciences. So, a lot is known about how our cortical areas e.g. auditory or visual, transduce and internally represent the external stimuli. These cortical sensory areas are defined based on cytoarchitectural properties. For very simple and primitive features of the external world, the final pattern of neural activity is isomorphic with the perceived reality. In this case, the internal representation of the world is mainly topographic. But the functional properties of neurons loose their specificity as soon as the internal representation is not isomorphic, so the topographic map is superseded by patterns of connectivity able to deal with more abstract features.

To sum up, the neural representation of holistic perceptions are embedded in recurrent patterns of connectivity. In order to understand the neural mechanisms in a working brain, we need to challenge the concept of functional representation. An agenda that acknowledges the crucial importance of the representational power of patterns of connectivity and its dynamics must be put in place.

7.6.2 Meaningful Patterns in the brain

A main premise is that all cognitive function takes place within and between neuronal networks, mostly in the cortex. This assumption and how to deal with it, is at the core of the gap between the neural sciences and the cognitive sciences. The mind/brain problem is the philosophical stand for the same gap.

The above assumption can turn into a *luogo comune*, a old ancient one indeed, when not addressed with the appropriate language. In the cognitive sciences, cohabit concepts extracted from levels of description as disparate as the mathematical models of activity of the neurons or even deeper, like the molecular composition of neurotransmitter to the macro level descriptions -i.e: in plain English- of emotions or social behavior. Surely, a new language is needed to help bridge that gap.

It is now easy to see that the term meaningful pattern, that was introduced above, relies on the characterisation of the cognitive structure of a neuronal network. Similar to the concept of meaningful patterns defined here are the term *cognit* coined by the UCLA based neurobiologist J. Fuster[279], the neural populations at the mesoscopic level in the olfactory bulb studied by Walter Freeman[20] or the concept of *neurocognitive networks* conceived by S. Bressler as large-scale systems of distributed and interconnected neuronal populations organised to perform cognitive functions [273].

Fuster's *cognits* are the structural substrate of cognitive operations and represents items of knowledge. Thus, a *cognit* is a cognitive representation that evolves over time due to the maturational and experience-based process undergone in the brain.

In Bressler's view, in order to understand how large scale neural structures organise to perform cognitive functions, we must be able to first identify the meaningful networks or patterns that express those cognitive functions.

Freeman's approach is mainly concerned with the introduction of an intermediate level, the mesoscopic, in which the dynamics of large number of neurons is studied through empirical models in the olfactory system. The olfactory system is the simplest and phylogenetically oldest of all mammalian sensory systems. The proper element for mesoscopic modeling is the *local population* rather than the more general concept of network of neurons. Contrary to *local populations*, the network of neurons are not necessarily local. Indeed their neurons can be separated by large distances in the brain, for example, networks connecting neurons of different hemispheres or cortico-talamic networks. The locality property utilised in Freeman's mesoscopic models, facilitate their own simplification, because the dendrites of neighbor neurons, as opposed to the axons, whose operation is non linear, can be treated as a linear integrator. So, focusing on locality we steer clear of the non linearity in long distance axons connections.

7.6.3 The multilevel hierarchy of meaningful patterns

Within the micro level of single neurons, the relevant quantities to be measured typically are the firing potential and membrane potential of neurons. Within this level there is a direct Input-Output recording that conveys a linear causality between two observed events. In other words, there is an isomorphism between the stimulus intensity and the firing rate of sensory neurons. More precisely, there is an isomorphism between the firing rate of motor neurons and the degree of contraction of the muscles. The models at neuronal levels benefit from the linear causality.

The micro level approach, is based on taking data from the I/O pairwise observation which is embedded into the linear causality schema. Despite its success in unraveling the structural and functional properties of neurons, it fails to explain how assemblies of neurons can express cognitive function. One major consequence of this is that the formation of patterns at the level of populations of neurons , that is the pattens P_{L_i} at level L_i , though arising from an inferior level in the hierarchy, L_{i-1}, for example single neurons, cannot be explained directly from the immediate inferior level patterns $P_{L_{i-1}}$.

The micro level patterns are occasioned by the coordinated and synchronised activity of single neurons. While at the micro (neuron) level the causality is linear; at superior levels it is circular, and the I-O relationships are not as clearly defined as in the microlevel. The epistemology in the macrolevel is necessarily different than that of the level of single neurons, which has clear-cut components. Contrarily, in the macrolevel, the components are inherently coupled with multiple feedback pathways.

It ought to be noted that meaningful patterns are not only associated with the macrolevel neural basis of cognitive functions such as language or logic reasoning. There are meaningful patterns at the level of single neurons as well, for example in simple sensorimotor actions. In the mesoscopic level of *local population*, a meaningful pattern, is that that precedes the stimulus that the subject has learned to expect. As it is referred in 7.2.3, the bulbar patterns do not relate directly to the stimulus, but to the meaning of the stimulus.

7.6.3.1 What is a pattern made of?

The role of patterns is of the greatest interest for the understanding of the functional architecture. However, one precautionary remark should be made before making headway in this direction.

A pattern, as almost any other structure, is made up of its components and relations between them. This very same postulate sets up the agenda of the science of systems. Put in neural terms, a pattern is made up of components and relationships, namely, neural assemblies and the connections between them.

Thus, the research proceeds by first finding a meaningful elementary unit and then developing consequences, that is to say, seize the linkage between those neural assemblies and understand how higher order assemblies *emerge* from the interactions between simpler components.

An attempt of this kind may be understood in at least two ways. In the *bottom-up* approach, the content of networks is reducible to their components, and these reducible to their inputs. In this approach, the periphery governs and the meaningful patterns are 1 to 1 maps of stimulus-neurons. On the other hand, in the *top-down* approach, the patterns, are not directly related to sensory and motor neurons as occurs in *bottom-up* approach, rather the representational networks are the product of complex, non-linear and highly coupled interactions between neural assemblies sparsely distributed.

Almost needless to say, the patterns are not fixed, but evolve over time, for that purpose, and in addition to the identification of the meaningful pattern we need to understand its dynamics. In chapter 7.7 a detailed account on neurodynamics is given.

7.6.3.2 Where the patterns are found?

In 1957 V. Mountcastle discovered the columnar organisation of the cortex, put in his own words *there is an elementary unit of organisation in the somatic cortex made up of a vertical group of cells.* However the concept of cortical columns was not completely new, for example Berger in 1929 using the rudimentary electrophysiological techniques of the time, showed that cells of similar function -e.g: Primary-Visual, Motor, Somatic, are grouped in vertical bands. For a historical and condemnatory account of the cortical columns see [280] [281], [282], [283].

The idea of the cortex arranged in the form of regular columns proved very appealing and subsequent research produced even more. Huber and Wiesel renowned studies in the visual cortex confirmed the discovery of Mountcastle. They reported the existence in the visual cortex of orientation columns. The problems came when trying to put borders on the columns. For Mountcastle, Huber and Wiesel, the cortical columns are discrete slabs with a sharp border between them. Inside each column, all the cells share some common salient

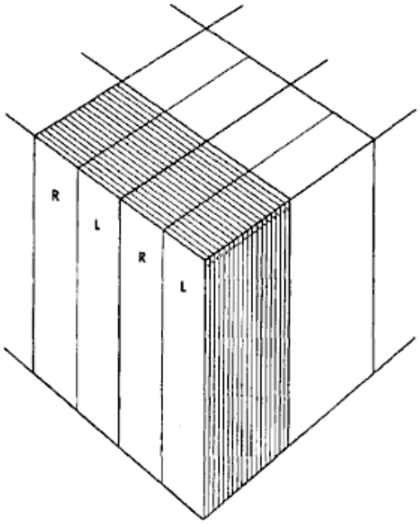

Fig. 7.7 Huber and Wiesel study of the visual field claimed that the orientation columns are 50μm width and represent a shift in angle of 10^{o}. A complete set of columns subserving the 180^{o} is called hypercolumn. There are two types of ocular dominance columns (L,R) and two kind of orientation columns. For purposes of illustration, Hubel and Wiesel depicted the two sets of columns are orthogonal but there is not evidence of such relationship between them [19].

property. Such a point of view has been today proved to be, at best misleading and at worst fallacious. In actual fact, these columns lack borders. This fact can be empirically noticed, orientation is not quantal, the 22 orientation shifts found by Huber and Wiesel are arbitrary, but varies smoothly across most of the cortex [284].

It is worthwhile repeating that the columns initially represented basic modular units of the cortex. In 1988, Rakic found what He coined "ontogenic columns" serving the basis for the radial unit hypothesis of cortical formation. [285]. As Mountcastle's original columns were larger than Rakic ones, Mountcastle coined a new term, minicolumn which was baptized as the new basic modular unit, dispensing the cortical column of that honor.

To sum up, minicolumn is the basic modular unit and the neurons within it encode similar features. The superposition of a group of minicolumns form a columnar module whose neurons share common functional properties. This now leads me to raise some concerns; first, how many minicolumns are typically in a column; and second, is the minicolumn a functional unit of the cortex, or it is instead the column to play that role? There is no common agreement on the first point, according to [273], "perhaps 50-80 minicolumns are aggregated into a macrocolumn" [3]. The answer to the second point is no. The functional role assigned to columns is not justified. The rationale of this is twofold. First, because single neurons have multidimensional receptive fields, this means that different properties that map onto a given cortical surface can never be simply detached into separate "modules". Secondly, it has been impossible to find the canonical circuit isomorphic with the cortical column. This does not disprove the existence of canonical or local circuits in the brain, rather it draws attention to the fact that it can not be as simple as clipping out a cylinder of cortical tissue in the hope that it contains such circuit. Systems biologists are committed to this task and network motifs such as negative auto-regulation and feed-forward loops have been identified in microorganisms [286]. In order to understand the success of systems biology in dealing with the design principles of biological systems, we need to acknowledge that, on one hand, from the technological improvements in high throughput sequencing techniques, a huge collection of data arose, and on the other hand, the use of mathematical terms has provided a whole new language able to deal in a more satisfactory way with issues, like molecular interactions in biochemical reactions, that were exclusively tackled with plain words. The transition from a purely descriptive biology to an axiomatic one, has just started and seemingly will keep going towards a direction of mathematisation. The strategy I am proposing to be pursued here, strives to track down the underlying principles, written in formal-mathematical terms, that explain and control the organisation of the cortex and other major areas in the brain.

To conclude, the search for meaningful patterns can not rely on anatomical structures seen as modular units, rather it is only through the understanding of organisational principles and their formulation in mathematical terms how the patterns and their dynamics, can be satisfactorily be apprehended. Mountcastle, the champion of the columnar organisation of the cortex, seems to be aware of this, since he writes[287]:

> These cortical sensory areas do not always fit classically-defined cytoarchitectural areas. They are defined by their patterns of connectivity and the functional properties of their neurons, as well as by their cyto- and myeloarchitecture.

---

[3] The cortical column is used in so many ways that a remark is necessary here. Cortical module, cortical column, Hypercolumn and Macrocolumn are synonymous, meaning a group of neurons which have nearly identical receptive fields.

7.7 The dynamics of the meaningful patterns. Neurodynamics

Dynamics has different connotations depending on the context is being used. While in classical and molecular physics, dynamics refers to the motion of bodies and the impressed forces that caused it, in neurosciences, dynamics or system dynamics is concerned with the behavior of complex systems, i.e. the brain.

In system dynamics, a dynamical system is a mathematical formalisation consisting of a 3-tuple, (T, M, Φ), where T is the time, M is the space state and Φ is the evolution of the system, $\Phi : T \times M \rightarrow M$. Thus, Φ associates each point of the space state $m \in M$, with an unique image $m' \in M$, for every instant $t \in T$.

The concept of a dynamic pattern is rooted in the famous aphorism, wrongly attributed to Heraclitus, "Panta rhei" or everything flows, which entails the idea that systems are made of flow patterns whose components are self-maintaining features of the flows.

In the natural sciences, e.g. Physics, Biology, Chemistry, dynamic patterns refer to those patterns that are created and evolved by a process of dissipation of energy. For Prigogine [288], dynamic patterns are dissipative structures which feed on energy to dissipate as heat. Notorious examples of dynamic systems are Lotka-Volterra study in the change of the size of the population of fish in the Adriatic sea, Turing's Theory of morphogenesis [289] or Hodgin-Huxley model that explains how action potentials in neurons are initiated and propagated. Brain functioning is based on organisation of large number of neurons into coherent dynamic patterns. In neurosciences, neurodynamics is based on spatiotemporal patterns in brain activity. The patterns are indeed dynamic structures [20].

The dynamic pattern theory is indeed a mathematical tool useful to describe the *qualitative change* in the creation of an emergent property in the system that is not present in any of the components that form the system. Changes take place by rapid and repeated jumps from each pattern to the next defining a flow.

Prior to trying to work out the dynamic model of a system, for example the brain, the state variables must be declared. The level at which we are accessing to these variables, microscopic, mesoscopic . . . necessarily imposes strong restrictions in the model. As noted above, the brain operates at different levels of organisation, each with different components and its particular epistemology. As a result, there is a gap between levels. While the state variables and their mappings at microscopic level are well known, those at masoscopic and macroscopic level are not. The mapping among the state variables of the micro and the macro levels, that explains accurately how local actions of single neurons give rise to the global brain states, is still missing.

7.7.1 The I/O Newtonian approach in brain dynamics modeling

The success in the obtention of models of the dynamics of single neurons is owing to the application of the classical I/O approach. The state of a neuron is defined in terms of variables or quantities that can be measured and controlled e.g. membrane potential, firing rate. The way in which the neurons interact with each other is calculated by a rule, typically a differential equation, able to express the relationship between the output and the input variables. In order to be solved, the experimenter sets the initial conditions of the equations. As the inputs of the equations are controlled by the experimenter, we can analytically obtain the I/O relations of the micro state variables included in the equations. For example, in the Hodgkin-Huxley equation the action potential in the giant squid axon is obtained through the ionic current and the capacitances.

$$\dot{V_m} = \frac{1}{-C_m}(\sum_i I_i) \tag{7.4}$$

Thus, the derivative of the potential across the membrane, $\dot{V_m}$, is directly proportional to the sum of the currents in the circuit, $\sum_i I_i$, and inversely proportional to the capacitance, C_m.

The microscopic studies of the brain have benefited from the fact that a neuron's output is simply the action potential it sends down its axon Herein lies the linear causality assumed at the microscopic level of the neurons which is inherited from classical Newtonian Physics. Similarly to this approach, the Newton-Euler equations used in industrial robotics for calculating the movement of the manipulator arm created by the applied torque, is similar to the linear ODE cascaded equations used in the model (K0 set) described by neurobiologists that represents the dynamics of a local non-interactive neurons. The KO and the rest of K-sets are described in 7.8

This approach operates appropriately at the microscopic level of sensory driven activity and it has made it possible to define the topographic pathways of sensory and motor areas in the cortex. The neural activity in those areas, is imposed by the stimulus input. Thus, the stimulus and the pattern that expresses synchronous actions between neurons, are isomorphic.

However, not all the stimuli are as easy as an electrical impulse directly applied to a sensory receptor. The perception of a landscape or the remembering of a past event, needs of a different approach than the topographic mapping of stimulus-response in single neurons. The I/O approach in neuroscience assisted in the obtention of the mapping of topographic pathways is useless when dealing with the creation and deletion of perceptions and memories.

To sum up, the dynamics of mesoscopic and macroscopic levels are built upon patterns that are not identical to external stimuli, and neither can be inferred from them, this is also the case in microscopic level. Within the mesoscopic and macroscopic levels, patterns capture the meaning of the stimulus in a given context. Therefore, a different approach is required which may be able to cope with activity patterns of large neural populations, rather than with single interacting particles.

7.7.2 Neurodynamics

Neurodynamics is mainly concerned with the study of spatiotemporal brain activity patterns. The dynamics here is about the changes that the brains have undergone by the jumps from one pattern to another. Hence, the main objective in neurodynamics is to extract patterns and then model the state transitions or jumps between these patterns. The link between microscopic sensory data activity and mesoscopic constructs may be explained using the state transition of global attractors. Chaos theory can help us to understand how mesoscopic *order* is created from microscopic *disorder*. For example, the mesoscopic activity on the olfactory system is chaotic.

7.7.3 EEG/ECoG recording and theoretical hypothesis

As it was shown before, the term pattern may have different meanings. In [261], Watts and Strogatz use pattern as equivalent to topological connectivity. The 302 neurons and 7000 synapses of the *C.elegans* are mapped into a graph. For the *C.elegans*, the pattern of neural activity is merely the topology of the graph. The reduced number of neurons and their interactions, suffices to explain the neural basis for cognition and behavior of this worm. In these kind of organisms, there is only microscopic level to deal with, the mesoscopic, and macroscopic levels do not exist.

In order to tackle the mesoscopic and macroscopic properties of the neural system, the depiction of a graph with the structural connectivity is not enough. Techniques like EEG or ECoG that capture the patterns of connectivity at the macroscopic level of thousands or even millions or neurons are needed.

Before one starts searching for meaningful patterns in the brain, two issues must be promptly addressed however. One is to determine the technique that is going to be used in the experimental analysis. The other, is in the theoretical ground, as we can easily get lost in the data. Therefore, sound hypothesis based on empirical work must be made. Certainly, the empirical and theoretical flows of information are bidirectional.

As it was shown in section 7.4.1, EEG/ECoG recordings provide data more closer to brain dynamics than the metabolic data obtained with a fMRI.The main difference between EEG and ECoG is that EEG is recorded from the scalp while ECoG recording is taken directly from the cortical surface. Both EEG and ECoG depict brain waves that represent synchronized activity over network of neurons. The waves are indeed the summation of synchronous activities of thousands or even millions of neurons. Thus EEG/ECoG waves are macroscopic/mesoscopic manifestations of the collective action of a myriad of neurons and their action potentials. It ought to be noted here, that brain science at the macroscopic level, is not strictly causal in the sense that the macroscopic waves are correlated and not causally connected with the microscopic action potentials.

The theoretical claim is as follows: neocortical connectivity and dynamics are scale free. The empirical basis for this statement was shown in section 7.6. Additional empirical support for this claim can be found in [18]. The scale free dynamics could explain how brains with huge differences in size e.g. the brain of a whale is 10^4 times bigger than the brain of a mouse, can still operate at the same time scales. Moreover, it provides a different interpretation of the hot spots shown in fMRI, rather than the physical localisation of cognitive functions, which can be understood as the hubs in scale free connectivity network.

> Despite the differences among nerve cells, the basic mechanisms of electrical signaling are surprisingly similar. This simplicity is fortunate for those who study the brain. By understanding the molecular mechanisms that produce signaling in one kind of nerve cell, we are well on the way to understanding these mechanisms in many other nerve cells [223].

The same idiosyncrasy can be found in the Theory of Neuronal Group Selection (TNGS) [15] or The Global Neuronal Workspace (GNW) [290], these theories capture macroscopic patterns, exploiting self-organising features of neural populations across the corticothalamic loop [291].

However, further investigations by neuroanatomists and neurophysiologists are still needed to precise the correctness of this hypothesis that for the time being seems well founded. There is a question that needs to be addressed, however. How do brains perceive the meaning of stimulus. The answer to the question is: through the construction and binding of percepts. The percepts are the *stuff* that makes possible perception, in other words, a percept is the neural correlation of a perception. This is a theoretical claim that requires an empirical counterpart. Placing ECoG electrodes in the cortical surface, it is possible to record brain waves, that reflect the electric activity beneath the electrodes. Clearly, the brain waves are not the neural activity per se, rather they are epiphenomenal to that activity.

So the brain waves alone cannot provide the answer to the above question, but they can help, because these wave packets carry the patterns or percepts, associated with the content of what is being perceived. Here, patterns are used as equivalent to the meaningful patterns previously described in section 7.6.2.

At this point it is very important to realise that the term pattern is an abstraction, a mathematical object, and not any particular physical instance. In order to be able to apply the concept of pattern, we need to instantiate it into real structures in the brain. Now it can be shown that the meaningful pattern specified in section 7.6.2 and the spatial patterns of ECoG waves that are going to be explored next, are two sides of the same coin. Thus, for example, in a *C.elegans*, the meaningful pattern is associated with the topological connectivity perceived with electron microscopy. And in a more complex nervous system e.g. the human cortex, the meaningful patterns is related to the wave patterns, more precisely the spatial amplitude modulation of ECoG waves. A meaningful pattern occurring in a perceptual process is a percept.

7.7.3.1 The AM patterns in EEG/ECoG recording

In this subheading I will show the way in which a percept or a meaningful pattern can be identified by the spatial pattern found in the ECoG wave packets.

It was already stated that stimulus and percept are not identical. The information processed by the sensory receptors corresponds to some features of the stimulus, but it is the percept that conveys the content of the stimulus. When a conditioned stimulus is presented to an animal, there is a latency between that instant and the moment in which the wave pattern appears. The temporal difference is due to the time required for the formation of a percept of meaningful pattern. For example when a cat has learned that a particular stimulus, for example, a specific odor that precedes the ingestion of food, a spatial pattern that captures the meaning of the situation is created.

In order to get the spatial pattern a ECoG weave burst we need to sample the ECoG signal at different points, for example, recording from a matrix of 8x8 electrodes placed in the cortical surface. Now, observing the characteristics of the waves in all the 64 points being measured, the waveform is similar everywhere but the amplitudes differ across the surface. The spatial patterns of amplitude modulation, AM patterns, are the best candidates to express the meaning of a conditioned stimulus. Freeman[92] found these patterns in the olfactory system and it is coherent to expect, based on the fact that the olfactory system is the oldest and its structure probably pioneered the development of other areas, to find them in the other sensory cortices. Thus, it may be a critical mechanism to understand perception.

There is evidence for the existence of locally coherent domains in the cortex. This means that there is a common waveform, reflecting high frequency oscillatory activity, across spatial extent. The amplitude and phase of the wave are both spatially modulated [292]. The spatial pattern of AM has been related as a categorical perception of conditioned stimuli [293], [294]. Spatial integration is inherent in the activity of a population of neurons and

EEG/ECoG is so far the best placed method to observe this phenomenon. The spatial "code" for olfaction is spatial, it has been proved also at the molecular level [295]

It is important to realise that AM spatial patterns do not encode the features of the stimuli, rather they categorise the stimuli according to the context in which the animal is perceiving the stimulus and its past experience. Indeed, two different AM patterns arise from the same stimulus, this is because AM patterns are generated endogenously based on the internal state of the animal e.g. motivation, memory etc. and therefore the AM pattern cannpt be obtained directly from the stimulus.

$\text{stimulus} \circ \text{AM pattern}^{-1} \neq id$

The pattern of phase modulation, PM patterns, does not seem to have a correlation with perception as AM patterns have, but can help to delineate the borders of the coherent domain within the AM patterns that encode the meaning of the stimuli. For example, in the olfactory bulb, the AM patterns are longer than the cortical columns; so the PM patterns can be seen as an alternative to the strict boundaries imposed by the anatomical areas like the cortical columns. The PM patterns have a conic form, which suggests that the AM patterns do not occur simultaneously everywhere in the domain, rather the epigenesis is the apex of the cone[4].

A typical AM pattern involves from 10^3 to 10^7 neurons [92]. Thus, an AM pattern is a manifestation, at the mesoscopic level, of the process of integration of the information arriving from different sensory modules e.g. auditory, visual, somatosensory etc. Irrelevant details, like the firing rate of a particular neuron, are removed. The AM pattern, expresses the percept, so the percept can be reconstructed through the pattern. To sum up, the wave is the carrier of what is being perceived, and the AM pattern of the wave expresses the content of what is being perceived, namely the percept.

In [296], it is hypothesised that AM patterns plays an even more important role in cognition than categorical perception, AM patterns would represent the state of the area in which they evolve, in relation with the other areas with which they are interacting. Thus, each area's spatial AM pattern would exert a mutual constraint and as a consequence of the interaction, a consensual state would emerge as the global neurocognitive state, representing the assessment of organismic state. In this view, consciousness is an emergent property of the interaction of cortical areas. Disruptions of this process of synchronisation between AM patterns could result in pathological issues like schizophrenia, Alzheimer's disease [297] or autism [298].

---

[4] The simulitude between the PM cones and the colimit in Category Theory can not be merely coincidental

7.8 The K-Sets hierarchy model

Freeman K-sets theory is a nested hierarchy of models that describe the dynamic behavior of neural masses at the mesoscopic level which comprises 10^4 neurons with 10^8 synapses.

> The definitive attribute of the neural mass is the set of functional connections which are the channels through which neural activity is transmitted[20] .

Thus, the K-sets are hierarchically organised. K0, KI and KII are electrophysiological models that fall under the domain of the experimentalist, namely, the anatomist and the physiologist. KIII is the lowest model in the hierarchy that copes with behavioral aspects which are mainly located in the limbic system -e.g: amygdala, hippocampus. On that account, KIII is within the experimental domain of the psychology. The dynamics of olfactory system are now fairly clear and can be modeled by a KIII set in the K-sets theory. KIV models cognitive abilities such as navigation. The hippocampus and its navigational and perceptual features have been modeled using KIV sets [299]. KV is proposed to model scale-free dynamics in mammalian cognition and as far as I know, is still underdeveloped.

Freeman in [20], provides a comprehensive and educational exposition of the theory of K-sets until the KIII level. The KIII is proposed as a model for key features in the olfactory system like learning and pattern classification. K-sets takes into account both the anatomical and functional aspects in the brain areas. The topology of connections is modeled using flow diagrams likely those used in Control Theory. The dynamics, that is to say, the variation of the state of the neurons, is expressed trough ordinary differential equations (ODE). The parameters of the equations are chosen in accordance with the topology of the network dynamics that is being modeled.

Owing to the mathematical nature of the K-sets, it is possible to measure, in quantitative terms, how good the dynamic model is, by comparing the state of the system predicted in the ODEs with the responses experimentally observed. K-Sets hierarchy designates a topological hierarchy for neural sets. Freeman identifies eight basic topologies that represent six possible patterns of connections among neurons. Convergence is the axon of two or more neurons joining on a different neuron, Divergence means the axon of one neuron ends in two or more neurons, Linear when the output is transmitted in succession, Parallel when neurons receive from a common source and propagate to a common target of neurons, Autofeedback is a neuron with feedback onto itself, Excitatory positive feedback, Inhibitory positive feedback and Negative feedback.

Only one neuron may have circa 10^5 synapses, therefore it seems not possible or worthy to describe all its possible connections. Besides, a neuron is not an uniform cell, there are cells with no axons, other with no dendrites . . .

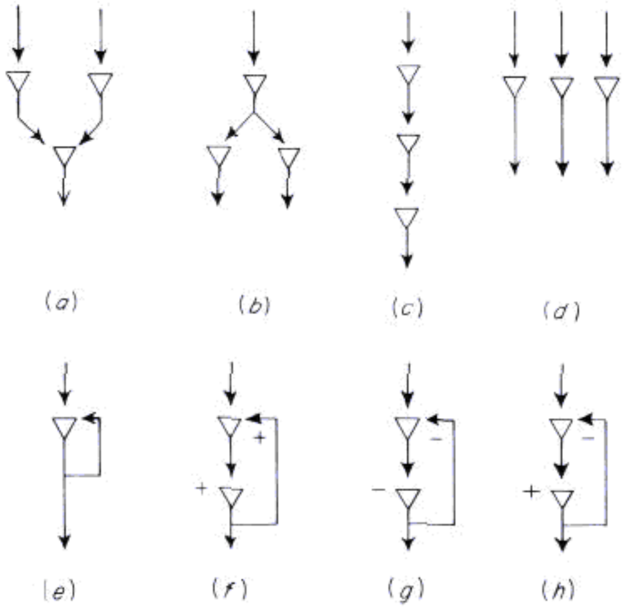

Fig. 7.8 Topological patterns of connection described in [20] (a) Convergence. (b) Divergence. (c) Serial transmission. (d) Parallel transmission. (e) Autofeedback. (f) Excitatory positive feedback. (g) Inhibitory positive feedback. (h) Negative feedback.

The K-sets transcend the anatomical and electrophysiological issues and put the focus on modelling the domain of cooperative activities. The main factor in the assignment of one or other level in the K-hierarchy seen above, is the number and scope of the feedback connections.

7.8.1 K0

K0 is a collection of non interactive neurons, that is to say, the neurons in a K0 set have no functional interconnections. Typically, a K0 set may contain $10^3 \sim 10^8$ neurons. The problem of assigning a value to each of these *points*, which is the normal procedure in the physical sciences when dealing with systems that are simple enough, is that it may result a computationally intractable problem.

Thus, it is assumed that the activity in a K0 set is continuously distributed in both time and space. The closer two neurons are, the more similar their state of activity. A K0 set can be modeled as a single average neuron whose activity function is given by a activity density function $\phi(x, y, t)$ Thus a K0 set has only two possible topologies, linear or parallel, sharing a common source of input (for example, a common stimulus like odor, light intensity or pressure on skin receptors) and common sign of output (+ for excitatory and - for inhibitory).

Two K0 sets connected without feedback channels form another K0 set. Inside a K0 the neurons are whether excitatory, forming KO_e or inhibitory, KO_i. K0 is the basic unit or primitive in the K-sets theory, and models the open loop dynamics of a neural population.

7.8.2 KI

KI, similarly to K0, has a common source of input and common sign of output, but conversely to K0 there is interaction within the set. There are two types of KI sets, KI_e, made up of mutually excitatory neurons and KI_i of inhibitory neurons. For example, a $K0_e$ set with interactive neurons is KI_e, likewise an interactive $K0_i$ set is a KI_i.

Owing to the interactions between the neurons, the active state of KI can not be represented as an average neuron, as seen with K0 sets. The ODEs for KI have an additional parameter, the feedback gain which describes the interaction within the system. The feedback gain is a collective or emergent property in the sense, that can not be measured at the level of the synapse between two neurons. A number of K0 and KI sets exclusively connected by forward mechanisms forms another KI set.

7.8.3 KII

KII models the feedback interaction of KI sets. There are four types of KII models. KII_{ee} if both KI are excitatory, KII_{ie} and KII_{ei} if one component is KI_i and the other KI_i and KII_{ii} when both are KI_i. Thus, in a KII set each excitatory neuron can interact with either excitatory or inhibitory neurons, the same goes for the inhibitory neurons.

7.8.4 KIII

KIII is a set of two or more KII units connected by feedforward and delay feedback connections. In [300], [301] a KIII based model of the olfactory system is applied to classify linearly non separable patterns. The advantage of KIII pattern classifier, over traditional neural network is the reduced training period that is required in the former. The drawback is that, of course, the ODEs have to be solved. At this level of complexity an analytic solution for the equations is unlikely, so the time required for the computational algorithm that gives a numerical solution can be a critic factor. However stochastic models like neuropercolation can help in this limitation.

The figure 7.9 shows the modelisation of the hippocampus using a KIII set. Each K set emphasizes a particular feature, for KIII is that of the creation of chaotic background activity [21].

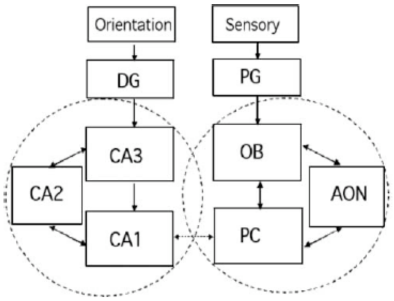

Fig. 7.9 The figue is from [21], and depicts the relationship between the hippocampal formation and the sensory cortex KIII sets. In the hippocampal side, DG is dentate gyrus, CA1-3 is cornu ammonis and on the olfactory cortical side, PG is periglomerular, OB olfactory bulb, AON anterior olfactory nucleus and PC prepyriform cortex. Chaotic behavior in KIII is the result of the competition between 3 KII components of the hippocampus, CA1, CA2 and CA3. The dentate gyrus (DG) is modelled as K-I set .

7.8.5 More complex models: KIV and KV

In [22], [302] the sensory cortex and the hippocampal formation are modeled as KIII sets and integrated in the KIV level. The KIV model is a bio inspired architecture of modeling sensory-perceptual-intentional action cycle, using nonlinear dynamical principles. KIV which is schematically depicted in figure 7.10, has been successfully implemented in mobile robots operating in virtual and real environments in classification, basic learning and navigation tasks [303], [304].

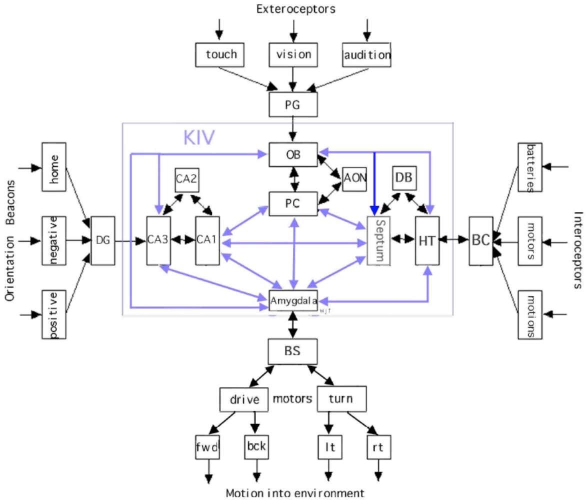

Fig. 7.10 The KIV consists of three KIII sets, which model the cortical and hippocampal areas, and the mid-line forebrain, respectively. [22]

7.8.6 Categorizing K-Sets

The main objective of this section is to show that any mathematical structure forms a category if we define a homomorphism for such structure. More precisely, the Category of Sets **Set** is explored and a formulation is provided for its use in the Freeman's K-Sets models. A categorisation of K-sets is sketched.

The hierarchy of models of K-sets epitomizes a very important statement that should never be overlooked. The levels of organisation in the brain, as was pointed out previously, are not necessarily two nor three, but many more, describing different levels of structural and functional operations. Therefore, a previously unique class of network model can never capture such diversity, instead what we need is a framework where scale free models can be put in place.

The category of sets, **Set**, can be employed to categorise the K0, KI, KII, KIII and KIV sets, as is shown bellow. In doing so, we do not just reformulate sets into more abstract objects, namely categories. Indeed, the final objective is to provide the mechanisms that enable us to formally describe the interactions between models at different levels. Figure 7.11 shows shematically the K-sets hierarchy.

Category theory is a generalisation of set theory, a set is defined in terms of the objects -their properties-, in category theory, a set is defined in a more general -categorical- way in terms of the morphisms that map to or from them.

The K-sets model theory follows the drawer-like paradigms: different models explain phenomena at different levels of description. As a consequence of this clear cut distinction between levels, a global view of the whole system encompassing all the levels is missing.

K0 are non interactive single neurons that can be modeled by ODE with second order non-linear transfer functions. Thus, the category K0, $\mathbf{C(K0)}$ can be put in categorical terms as the category whose objects are single neurons N, and the arrows are endomorphisms or arrows with the same source and target $a : N \to N$. There are two types of neurons in N, excitatory neurons N_e and inhibitory neurons N_i. Of course no arrow between two N_e or two N_i exists. The category $\mathbf{C(K0)}$ is discrete because the only morphisms are the identity morphisms. Any given set S is a discrete category if the only morphism defined is the identity.

KI sets represent the coupling of either two excitatory neurons or two inhibitory neurons in $\mathbf{C(K0)}$. Thus, the category K1 $\mathbf{C(K1)}$ is the category whose objects are the same objects of $\mathbf{C(K0)}$, and two kind of arrows $s_e : KO \to K0$ and $s_i : K0 \to K0$.

KII sets have either KOe and KOi nodes. Put in categorical terms, KII is the functor between two categories in $K0$, $K0e$ and $K0i$. Thus $F : K0e \to K0i$

KIII can be categorised as the natural transformation $N : F \to F'$ between the two functors $F : KOe \to KOe$ and $F' : KOi \to KOi$, as previously

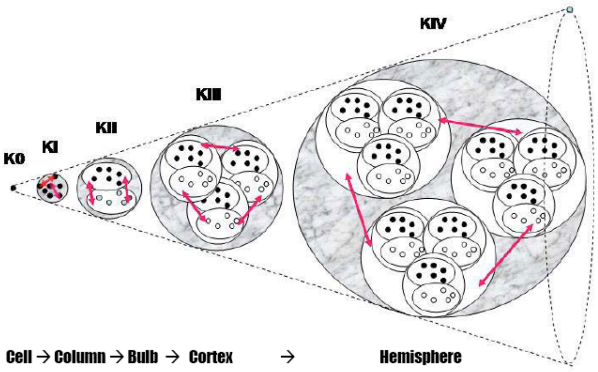

Fig. 7.11 K-Sets hierarchy designates a topological hierarchy for neural sets. In [23].

defined in $\mathbf{C(KII)}$. The dynamics generated by this kind of structure is non-zero fixed point or limit cycle behavior.

In order to categorise the top set in the hierarchy KV, we do not need anything more abstract than natural transformation. In order to categorise KV we must exploit the universal constructs seen in section 3.2.2.1. Notably, the colimit, as the object that acts as the "glue" of patterns of neural connectivity. This idea deserves to be carefully explored in future works.

Chapter 8

A theory of hippocampus structure and function based on Category Theory

8.1 The nature of spatial representation

Prior to get into the biological basis of space and the major role of the hippocampus in spatial representation, I will give a brief keynotes on the different conceptions of space in modern science (since Galileo). Those indifferent to philosophical and theoretical debates may skip this heading and pass directly to 8.2.

In physics, the main concern is with the properties of the physical space, which is mean to be that inhabited by things, belonging to the external world and independent of the minds that perceive them. One of the oldest disquisitions in natural philosophy[1] is the absoluteness or relativeness of the physical space. While there is a common agreement in that the psychological space necessarily needs to match aspects of the physical external universe, the disagreements begin when one asks about the nature of space.

Newton, built his laws around the metaphysical concept of absolute space as a "thing" with no relation to anything external to it and inaccessible to the senses. In Newton's dynamics, the concept of absolute motion leads naturally to that absolute space, the last must exist because the former is real. For example, the centrifugal force creates an centrifugal motion in a body which is an absolute motion that can only be understood assuming that it occur within an absolute space. Even though we can not experiment the absolute space because is inaccessible to our mind, we can measure the effects of the bodies embedded in such a space.

Leibniz is the champion of the notion of relative space which is, of course, neatly at odds with Newton absolute space. For Leibniz, space cannot exist in the absence of objects, not any object, but monads which a very special object conceived by Leibniz as are a sort of metaphysical object because have no mass nor extension. The main characteristic of monads is their continuous changing internal state. In Leibniz, the space arises form the relationship

---

[1] the actual physics is the modern natural philosophy

between monads. Thus the space is the whole collection of places. We say that i has the same place as j had before when i has the same relations than other thing j had.

In Kant, the conception of space lays on the one hand on Newton absolute (physical) space and on the other on Euler hypothesis of space as purely psychological but not derived from sensations of the external world. Kant postulated his famed “space is synthetic a priori”. The space is pure intuition and is given prior to all other perceptions, and contains in itself the principles that relate the perceived objects[305]. Now is time to turn to those principles.

In Euclidean geometry we are concerned with three type of objects: points, lines and planes which are not defined and neither can be discovered by repetitive acts, rather they are intuitively given. In this rubric our minds are not only hardwired with these three notions also the basic relations between them are also given[306].

After the discoveries of new geometries in the XIX th century by Riemann, Lobachetsky and others, the Kantian assumption that Euclidean geometry necessarily reflected the structure of the physical world was no longer tenable.

However, it might be said that the existence of Rienmann geometry does not invalid the idea that the spatial framework is innately given, a synthetic a priory, in Kant’s words, rather it is the metric of the framework that must be revised.

Poincaré’s model of space states that the geometrical space is a matter of convenience, we choose one geometry or other depending on our needs. In Poincaré the space is representative, which means the space of introspection that lacks the neutral qualities of the Kantian’s space: infinite, continuous, homogenous and isotropic. In this conception, the space is more egocentric and is related to the perceptual experience given by the senses .e.g: tactile space, motor space . . . Hull used the Poincare’s ideas in his study of spatial behavior in rats. He coined the “habit-family hierarchy” which is an application of the mathematical group theory used by Poincaré, displacement group, the group of the different movement patterns that can bring the system .e,g: a rat, to a previous situation[307]. Thus spatial concepts would arise in the mind through the equivalent movements (physical or thought) for the same external change.

Gestalt theory borrowed terms from relativity physics, notably the concept of field which is an overload term. The neural processes conform a psychological field, the individual parts of the field lack phenomenological meaning, rather the entire pattern or field is isomorphic to a percept. Gestalt theory can not provide a basis for objective spatial representation and is unable to cope with learning [308].

8.2 The hippocampus as a representational device

How does the mind represent physical space? This is a question that has kept philosophers busy for centuries. In 1975, the philosophical discussions about space representation acquired a extremely powerful and fresh insight when O'Keefe and Nadel, discovered the place cells in the hippocampus of the rat.

The experimental study of spatial representation has since then exploded. The 70's was the decade of the place cells, neurons that discharge when the rat is in a particular position. In the 80's head direction cells, neurons that discharge significantly whenever the rat's head changes direction, acquired the attention of scholars. Since 2005 we have been in the grid cell "era".

These discoveries are of major importance in different research fields. Indeed the theory of the cognitive map [308] is rooted in the discovery of place cells in the hippocampus. One derivative of this theory is the map-based navigation capability, that some animals have, and that engineers have been able to replicate in robots[309].

8.2.1 The cognitve map hypothesis

The concept of cognitive or mental map arises naturally from the place cells. The existence of these cells seem to somehow validate the Kantian conception of the space as an a priori intuition independent of the experience. The cognitive or mental map can be seen as the biological realisation of the mental space that in Kant epistemology was an a priori intuition.

A remarkably usual argument against Kant a priories -space, time- is that there are other geometries than Euclidean geometry like for example Rienmann. Kant, was a son of his time and could not possibly know about the future theoretical developments in mathematics and physics, of course Kant could never guess what Rienmann mind would fabricate!. The problem with this criticism, as I see it, is that there is the expectation that the theory postulated by Euclides -axioms and theories- must be implemented in the brain to validate Kant's view. It goes without saying that axioms can hardly be hardwired in the brain, but, and I think this is where Kant is correct, the concept of distance or metric actually is embedded in our neuronal tissue as the grid cells show.

The concept of mental map is presented as the tool that allows animals to explore the environment, search for food or find their way back to their nests. An extreme case is to find the nest after long migrations, like for example in the artic tern, which is seabird that makes a round trip of 70,900 km between the artic and the antartic poles every year.

It might be mentioned that the ability to build cognitive maps that represent the world was known before the discovery pf place cells; notably E.C. Tolman, working with rats, proved that this animal construct field maps of the environment[310]. More precisely, Tolman demonstrated that rats have place learning ability in more or less complex mazes.

In the formation of the cognitive maps, the landmarks or external cues, play a major role. However, it might be remarked that place cells do not merely reflect sensory stimuli, like for example visual cues. It is known that for a rat in a familiar environment, the landmarks can be removed and the place cells in the hippocampus of the rat will continue firing in the same firing fields. It is evident that for an animal, to navigate in a efficient way obtaining food, hiding from potential enemies or finding the nest, it is necessary more than having a mental representation of individual places. These places must be included into a metric system that allows the animal to for example calculate the distances between two points or know the direction taken.

The cognitive map was challenged by a more complex conceptualisation of the hippocampus as the locus of path integrator. Information like velocity, direction is integrated into a navigational position.

The omnidirectionality property which at the root of the idea of the allocentric map is not valid when the rat is an unidirectional space -e.g: moving in a straight alley. In this case, place cells are not omnidirectional but direction-dependent, put simply, different neurons are active for the same place, depending on the direction taken.

This finding does not discredit the idea of the cognitive map, rather it forces us to question about how the maps are effectively created in the brain. The map can not be created exclusively based on visual cues, other inputs must be considered like for example speed or direction. The cognitive map mechanism needs to be thought in a wider context, to be incorporated into a more dynamic and systemic perspective. A map is a static object that does not match in a changing world.

Besides, there is the problem of the overlapping maps stored in the very same population of place cells. If we think the problem of navigation and spatial representation in terms of dead reckoning or path integration, the problems disappears. The animal needs are other than visual cues or landmarks, self-motion are necessary too in order to have a metric that allow the animal to calculate distances and optimal routes.

The debate about whether the brain generates a map like or not, seems to have shifted in favour of those who back the cognitive map theory. Indeed the discovery of place cells, head cells and recently, grid cells suggest so. Yet the underlying nature of the cognitive map remains elusive. Is the representation purely metrical or is topological? or are the maps constructed in the hippocampus without paying attention to the features of the environment i.e:metrical maps, or do they reflect the relationships between the environmental features i.e:topological maps?

The role of the hippocampus is to associate internal and external coordinates systems and to accommodate cue conflict situations (reinstantiate the context when there is a mismatch between internal and external relationships). Rather than debating whether the hippocampus is the depositary of the declarative memory or the index access of a collection of maps, it may be more productive to ask What is the role of hippocampus in navigation and memory? With this mind, in “The hippocampal debate: Are we asking the right questions?” [311], Redish suggests that there are multiple memory systems in the brain and multiple navigation systems.

8.2.2 The hippocampus: anatomy and connectivity

In order to understand the function, one has to deal with the structure first. This motto is elaborated here, and both the anatomy and the architecture of the hippocampus are explored.

The hippocampus is an allocortical structure in the medial temporal lobe. The most studied cell in the nervous system is the pyramidal cell of the hippocampus. One of the main reasons why it has been invested so much time and effort in investigating the anatomical and functional aspects aspects of the hippocampus, is its simple architecture of three layers, simple if we compare it with the six-layered cortical structure.

The modern name to refer to the allocortical structures of the brain, is limbic system. The limbic system includes the amygdala, the hippocampus, the entorhinal cortex and the hypothalamus. It is commonly assumed that the limbic system plays a major role in the emotional aspects of the animal, so emotions and feelings like love, hate, fear ... would be “located” in this allocortical areas of the brain.

It is more correct to talk about hippocampal formation than hippocampus. The hippocampal formation, is a group of brain areas consisting of dentate gyrus, hippocampus, subiculum, presubiculum, parasubiculum and entorhinal cortex. The connectivity in the hippocampus follows a random law, this is very different to the power law connectivity of the cortex.

The pioneer work of David Marr in computational hippocampal models, assumed the unidirectionality of connections. Anatomists today, has shown that this assumption is not completely accurate, there is bidirectionality between, for example, the dentate gyrus (DG) and the CA3. However, the flow of information is largely unidirectional. The general picture of the hippocampus connectivity is as follows: the signal input to the hippocampus from the cortex arrives via the entorhinal cortex (EC) and other parahippocampal regions. The output projects back to parahippocampal cortices primarily via the subiculum and projects to subcortical structures via the fornix. The output signal is not conveyed exclusively to the EC, additional output pathways go to other cortical areas like the prefrontal cortex, and large output goes to the limbic system .e.g: lateral septal area.

This scheme of connections can be simplified if we cluster the regions: entorhinal, perirhinal, parahippocampal cortices and the amygdala under the label medial temporal lobe (MTL). The MTL regions shape hippocampal processing and are in turn often shaped by it.

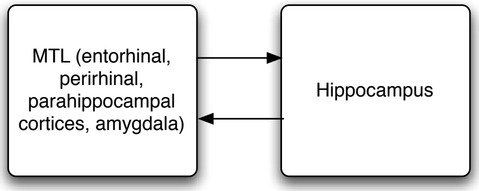

Fig. 8.1 Simplification of the usual schema of connections modules in the hippocampus. The medial temporal lobe (MTL) is influenced by and influences the hippocampus proper.

8.2.3 Place cells

Place cells are neurons located in the brain hippocampus that fire in complex burst whenever an animal, for example a rat, moves through a specific location in an environment.

The striking thing about place cells is that they code the Cartesian position, irrespective of either the direction from which the position is reached or the behavior of the rat at any precise instant. Thus, there is a direct link between the neural activity of a single cell to a Cartesian position.

How does the animal know that it is in a particular position? Apparently this could be done by computing the allocentric space, landmark or visual cues. The most important property of these cells is the omnidirectionality property and it can be observed in their conical shape (the firing rate increases when the rat approaches the location, independently of the direction is heading when it does it). Thus, the place cells are necessarily, coding explicit (no contextual) locations in the environment and not particular sensorial cues.

The region in which a place cell fires the most is called place field. Thus, there is a correspondence place field/place cell. One cell has a place field when the firing rate within the field is much higher than outside -e.g: from 20 Hz to 0.1 Hz. For a given environment, we can determine a collection of place cells whose associated place fields cover the whole environment. Nobody denies that under certain circumstances, the hippocampal pyramidal cells show place fields -location-based firing.

Notwithstanding, it is less clear what they really represent; there are those who argue that place cells can be an epiphenomenon, produced by the spatial nature of the experiments where these cells are discovered. Granted that place

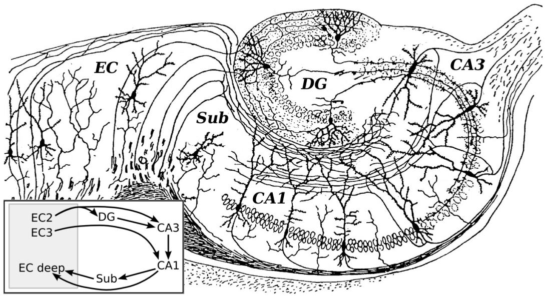

Fig. 8.2 Drawn of the hippocampus and its circuitry drawn by Ramón y Cajal. The superficial layers of the EC (layers 1,2,3) provide the most prominent input to the hippocampus, and the deep layers (layers 4,5,6) of the EC, receive the most prominent output.

cells are correlated to space, the question that arises is: Are the place cells the only neurons correlated to space? The other possible representational content of theses cells and of the assemblies they form, can serve to further question how the hippocampus contributes to spatial representation, navigation and episodic memory.

8.2.3.1 Place cells as representational entities

The interest in these cells is rooted in the fact that they are candidates to be the direct representation of the external space. A place cell, fires maximally when the animal is in a specific position or place field, so the firing rate of a cell can be used to decode the position of the animal within the environment with striking accuracy.

The existence of place cells was not accepted until Muller [312] came out with the numerical method that allow to quantify the place fields.

At this point, we need to formally define the term place field. A place field $\mathcal{F}$, for a place cell, is an open ball of radius r and center x in a normed vector space $\mathcal{V}$ such that $fr(\mathcal{F}) > k$. Being k a constant that represents the threshold or firing rate, and fr a function that returns the minimum firing rate for all the pixels (vectors) that fall into the vector space $\mathcal{F}$ [2].

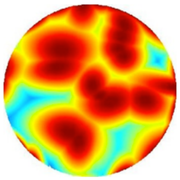

Fig. 8.3 The picture shows the 36 place fields one of each of 36 rat hippocampal CA1 place cells. Thus, as it is shown in the figure the position of the rat is contained in the firing of these cells. The place fields are cone shapes, this means that that the firing rates increases irrespective of the direction from which the rat arrives.

8.2.4 Grid cells

Grid cells, likewise place cells, are place-modulated neurons. The firing location of a grid cell is multiple, contrary to the place cells which are mono field. The multiple firing location of a grid cell is indeed a grid with a most striking property, it is an array of equilateral triangles.

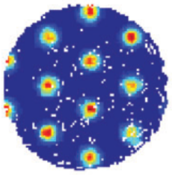

Fig. 8.4 Grid map of a rat running on a linear track after 10 min [24].

---

[2] In the initial notation of Muller, the function fr uses 6 colors to represent the firing rate, the darker the colour is, the higher is the firing rate of the cell at that pixel

It might be noted that grid cells were discovered, while researchers investigated whether place cells activity was endogenous to the hippocampus.

The hypothesis was that CA3 and DG are the intra-hippocampal inputs to CA1, which is the area in the hippocampus where one can find most of the place fields. This idea was proven wrong, after removing CA3 and DB, the CA1 inputs, the spatial firing in CA1 persisted. So place signals did not exclusively arise within the hippocampus, spatial signal was brought to the CA1 from outside the hippocampus.

In 2004 Fyhn et al.[313] discovered a group of neurons in the medial entorhinal cortex (mEC) that shows spatial related firing. The mEC cells have sharply tuned spatial firing, much like the hippocampal place cells do, but with one difference, each of the mEC cells, has multiple firing fields rather than one as is the case in the place cells. One year later, Hafting et al.[314] discovered that the many firing fields of each neuron, generate a grid. Thus as the animal moves, the grid cells tile the environment with periodic triangles that reflect the changing position.

The grid cells have been found in the mEC, from the six layers of this cortical structure, it is in layer II where we find the highest density of this kind of cells.

The neurons in the layer II of the medial entorhinal cortex (mEC-II) are the main input of the place cells, but in contrast the entorhinal cells are activated throughout the environmental terrain, whenever the animal is at the vertex of some equilateral triangle, forming a tessellation or grid.

In short, both place cells and grid cells are neurons with spatial located firing, in other words, they have spatial representational power, allowing the animal to know its position and to navigate in an environment, for example find the way back home after eating. The difference, apart from the fact that place cells are hippocampal neurons and grid cells are in the mEC, is that whereas a place cell has a single firing location, a grid cell has multiple firing fields with a striking geometric regularity; the firing fields form periodic triangular arrays, tiling the entire environment available to the animal.

8.2.4.1 Grid field

Three parameters totally describe the grid of a cell (grid cell): *spacing* is the distance between contiguous fields, *orientation* the rotation angle of the grid referred to the reference axis, *spatial phase* is how much the grid is translated relative to an external reference point.

A grid field for a grid cell is a set of open balls $G_i : 1..n$, where for every ball G_i, $fr(G_i) > k$, that is to say, it has a significative firing rate. Thus, so far, the definition of G_i is identical to place field seen in 8.2.3.1.

Additionally, every ball G_j of a grid field, form a equilateral triangle with its two other closest balls G_i and G_k in $\cup^1 n_n \mathcal{G}$ or G to abbreviate. The grid field G is identified by the three parameters, spacing, orientation and phase that can be trivially obtained from the metric of the space defined above.

The processing of the place signal is therefore not an exclusive privilege of the hippocampus, the mEC plays a major role in the spatial representational.

Table 8.1 Place cells and Grid cells, similarities and differences

	Brain area	Type of map	Activation
Place cells	Hippocampus	static	need input from mEC
Grid cells	mEC	dynamic	active instantaneously in any novel environment

The majority of cells in mEC-II and mEC-III have grid properties, this means that most of the cortical input to the hippocampal place cells that have to do with spatial representation, come from grid cells in the mEC. Grid cells can be found just one synapse upstream of the place cells[314]. Therefore, acknowledging that grid cells and place cells are intrinsically connected, to claim that place field may be extracted from grid fields[315] deserves to be appropriately explored.

Mathematically, using Fourier's analysis, several grid fields with different spacing can combine linearly to yield a place field. In [316] is presented a computational model showing that place field can arise by the sum of 10-50 grid cells. When the spatial phase variation in the grid-cell input was higher, multiple, and irregularly spaced firing fields were formed. This idea has been very appealing in the hippocampus community, has produced a large number of models with a common tenet: place cells in the hippocampus compete to receive the summed firing pattern activity of the cortical grid cells.

The problem with these kind of models that transform grid patterns into place pattern is that they do not tell us that much about the mechanisms that underlie the spatial firing pattern of grid cells and place cells. Besides, it is debatable that a linear sum of grid cell pattern which has a metric is the correct way to model a place cell pattern which represents topologically the environment without metrical relations.

It might be remarked that the models of grid field formation deal with timing rather than with structure or connectivity, and this is because they assume that the structure is already known, a single cell, whose firing activity needs to be understood, in the best case scenario, or merely simulated in a computational models, in the majority of cases.

8.2.5 Head cells

Head direction cells (HD) are neurons with direction-based firing, that is to say, these cells fire when the head of the rat is pointing a particular direction regardless of the location of the rat or even the angle of the neck. The head direction neurons are found in the postsubiculum (between the hippocampus and the EC) but also in some limbic structures.

It has been conjectured that HD work as a compass, but the attractor magnetic field is not the earth's magnetic pole but some arbitrary reference direction. HD provide information of the direction that may be integrated with information of the speed and location and would allow the animal to navigate using path integration.

The direction system poses important problems to the cognitive map theory.

Experiments with rats in rectangular boxes showed that place cells do not remain invariant to the modification of the box. Indeed place fields changes sizes, split into two and even disappear. O'Keefe suggestion to this disconformity with place cell location selective discharges that himself discovered is that the rat calculated the size of the box by triangulation on the basis of visual cues, the heights of the wall. McNaughton, less confident on rats geometrical skills, suggests that distance is calculated based on self-motion cues. As the rat moves, it remember the physical contact with the walls and compute the vectorial distance (amount and direction of movement). The idea is that direction is obtained outside the hippocampus and delivered to the place cell system. The hippocampus would act as an integrator that calculates shortest distances between landmarks based on direction, velocity and position information. We must take into account other cues, notably the velocity, for [317] the firing rate of single cells is a function of location and also speed.

8.3 A theory of brain spatial representation based on category theory

One of the highlights of this work is that it exemplifies the theory of categories in strong non-algebraic categories. Indeed, the crucial aspect and novelty in this work needs to be met in the categorical construction of biological (non algebraic) categories.

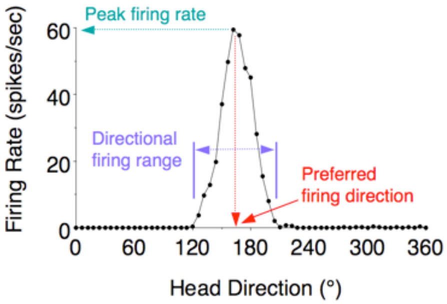

Fig. 8.5 Firing rate vs. head direction plot that shows the preference of this cell around the 180° [25].

8.3.1 *The category of neurons*

Let us define a category of neurons **CAT-Neur** as a category whose objects are either neurons or sets of neurons. **CAT-Neur** as any other category, consists of three things, *i.* a set of objects O, *ii.* a set of morphisms $Mor(A, B)$ for any two objects A,B of O, and *iii.* a rule of composition that fulfills the properties of associativity and identity .

We identify three possible categories for the category **CAT-Neur** that are useful for the development of the theory introduced in this paper.

i the category **Neur**.
ii the category **Neur+**.
iii the category **Neur***.

The category **Neur**, whose objects are neurons and the morphisms are the synaptic paths between them, with the convolution of paths as composition.

The category **Neur*** which is the category of neurons where the objects are topological spaces of neurons (N, θ) and the morphisms are continuous maps.

A function from two topological spaces $f : (N, \theta) \to (M, \upsilon)$ is continuous if $f^{-1}(B) \in \theta$ whenever $B \in \upsilon$.

And the category **Neur+**, which has as objects, metric spaces, and as morphisms, Lipschitz maps for $\lambda = 1$ that preserve distances.

Note that a Lipschitz map is always continuous but the contrary is not true. The morphisms in **Neur+** preserve distances between metric spaces which exemplify neural assemblies.

8.3.2 The category of metric fields

Now we will define the category of place fields, that is the physical locations that produce the spike firing in the grid cells and place cells.

The category **Field+** has as objects metric spaces (including hexagonal grids) and as morphisms contractions. And the category **Field*** is composed of topological spaces and continuous functions.

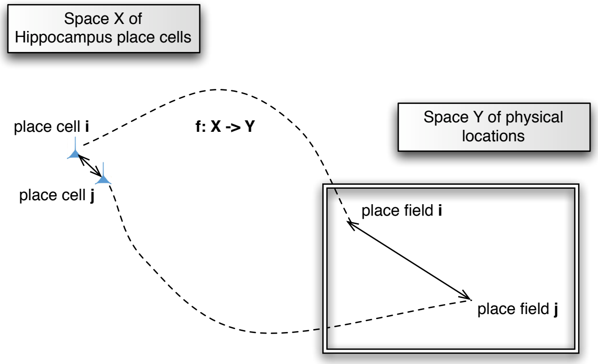

Fig. 8.6 (X, d) is a metric space where X is the set of place cells in the hippocampus and d the euclidean metric distance, (Y, e) is a metric space in the bidimensional plane with identical distance $e = d$. The mapping between the metric spaces $f : X \to Y$ preserves the distances if $e(f(x_1), f(x_2)) = d(x_1, x_2)$. f is said to be an isometry and is immediately a monomorphism (Demo: $x_1 \neq x_2, e(f(x_1), f(x_2)) = d(x_1, x_2) \neq 0 \Rightarrow f(x_1) \neq f(x_2)$). Then an isometry that is an epimorphism is an isomorphism

Empirically speaking, it makes no sense to take the space (Y, e) in figure 8.6 as a metric space, for example closed points in the space are not necessarily mapped onto closed place cells in the hippocampus.

Thus, objects in the category of neurons can not be mapped directly onto metric spaces.

The category of metric spaces is of course defined by objects and morphisms. An object is a metric space (X, d) and a morphism is a mapping between two metric spaces $(X, d) \rightarrow (X', d')$. As in any other category, the composition of morphisms must satisfy associativity and identity.

Let us define first a metric space. The structure (X, d) for X a set and the function $d : X \times X \rightarrow \mathbf{R}^+$ satisfies:

1. $d(x, y) = 0$ when $x = y$
2. $d(x, y) = d(y, x)$ and
3. $d(x, z) \leq d(x, y) + d(y, z)$

The Euclidean distance is a map $d : \mathbf{R}^n \times \mathbf{R}^n \rightarrow \mathbf{R}^+$. For $n = 2$ the distance is $d((x_1, y_1), (x_2, y_2)) = \sqrt{((x_1 - x_2)^2 + (y_1 - y_2)^2)}$.

Typically the function d is assumed to be the Euclidean distance, but of course, other distances are possible. One example of a metric that satisfies the three axions above is the "Manhattan distance" or $d : \mathbf{R}^n \times \mathbf{R}^n \rightarrow \mathbf{R}^+$, so for a two dimension space, $d((x_1, y_1), (x_2, y_2)) = |x_1 - x_2| + |y_1 - y_2|$.

Definition 8.1. A mapping $f : (X, d) \rightarrow (X', d')$ preserves distances if for all pair of points, $x_1, x_2 \in X$, it holds $d(x_1, x_2) = e(f(x_1)f(x_2))$.

Definition 8.2. A function $f : (X, d) \rightarrow (Y, e)$ between two metric spaces is continuous at $x_0 \in X$ if for all $\epsilon > 0$ there exists $\delta > 0$ such that if $d(x, x_0) < \delta$ then $e(f(x_0), f(x)) < \epsilon$

A contraction is a Lipschitz map with $\lambda < 1$, while a map between two metric spaces $f : (X, d) \rightarrow (X', e)$, is such that $d(x_1, x_2) = e(f(x_1)f(x_2))$, is a distance preserving map . Note that every Liptschitz function is continuous and as a contraction is a Lipschitz map with $\lambda < 1$, contractions are continuous [133].

Now we are able to define the category **Met** of metric spaces and Lipschitz maps that are structure preserving maps. The composition of Lipschitz maps, gf, is a Lipschitz map and the properties associativity of composition and identity $id_x : (X, d) \rightarrow (X, d)$, are trivially demonstrated.

The topological spaces are useful when we are interested in closeness and continuity rather than in distance as it is the case in metric spaces. The category of topological spaces **Top** is one that has topological spaces as objects and continuous maps as morphisms.

8.3.3 Functor between Neur and Field

At this point we wish to define the functor between the categories that have been defined.

Let us suppose that **Neur+** is a category whose objects are sets of neurons and the arrows all the functions between them, so in the case that there is only one place cell, the category **Neur+** is a set of a single element or singleton. For an object of category **C**, there is an unique functor $F : C \to 1$. Thus, there is an unique functor from the category of metric spaces and Lipschitz-distance perserving maps, **Field+**, and the category of one place cell **1**.

Functors preserve isomorphisms, so given the functor $F : C \to D$, the isomorphisms in category **C** are preserved in category **D**.

An interesting feature of functors is that they may preserve properties. For example, since functors preserve composition of morphisms $\circ$, and identities, id, they preserve every property that can be positively expressed in the language of $\circ$ and id. In particular they preserve commutativity of diagrams [318]. So given a functor $F : C \to D$, for certain objects, arrows or composition of arrows in category **C**, that have the property p, the functor F brings such property to the F-image.

Definition 8.3. Let **C** and **C′** two categories, a covariant functor F from C to C' is defined as a rule which associates for every object A in **C** an object $F(A)$ in the category **C′** plus a rule that associates for every morphism $\alpha : A \to B$ in **C** a morphism $F(\alpha) : F(A) \to F(B)$ in the category **C′**. Then F must satisfy the following two conditions:

ii.a The composition is preserved: for the diagram $A \xrightarrow{\alpha} B \xrightarrow{\beta} C$ in **C**, $F(\alpha \circ \beta) = F(\alpha) \circ F(\beta)$

ii.b Identities are preserved: for any object A in the category **C**, $F(id_A) = id(F_A)$

Now, the functor (more precisely a covariant functor) from a category of neurons **CAT-Neur** to the category **Met** of metric spaces, $F : CAT - Neur \to Met$ is such that i and ii.

i every object N in the category of neurons **CAT-Neur** is mapped onto an object $F(N)$ in the category **Met**,

ii every morphism $\alpha : N \to N'$ in **CAT-Neur** is mapped onto a morphism $F(\alpha) : F(N) \to F(N')$ in the category **Met**. F preserves composition and identity.

  ii.a The composition is preserved, so $A \xrightarrow{\alpha} B \xrightarrow{\beta} C$ in **CAT-Neur**, $F(\alpha \circ_N \beta) = F(\alpha) \circ_M F(\beta)$ (both sides of the equation are morphisms in **Met**)

  ii.b Identities are preserved, so for any object A in the category **CAT-Neur**, $F(id_A) = id(F_a)$ (both sides of the equation are morphisms in **Met**)

The physiological interpretation of the functor is as follows. i means that it is possible for any object N in the category of neurons **CAT-Neur** to have associated a metric space (X, d). As it was stated in 8.3.1, the objects in the sets of category **CAT-Neur** are sets of neurons.

Note that this is different to assign a location to a set of neurons, rather we are associating a set of neurons with a metric space $N \to (X, d)$, where $d : X \times X \to \mathbf{R}^+$.

For example, let **Met1** be the category of metric planar space of diameter 1, (M, υ), that is, $d(m, m') \leq 1$ for all $m, m' \in$ the open ball $\mathcal{M}$. Then $F(N)$, $F : N \to (M, \upsilon)$, represents that the object N, a set of neurons, falls into a circumference of diameter 1 in the two-dimensional space M.

On the other hand, if we take for the category **CAT-Neur** the category **Neur**, then condition *ii* can be interpreted as follows, whenever there is a synapse between two neurons n, n', $\alpha : n \to n'$, there is a relationship between the metric spaces associated to each of the synaptic neurons, $F(\alpha) : F(N) \to F(N')$, such that F is a map that preserves composition and identity.

Let $A \xrightarrow{\alpha} B \xrightarrow{\beta} C$, then $F(\alpha \circ \beta) = F(\alpha) \circ F(\beta)$ simply means that the map associated to a synaptic path is equivalent to the map associted to the synapses.

The other requirement, identity is preserved, can be interpreted as there is always a metric space for any neuron.

It might be remarked that the functor F defined here, does not preserve the metric space defined in the category **Met**. This is in accordance with the empirical fact that the brain has no metric or at least not a Euclidean-like metric based on the distance.

Indeed, what F does is to bring the structure of the category of neurons over to the category of metric spaces **Met**. The very different nature of the two categories that are being mapped by F, makes difficult to see how F works.

With an example we will try to make this point more clear. Let the objects of **Neur** be place cells, that is, neurons that fire when the brain occupies a position in a plane surface like for example a maze or a box. The metric space for the environment is given by the category **Met**. For every synapse α coupling two place place cells, N and N' in **Neur**. $F(N)$ and $F(N')$ are called the place fields of N and N' respectively in the category **Met**

Thus, the mapping F, in order to be a functor needs to be a structure preserving map between **Neur** and **Met**, the two categories being mapped by F. In the case that **CAT-Neur** is **Neur** whose objects are neurons, the relationship between the place field of the postsynaptic cell $F(N')$ and the place field of the presynaptic cell $F(N)$ may be exemplified by $d(F(N_i), F(N_j)) \leq d(N'_i, N'_j)$, where N_i, N_j are in category **Neur**, and N'_i, N'_j in category **Met**.

8.4 A new framework for place and grid cells

Here we propose a radically new theoretical framework for the formation of place cells from grid cells. The computational models of the hippocampus

[319], [320], [321] state that the sum of a set of elements, grid cells, directly produce another element, a place cell. In doing so, these models take for granted that the properties of the sum are directly reducible to those of its components. This strict form of reductionism is at odds with the nature of complex systems. It is necessary to tackle the system as a whole, and bring to light the way in which the components interact, producing higher levels of complexity, exemplified in new systemic properties that are not present in the single components.

It might be remarked here, that this is not a criticism of the reductionist approach. Indeed the reductionist analysis is arguably the best plan of action that one may follow in order to understand how a system works. But this is just the half of the work, the synthetic endeavor must follow after the analysis.

In what follows, we describe the effect in injecting the concepts of coproduct and colimit from category theory into the problem of place cell formation in the hippocampus.

The classical reductionism credo states that the whole is no more than the sum of its parts. Therefore the properties of the sum are reduced to those of its components, without introducing new properties. This is what the categorical concept coproduct exemplifies.

In a given category, all one needs to know is about the components A_i in the coproduct $\coprod_i A_i$, this is possible because all the components play a symmetrical role in the construction coproduct.

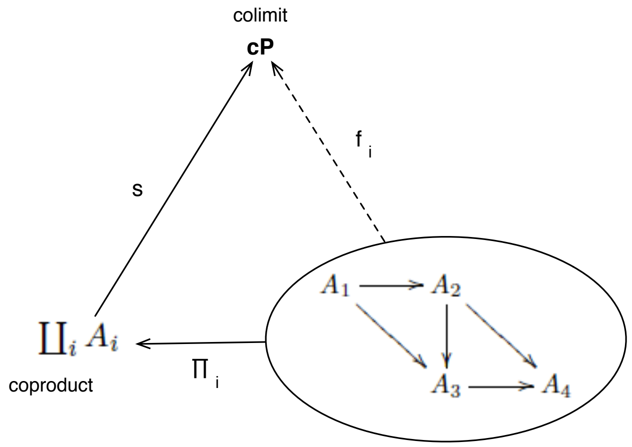

Fig. 8.7 The family of objects A_1, A_2, A_3, A_4 has both a colimit **cP** and a coproduct $\coprod_i A_i$. The coproduct is linked by s to the colimit. The link s express the transit from the coproduct to the colimit and embodies the symmetry breaking in the relationship between the family of objects A_i and the colimit

Definition 8.4. A coproduct of two objects A and B is a an object $A+B$ together with the arrows $A \xrightarrow{\iota_1} A+B$ and $B \xrightarrow{\iota_2} A+B$, such that for any object C and the pair of arrows $A \xrightarrow{\alpha} C$, $B \xrightarrow{\beta} C$, it exists an unique morphism π that makes the diagram commutes.

$$\begin{tikzcd}
A \arrow[dr, "\alpha"] \arrow[ddr, "\iota_1"'] & & B \arrow[dl, "\beta"'] \arrow[ddl, "\iota_2"] \\
& A+B \arrow[d, "\pi"] & \\
& C &
\end{tikzcd}$$

A generalised diagram of coproduct also called direct sum is shown next.

$$\begin{tikzcd}
C & \\
\coprod_i A_i \arrow[u, dotted, "(h)"] & A_i \arrow[l, "\pi_i"'] \arrow[ul, "\alpha_i"']
\end{tikzcd}$$

On the other hand, the colimit *cp* embodies the collective operations made by the family of components A_i which are made possible because the components cooperate by means of the links that connect them [12].

The colimit in a category of a family of components A_i without any arrow between them is the coproduct.

The colimit, contrary to the coproduct, entails a non symmetric relationship with its components. As the figure 8.7 depicts, the coproduct can be compared to the colimit cP. This symmetry breaking process may be somehow quantified by the arrow s.

8.4.1 Place field as colimit of grid fields

The cooperation of the grid fields gives rise to the colimit which is a place field. Thus the colimit of the metric system depicted can be seen as an integrator of the information contained in the metric system components. It might be remarked that the colimit is an object of the category **Field**, a sort of complex object that actualizes the internal organisation of the objects that is binding. Colimits and limits do not exist for all diagrams in all categories, but if they exist, they are unique up to isomorphism. The mathematical definition of colimit needs a prior definition, that of diagram.

Definition 8.5. A diagram D in a category $\mathbf{C}$ is a collection of vertices and directed edges consistently labeled with objects and arrows of the category $\mathbf{C}$. Thus, if an edge in the diagram D is labeled with an arrow f such that $f : A \to B$ in $\mathbf{C}$, then the vertices this edge in the Diagram D, must be labeled A and B [127].

Definition 8.6. Let D be a diagram in a category $\mathbf{C}$ with objects labeled D_i and morphisms labeled $f_k : D_i \to D_j$. We call cocone K for diagram D to the apical object B, together with the set of morphisms $g_i : D_i \to B$ forming a commutative diagram, that is, $g_j \circ f_k = g_i$

Given the cocones K' and K'' for D, a cocone morphism $h : B' \to B''$ is a morphism in $\mathbf{C}$ such that $g_i^{''} = h \circ g_i^{'}$. To simplify the notation we denote the cocone morphism determined by h as $h : K \to K'$. Directly, the cocones form a category, the category of cocones $\mathbf{coc_D}$.

Definition 8.7. A colimit for the diagram D is an initial object K in the category $\mathbf{coc_D}$, that is, for any other cocone K' for diagram D, there exists a unique cocone morphism $h : K \to K'$.

It follows from the definition that all colimits are isomorphic because all initial objects are isomorphic.

The figure 8.8 shows that grid fields and grid cells in the medial entorhinal cortex (mEC), are linked by a map, likewise there is a map between place cells and place fields. Therefore for each grid cell there is grid field which is a metric space with the form of a regular hexagon, and for each place cell there is one place field which is also an object of the category of metric spaces, **Field**, but in this case, its geometry a simple point rather than a hexagon.

We can assume that the neurons -place cells and grid cells- depicted in the bottom of the figure, are in the category **Neur** having as objects neurons and as morphisms synaptic connections.

However, this is not always the case. For example, a category of neurons whose objects contain several neurons connected between them forming populations of neurons, rather than single neurons. In this line, it is particularly valuable to shed light on how populations of grid cells contribute to the formation of one place cell. The colimit is the mathematical structure that allows us to encode the emergence of place field and the relationship between grid fields.

Now let us focus on the grid fields depicted as hexagons in figure 8.8 and their morphisms. It has been said above that regular hexagons are objects in the category **Field**, now we need to investigate the morphisms between the grid-field object in this category.

A contraction between two grid-field objects G_1, d, o, ψ, G_2, d, o, ψ is a continuous function $f : (G_1, d, o, \psi) \to (G_2, d, o, \psi)$, satisfying $d(f(x, y)) \leq d(x, y)$ and $o(f(x, y)) \leq o(x, y)$.

This restriction is in accordance with the experimental finding that shows that spacing in grid fields, increases along the dorsoventral axis in the medial

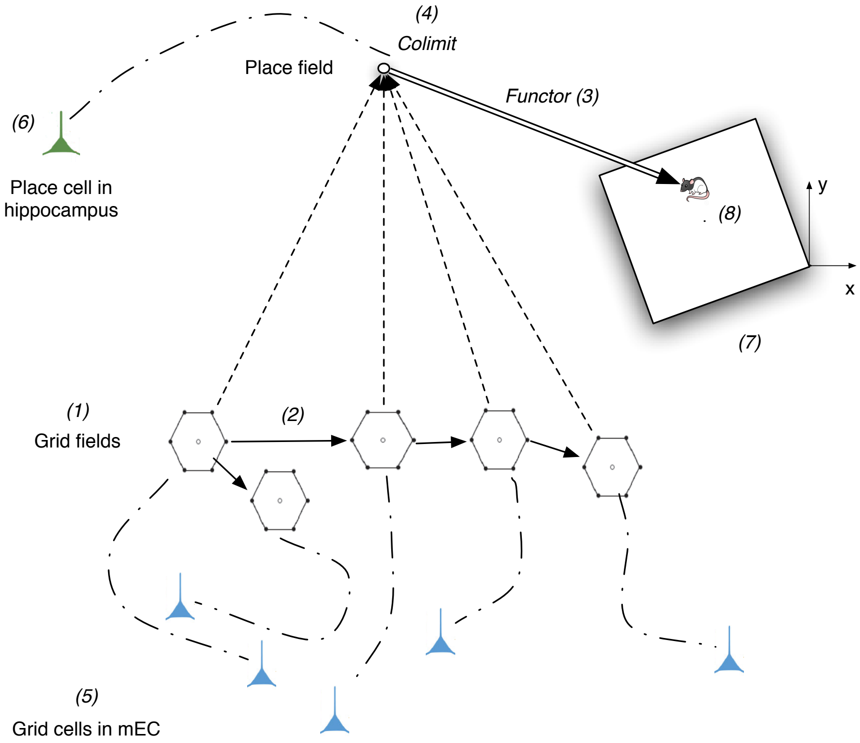

Fig. 8.8 The figure depicts a colimit where (4) acts as the place field of a place cell(6) in the hippocampus. The colimit is produced by several grid fields (one grid field(1) is produced by one grid cell(5)). (3) is a Functor between the colimit (4) and the physical location (8). (2) are connections among grid fields (1), generating produces the tesellation of the arena (7).

entorhinal cortex (mEC). This fact appears to be correlated with the increase in size of place fields along the dorsoventral axis of the hippocampus [322], [323] .

Neighbor cells in the mEC have similar spacing and orientation. However, there is no evidence that anatomical cell clusters, correspond to functionally segregated grid maps with their own spacing and orientation[315].

On the other hand, the phase of the grid, does not follow the restriction of continuity that spacing and orientation have. Indeed, firing vertices of colocalized grid cells are shifted randomly, that is to say, the mapping between vertices in the grid field and the external reference grid is not continuous. This is in fact how fields of neighboring hippocampal place cells behave.

The colimit is a universal property which is a remarkable fact that deserves to be explained. When a mathematical construction, in our case a colimit, satisfies an universal property, one can forget the details of the structure and focus on the universal property because all that has to be known about the colimit, is exemplified in the universal property.

One important point that needs emphasis is that the existence of a colimit, which imposes constraints, not only on the diagram of grid cells that determine the colimit, but also on all the objects of the category. Besides, the colimit, if it exists, is uniquely determined (up to isomorphism) but the reverse is not true, one colimit can have several decompositions. Put it in the context of figure 8.8, this means that when the coordinated activity of a group of grid cells produce a place cell, this is a colimit and it is unique. But given a place cell, its place field cannot be uniquely determined by a group of grid cells, as a matter of fact, several grid fields are possible for that place field.

Chapter 9

From cells to memories. A categorical approach

The original and innovative aspects of this thesis reside in its radically new insight into the structure, function and dynamics of the brain. I focus on the most complex system, the human brain. The theory that is described in this chapter, although at first sight may seem too pure and universal in contrast with the spurious biological realm, where the particular prevails over the universal; it may lead to a new and deeper insight into the structure and the representational power of the brain.

A theory that fully explains and predicts the highly complex cognitive abilities like perception, memory or learning has not been produced yet. Our society needs to deal with diseases like for example Alzheimer's disease that is ravaging a big sector of the population. It goes without saying that to shed light on the role played by the hippocampal system in cognitive functions like memory and learning can be of extraordinary value for the future of our own species.

We must exploit the important fact that from the point of view of neurobiological knowledge, memory and perception share the same neural substrate. The time is ripe for a mature and rigorous approach to brain structure and function that sets the basis for a shareable scientific framework, able to carry out knowledge, commonly understandable among the different actors in the brain sciences.

In this chapter I present a mathematical formalism based on the theory of categories for modeling declarative memory.

9.1 Types of memory

Just as any other higher cognitive function, to try to give a definition of memory seems hopeless. The definition in the MIT encyclopedia of cognitive sciences [324] is open enough to satisfy everyone: “the term memory implies the capacity to encode, store, and retrieve information”. However, it is also too unrestricted to provide a clear idea about what memory is and how it works.

Certainly, memory is not an univocal term, it has several forms that depend on different brain mechanisms. So a well-founded strategy to get an understanding of how memory works, is to investigate how such cognitive process is implemented in the brain. The idea behind this is that the layman's view of memory, which is still commonly used, which will become irrelevant once the biological mechanisms of memory have been uncovered and, if possible, described in mathematical terms.

Long term memory involves an anatomical change in the brain, any new memory "engenders", somehow, a new different head that you had before the memory is formed. Of course memory and the processes involved like for example learning can be tackled at different levels, like for example biochemical -PKIZ is a molecule that plays a critical role in erasing memories-, anatomical . . . .

A great deal of what it is today known about memory in humans is due to one patient called Henry Muleson, better noticed as H.M. . H.M. suffered violent epileptic seizures and in 1953 the epilepsy was surgically treated by bilateral excision of the hippocampus. As a result, the epilepsy disappeared but the secondary effect was dramatic, H.M. lost completely his capacity to recognize people that He saw everyday, the ability to remember the content of anything he read or saw vanished after the surgery. Scoville and Milner [325] concluded that the hippocampus, as H.M. case clearly showed, was related to memory functions. In essence, the hippocampus made long term memories out of short term ones. H.M. could remember his childhood and all the important events that happened prior the operation, but he could not hold on new information for more than a few minutes. The damage of hippocampus and some surrounding structure in H.M. entailed the impossibility to create new episodic-declarative knowledge. Nevertheless the patient's procedural memory, remained intact. In other words, H.M. could remember motor skills, and this explained that the manual works requested to H.M. in the laboratory improved with practice; but He could not remember at all what he did one hour ago. Thus, the distinction between declarative memory and procedural memory was clearly established.

Declarative memory can be episodic or semantic, the former encode events of one's past and the last encodes semantic knowledge that can be consciously retrieved and formulated. Declarative memory is involved in modeling the external world, that is to say, in storing representations of objects, episodes, and facts. Episodic memory underlies the capacity to reference personal experiences in a given context of time and space.

Semantic knowledge, on the other hand, is largely a context-free kind of memory. An example of semantic memory is when we build a class of similar items; if we know how a chair looks like, we will be able to, every time we see a chair, recognize it as another element of the class chair, and we do so without remember the details of the first chair we ever saw which was the first instance in the class chair. Put it simply, the semantic memory is the meaning of things.

Procedural memory is also called non declarative memory and is more an umbrella term for all those memories that are not declarative. Procedural memory includes procedural skills and the development of habits through repetition of appropriate ways to respond to stimuli. It might be noted that procedural memories are not real memories at least not in the idea of memory reminiscent to Marcel Proust *madeleine*, as the retrieval of an event from the past; rather they are skills, like for example, drive a bike or play video games, that are done increasingly better as a result of the experience.

Another major difference between these two types of memory is that while declarative memory is believed to depend on the hippocampal-entorhinal system; non declarative memory is not a construct of the hippocampal system itself, rather it involves other areas areas like the amygdala or the cerebelum. In some sense, memory is the persistence of perception. Memory, knowledge and perception are three terms that refer to a single thing, the formation of a neural structure with significant informational content, that in general can be consciously retrieved.

$$Memory = \begin{cases} Declarative\ memory & \begin{cases} Episodic \\ Semantic \end{cases} \\ \\ Procedural\ memory \end{cases}$$

9.2 A theory of declarative memory (episodic and semantic) based on category theory

The dual role of the hippocampus in formation and retrieval of concepts is not surprising, especially considering that the formation of new memory (knowledge) requires the retrieval of the old one. Thus, memory is knowledge, and perception is a condition of possibility of memory and therefore of knowledge.

The main point that is being explored in this heading is that despite the diverse nature of episodic and semantic memory, it is possible to connect them via categorical objects like product, pullback or colimit.

Let us begin by the introduction of the categorical concept of product and its application in a navigational task in one dimension, after the results will be expanded to navigation in a two-dimensional arena and the use of the categorical concept pullback. The heading finishes with a general formulation in categorical terms of the creation of semantic memories from episodic ones.

9.2.1 Categorical product in acquisition of middle point concept in 1D navigation

Suppose a rat is placed in a track (one dimensional environment), the animal immediately starts moving back and forth in order to get and idea of the dimensions of the environment. As the rat moves from one point to the other, episodic memories are created. Thus the animal is able to make the association of self-centered information with the temporal order in which the different positions are reached.

Episodic memories are not explicit, contrary to the episodic memory, the explicit ones may be retrievable independent of the internal state of the rat. Suppose there is no any particular visual or smell stimulus that can make the rat remember any particular position. One may think that after a while, the rat will acquire an explicit memory, for example the concept of middle point which exemplifies the position in the track, from where it needs the same amount of time to get any of the extremes. A cognitive behavior involves integration of information. The categorical concept of product is a formalisation of integration. Moreover, as it will be shown later, a product in a category that admits a final object, is an instance of a more general categorical form, pullback.

Definition 9.1. In a given category **C**, a product of two objects A and B, is another object P equipped with two morphisms, $P \xrightarrow{p_1} A$ and $P \xrightarrow{q_1} B$, such that for any pair of morphisms, $X \xrightarrow{f} A$ and $X \xrightarrow{g} B$ there is an unique morphism h making the following diagram commutes.

$$\begin{tikzcd}
& X \arrow[dl, "x_1"'] \arrow[d, dashed, "h"'] \arrow[dr, "x_2"] & \\
A & P \arrow[l, "p_1"] \arrow[r, "p_2"'] & B
\end{tikzcd}$$

Note that the broken arrow h means that it is unique, and the morphisms p_1, p_2 are usually called projection morphisms. The main characteristic of a product is that the constituents are retrievable via the projection morphism. The following diagram indicates the categorical product for the acquisition of the middle point.

$$\begin{tikzcd}
& X \arrow[dl, "x_1"'] \arrow[d, dashed, "h"'] \arrow[dr, "x_2"] & \\
W_A & P \arrow[l, "p_1"] \arrow[r, "p_2"'] & W_B
\end{tikzcd}$$

For our purpose, the categorical product given by the object P and the morphisms p_1, p_2 is a statement about a cognitive behavior of the rat, whereas X and x_1, x_2 is a constraint on what constitutes a valid product, rather than a specific claim about cognition.

Note that there is not any particular commitment in the morphisms p_1, p_2. In fact, p_1 can mean the travel time to reach the wall A, W_A, but also the number of steps needed.

In figure 9.1, is represented one possible experimental setting, that could be use to explore the biological plausibility of our theory in the acquisition of the middle point concept in a rat moving in a single track maze. Thus P, W_A and W_B are objects in the category $\mathbf{C}$ of memories or mental objects that will be described with more detail in future works.

9.2.2 Categorical pullback in acquisition of middle point concept in 2D navigation

Now suppose the rat is removed from the one dimensional track depicted in figure 9.1 and put upon a plane. The rat's capacity to build the explicit memory for the middle point of the arena can be seen as analogous to the generalised product -i.e: a pullback.

Definition 9.2. In a category $\mathbf{C}$, a pullback of two morphisms with common codomain $A \rightarrow^{f} C \leftarrow^{g} B$ is an object P together with a pair of morphisms $P \xrightarrow{p_1} A$ and $P \xrightarrow{p_2} B$ that form a commutative diagram $f \circ p_1 = g \circ p_2$.

Moreover, the morphisms are universal among such squares because for any pair of morphisms $Z \xrightarrow{z_1} A$ and $Z \xrightarrow{z_2} B$ such that $f \circ z_1 = g \circ z_2$, there is an unique morphism h such that the following diagram commutes

A pullback may be seen as a constrained product, being the constraint given by f and g, $f \circ p_1 = g \circ p_2$.

$$\begin{tikzcd}
P \arrow[r, "p_2"'] \arrow[d, "p_1"'] & W_B \arrow[d, "g"'] \\
W_A \arrow[r, "f"'] & W_C
\end{tikzcd}$$

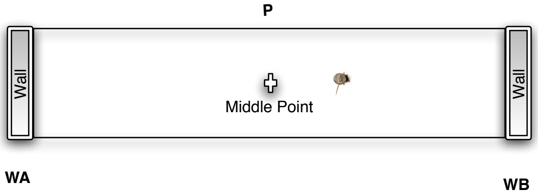

Fig. 9.1 $\mathbf{W_A}$ and $\mathbf{W_B}$ are the two walls that the rat will meet when moving in a single track maze. After reaching both walls, the animal would develop the concept of middle point **P**

9.2.3 Pullback and grid cell formation

The concept of pullback may be useful in dealing with grid cells. The way in which grid cells are calculated in literature is tricky. One of the three parameters refers to how accurate the representation outside the cell is. In doing so you are putting the outcome of the system in the input.

$$\begin{tikzcd}
P \arrow[r, "p_2"'] \arrow[d, "p_1"'] & B \arrow[d, "g"] \\
A \arrow[r, "f"'] & C
\end{tikzcd}$$

In the diagram above, P can be seen as a grid cell where the projection morphisms p_1 and p_2 refer to the orienting and the spacing respectively. The morphisms f and g impose additional restrictions in the grid cell, like for example the constant value of those parameters all over the arena.

$$\begin{tikzcd}
Z \arrow[dr, dotted, "h"] \arrow[drr, bend left, "z_2"] \arrow[ddr, bend right, "z_1"'] & & \\
& P \arrow[r, "p_2"'] \arrow[d, "p_1"] & B \arrow[d, "g"'] \\
& A \arrow[r, "f"] & C
\end{tikzcd}$$

Chapter 10

Discussion and Future work

Brain science is still in adolescence[26]. The brain networks associated with higher cognitive functions, neurocognitive networks in Bressler's words[296], arise over broadly distributed connections of large number of components, highly coupled with non linear dynamics. As a consequence, the resulting behavior of the neurocognitive networks are, in many cases, impossible to control and predict.

A mathematical explanation of how the brain is structured and how it achieves cognitive functions like perception, conceptualization or learning is seen for cognitive scientists, especially for those of humanistic background, as an extreme reductionism that obviates essential human capabilities like agency, intention, intuition, emotions or feelings.

The so called "hard sciences" have achieved mathematical certainty about the world, and although it has been always the case that formal systems fall short of capturing reality completely, this is not because these systems (mind included) are impossible to formally explain, rather it is us, as modelers who seem to have limited (perceptual) access to reality, so that we, unfortunately, usually only get partially valid descriptions.

The main claim of this thesis is that the object of study of the cognitive sciences, how the brain works, is still in a pre-scientific stage. The progress from an immature science to a mature one must pass through the construction of appropriate mechanisms with explanatory power. In brain science we must learn from the lessons geneticists have learnt in the hard way. The genome tells us nothing about how proteins fold and interact with each other or whether a particular gene plays in different functions. We face a similar problem in cognitive and brain sciences.

This thesis strives to carefully re-evaluate whether the reductionistic approach is reaching its own limits and will provide a new impulse to the scientific enquiry of brain and mind. The brain's structural and functional systems have complex features but all share commonalities. I identify three major features hold by biological networks (BN): BN evolve over time, BN are robust and BN are purposive or intentional. The topology of the biological network changes with time, that is to say, the components and the links between them, are not static but are continuously changing due to their our internal laws of functioning and to external environmental conditions as well.

The time is ripe for developing new alternatives to those short-sighted strategies and conceptions. The strategy I am proposing to be pursued here, aims to track down the underlying principles, written in formal-mathematical terms, that explain and control the organisation of the cortex and other major areas in the brain. In this vein, category theory is an analytical tool for investigating the common properties of certain objects. A pattern, as almost any other structure, is made up of its components and relations between them. This very same postulate sets up the agenda of the science of systems. What we need is to capture the patterns that occur at all the levels and put them in mathematical terms. As it was conjectured before in 7.5, the brain functioning and dynamics is scale free and on that account, category theory is the best equipped mathematical language to deal with such feature.

The time has come to set the agenda for a "hard cognitive science". To that end, I propose to translate into mathematical terms, some key concepts like perception or mental objects that until now have been used loosely, and are lacking mathematical structure. The applications shown in this work, depict the suitability of **Category Theory** as a language for complex system modeling, and as sophisticated toolkit for mental theories.

The hypothesis that has been explored and developed in this thesis is that neocortical connectivity and dynamics are scale free. The scale free dynamics could explain how brains with huge differences in size -e.g: the brain of a whale is 10,000 times bigger than the brain of a mouse, can still operate at the same time scales. The hypothesis provides a radical new interpretation to the hot spots shown in fMRI, rather than the physical localisation of cognitive functions, and may be understood as the hubs in a scale free connectivity network.

The thesis has developed three specific objectives:

1. to identify the meaningful patterns, which could serve as the building blocks for a unified theory of the brain/mind.
2. to work toward the discovery of the general principles, that mediates in the formation of meaningful patterns.
3. to provide is a real case scenario application of the theory.

The mathematical theory of categories is presented as the theoretical framework to achieve these objectives. In particular, in chapter 8, I propose a radically new theoretical framework for the formation of place cells from grid cells and in chapter 9 I present a mathematical formalism based on the theory of categories for modeling declarative memory.

This thesis is an attempt to shed some light on the nature of brain (and mind), creating new pathways and exploring fresh insights, that will lead to the comprehension of the recurrent patterns and structures that are necessarily involved in any cognitive process.

10.1 Future work

The study of the structure-function problem in the brain requires a systemic theory based approach that can attain a system level understanding that favours the construction of models that are not merely descriptive, but explanatorily powerful and scalable.

The complex sciences promote the use of mathematical theories and will bring rigor and comprehensive tools in the multilateral neuroscientific community. The theory provided in this thesis will allow to set the necessary to build on its strengths so as to develop the theories and models able to deal with multilevel interactions of coupled elements with non linear dynamics.

The work outlined in the present document is a extremely rich either in the purely conceptual mathematical side as in the experimental one. It may diverts towards many different possible ramifications. We can identify four critical challenges that are covered in the thesis and that they will need to be fully explored in the future

1. Fundamental issues and conceptual analysis
2. New mathematics of the brain
3. Complete mapping of the human brain
4. Integration and Implications

In particular, we can identify the following five future developments of the work presented here:

- To characterize the polychromous neural groups by a categorical property able to extend the Cohen-Grossberg-Hopfield differential (including delay).
- To define patterns of connections that serve as building blocks of the networks, and colimits of patterns that are meaningful patterns or neural masses with representational power
- Investigate which are the general principles that rule the emergence of meaningful patterns in the brain. The mechanisms by which patterns are created and maintained, need to be tackled based on the mathematical formulation of terns like stability, robustness or autoregulation studied in Control Theory.
- The hypothesis established in the thesis: co-operation of the grid fields gives rise to the colimit which is a place field, must be empirically validated with humans, analysing EEG data streams from the hippocampus and other brain areas involved.
- Despite the diverse nature of episodic and semantic memory, it is possible to connect memories via categorical objects like product, pullback or colimit. Place and grid cells will be tested as possible building blocks for episodic memory.

In science, many problems are still unsolved, otherwise there would not be scientists, but technicians. And many questions that we want to solve today will never have an answer, and not necessarily because they are too complex to be solved, but because they will just disappear from irrelevance. In the end, brain function and behaviour depend upon the interaction of elements (cells, networks, brains). The neural and cognitive sciences, deal with complex components (assemblies of cells, brains) and out of necessity will follow a mathematical strand. Science moves in little steps, but also makes its progress with revolutionary discoveries and concepts that sweep away whole and entire edifices of thinking and replace them with new theories that explain more with less.

The participation of experimentalists that provide EEG data analyses, and carry on test in order to validate and contrast the methodologies and theories exposed here, is the immediate future of this work and its real value for the future.

References

1. T. S. Elliot, *Four Quartets: Quartet No. 2: East Coker.* 1944.
2. M. Jeannerod, “A dichotomous visual brain?,” *Psyche*, vol. 5, no. 25, 1999.
3. D. W. Thompson, *On Growth and Form.* Cambridge University Press, 1992.
4. J. Kendrew, “The three-dimensional structure of a protein molecule,” *Scientific American*, no. 205, pp. 96–110, 1961.
5. P. Cull, “The mathematical biophysics of nicolas rashevsky,” *Biosystems*, vol. 88, no. 3, pp. 178 – 184, 2007. BIOCOMP 2005: Selected papers presented at the International Conference - Diffusion Processes in Neurobiology and Subcellular Biology, BIOCOMP2006: Diffusion Processes in Neurobiology and Subcellular Biology.
6. P. Simeonov, “Integral a post-newtonian view into the logos of bios,” *Progress in Biophysics and Molecular Biology*, vol. 102, no. 2-3, pp. 85–121, 2010.
7. E. Mayr, *This is biology.* Harvard University Press, 1997.
8. E. Schrödinger, *What Is Life? Mind and Matter.* Cambridge University Press, 1944.
9. P. Wellstead, *Systems and Control Theory for Medical Systems Biology.* IGI Global, 2009.
10. Wikipedia, “Action potential.” `http://en.wikipedia.org/wiki/Action_potential`.
11. M. Golubitsky, “Symmetry and neuroscience,” *Bulletin of the AMS, Amer. Math. Soc.*, no. January 14, 2006.
12. A. Ehresmann and J.-P. Vanbremeersch, *Memory Evolutive Systems; Hierarchy, Emergence, Cognition.* Elsevier, 2007.
13. J. Healy and T. Caudell, “Ontologies and worlds in category theory: implications for neural systems,” *Axiomathes*, vol. 16, pp. 165–214, 2006.
14. Y. Neuman and O. Nave, “On the semio-mathematical nature of codes,” *Biosemiotics*, vol. 1, pp. 99–111, 2008.
15. G. Edelman and G. Tononi, *Consciousness: how matter becomes Imagination.* London: Penguin, 2001.
16. E. Kandel, *In search of memory : the emergence of a new science of mind.* New York, NY, USA: W.W. Norton Co., 2006.
17. G. McCollum, “Social barriers to a theoretical neuroscience,” *Trends in Neurosciences*, vol. 23, no. 8, pp. 34–336(3), 2000.
18. W. Freeman, “Scale-free neocortical dynamics,” *Scholarpedia*, vol. 2, no. 2, p. 1357, 2007.
19. T. Hubel, D.H. Wiesel, *Brain and visual perception: the story of a 25-year collaboration.* Oxford University Press, 2005.
20. W. Freeman, *Mass action in the nervous system.* New York: Academic Press, 1975.
21. R. Kozma and W. Freeman, “Learning environmental clues in the kiv model of the cortico-hippocampal formation,” *Neurocomputing*, vol. 58-60, no. 6, pp. 721–728, 2004.

22. R. Kozma and W. Freeman, "Basic principles of the kiv model and its application to the navigation problem," *J Integrative Neurosci*, vol. 2, pp. 125–145, 2003.
23. R. Kozma, "Intentional dynamic systems. ncarai seminar series may 9, 2005, washington dc," 2005.
24. M. Fuhs and D. Touretzky, "A spin glass model of path integration in rat medial entorhinal cortex," *J. Neurosci.*, vol. 26, pp. 4266–76, 2006.
25. J. Taube, "Head direction cells," *Scholarpedia*, vol. 4, no. 12, p. 1787, 2009.
26. W. J. Freeman and R. Kozma, "Freeman's mass action," *Scholarpedia*, vol. 5, no. 1, p. 8040, 2010.
27. D. Noble, "Biophysics and systems biology," *Phil. Trans. R. Soc. A*, vol. 368, no. 1914, pp. 1125–1139, 2010.
28. J. M. Bower, "Looking for newton: Realistic modeling in modern biology," *Brains, Minds and Media*, no. 2, 2005.
29. M. Barr and C. Wells, *This is biology: the science of the living world.* Cambridge: Belknap Press of Harvard University Press, 1997.
30. S. Brenner, "Theoretical biology in the third millennium," *Phil. Trans. R. Soc. Lond. B*, vol. 7, pp. 1963–1965, 1999.
31. J. Knight, "Bridging the culture gap," *Nature*, vol. 419, pp. 244–246, 2002.
32. T. Kuhn, *The Structure of Scientific Revolutions.* University of Chicago Press, 1962.
33. T. Kuhn, *The Road Since Structure: Philosophical Essays:1970-1993.* University of Chicago Press, 2000.
34. F. Crick, "Central dogma of molecular biology," *Nature*, vol. 227, pp. 561–563, 1996.
35. S. B. Prusiner, "Molecular biology of prion diseases," *Science*, vol. 252, no. 5012, pp. 1515–1522, 1991.
36. R. Carnap, *The Logical Structure of the World. Pseudoproblems in Philosophy.* University of California Press., 1967.
37. B. Jacrot, *Physique et biologie.* Edp Sciences, 2006.
38. R. Gatenby and P. Maini, "Cancer summed up," *Nature*, vol. 421, p. 321, 2003.
39. G. Deco, V. Jirsa, P. Robinson, M. Breakspear, and K. Friston, "The dynamic brain : From spiking neurons to neural masses and cortical fields," *PLoS Comput Biol.*, vol. 4, no. 8, pp. 23–37, 2008.
40. R. Brown, "What is a brain state?," *Philosophical Psychology*, vol. 19, no. 6, pp. 729–742, 2006.
41. T. Polger, *Natural Minds.* The MIT Press, 2004.
42. W. Bechtel and J. Mundale, "Multiple realizability revisited: Linking cognitive and neural states," *Philosophy of Science*, vol. 66, pp. 175–207, 1999.
43. G. Buzsaki, *Rhythms of the Brain.* Oxford University Press, 2006.
44. D. Pare and R. Llinas, "Conscious and pre-conscious processes as seen from the standpoint of sleep-waking cycle neurophysiology," *Neuropsycholgia*, vol. 33, no. 9, pp. 1155–1168, 1995.
45. P. L. Nunez, *Neocortical Dynamics and Human EEG Rhythms.* Oxford University Press, 1995.
46. P. L. Nunez, "Toward a quantitative description of large-scale neocortical dynamic function and eeg.," *Behav Brain Sci.*, vol. 23, no. 3, pp. 371–98, 2000.
47. D. Chalmers, "Facing up to the problem of consciousness," *Journal of Consciousness Studies*, vol. 2, no. 3, pp. 200–219, 1995.
48. J. Fell, "Identifying neural correlates of consciousness: The state space approach.," *onsciousness and Cognition*, vol. 13, no. 4, pp. 709–729, 2004.
49. G. Werner, "Viewing brain processes as critical state transitions across levels of organization: Neural events in cognition and consciousness, and general principles," *Biosystems*, vol. 96, no. 1, pp. 114 – 119, 2009.
50. F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain," *Cornell Aeronautical Laboratory, Psychological Review*, vol. 65, no. 6, pp. 386–408, 1958.

51. A. Hodgkin and A. Huxley, "A quantitative description of membrane current and its application to conduction and excitation in nerve," *Journal of Physiology*, vol. 117, pp. 500–544, 1982.
52. W. Rall, "Rall model," *Scholarpedia*, vol. 4, no. 4, p. 1369, 2009.
53. E. De Schutter, "Why are computational neuroscience and systems biology so separate?," *PLoS Computational Biology*, vol. 4, no. 5, 2008.
54. S. Solinas, R. Maex, and E. De Schutter, "Dendritic amplification of inhibitory postsynaptic potentials in a model purkinje cell," *European Journal of Neuroscience*, vol. 23, no. 4, pp. 1207–1218, 2006.
55. W. Rall, *Core conductor theory and cable properties of neurons.* Bethesda MD: American Physiology society, 1977.
56. W. Gerstner and W. M. Kistler, *Spiking Neuron Models. Single Neurons, Populations, Plasticity.* Cambridge University Press, 2002.
57. P. Bressloff, "A green's function approach to analysing the effects of random synaptic background activity in a model neural network," *J. Phys. A: Math. Gen.*, vol. 27, no. 12, p. 4097, 2008.
58. L. Abbott, "Lapique's introduction of the integrate-and-fire model neuron," *Brain Research Bulletin*, vol. 50, no. 5-6, pp. 303–304, 1999.
59. W. Gerstner and R. Naud, "Neuroscience: How good are neuron models?," *Science*, vol. 326, no. 5951, pp. 379–380, 2009.
60. C. Meunier and I. Segev, "Playing the devil's advocate: is the hodgkin-huxley model useful?," *Trends Neurosci.*, vol. 25, no. 11, pp. 558–63, 2002.
61. A. Herz, T. Gollisch, C. K. Machens, and D. Jaeger, "Modeling Single-Neuron Dynamics and Computations: A Balance of Detail and Abstraction," *Science*, vol. 314, no. 5796, pp. 80–85, 2006.
62. H. Markram, "The blue brain project," *Nature Neuroscience*, vol. 7, pp. 153–8, 2006.
63. W. McCulloch and W. H. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bulletin of Mathematical Biophysics*, vol. 5, pp. 115–133, 1943.
64. J. Searle, "Minds, brains and programs," *Behavioral and Brain Sciences*, no. 3, pp. 417–424, 1980.
65. S. Harnad, "The symbol grounding problem," *Physica D*, vol. 42, no. 10, pp. 335–346, 1990.
66. P. R. Cohen, "If not turing's test, then what?," *AI magazine*, vol. 26, no. 4, pp. 335–346, 2006.
67. W. Hoffman, "Are neural networks a good model for cognition?," *Psycoloquy*, vol. 9, no. 12, 1998.
68. C. Eliasmith, "Attractor network," *Scholarpedia*, vol. 2, no. 10, p. 1380, 2007.
69. M. Fox and M. Raichle, "Spontaneous fluctuation in brain activity observed with functional magnetic resonance imaging," *Nat Neurosci Rev*, vol. 8, pp. 700–711, 2007.
70. P. Dayan and L. Abbott, *Theoretical neuroscience: computational and mathematical modeling of neural systems.* MIT Press., 2002.
71. G. West, J. Brown, and B. Enquist, "A general model for the origin of allometric scaling laws in biology," *Science*, pp. 122–126, 1997.
72. C. Castellano, F. S., and V. Loreto, "Statistical physics of social dynamics," *Rev. Mod. Phys.*, vol. 81, pp. 591–646, 2009.
73. S. Strogatz and I. Stewart, "Coupled oscillators and biological synchronization," *Scientific American*, vol. 12, no. 1993, pp. 102–109, 1993.
74. H. Strogatz, *Nonlinear Dynamics and Chaos.* Addison Wesley publishing company, 1994.
75. Y. Manor, J. Rinzel, I. Segev, and Y. Yarom, "Low-amplitude oscillations in the inferior olive: a model based on electrical coupling of neurons with heterogeneous channel densities," *J. Neurophysiol.*, vol. 77, pp. 2736–2752, 1997.
76. M. Steriade, L. Domich, G. Dakson, and M. Deschenes, "The deafferented reticular thalamic nucleus generates spindle rhythmicity," *J. Neurophysiol.*, vol. 57, pp. 260–273, 1987.

77. J. Caplan, "Human θ oscillations related to sensorimotor integration and spatial learning," *J. Neurosci.*, vol. 23, pp. 4726–4736, 2003.
78. B. Hutcheon and Y. Yarom, "Resonance, oscillation and the intrinsic frequency preferences of neurons," *Trends in Neurosciences*, vol. 23, no. 5, pp. 216 – 222, 2000.
79. A. Boehlen, U. Heinemann, and I. Erchova, "The range of intrinsic frequencies represented by medial entorhinal cortex stellate cells extends with age.," *J Neurosci.*, vol. 30, no. 13, pp. 4585–9, 2010.
80. J. Perez Velazquez, "Brain, behaviour and mathematics: are we using the right approaches?," *Physica D*, vol. 212, pp. 161–182, 2005.
81. T. Frank, A. Daffertshofer, C. Peper, P. Beek, and H. Haken, "Towards a comprehensive theory of brain activity: coupled oscillator systems under external forces," *Physica D*, vol. 14, pp. 62–86, 2000.
82. F. Hoppensteadt and E. Izhikevich, *Weakly Connected Neural Networks.* Springer, 1997.
83. J. Collins and I. Stewart, "Coupled nonlinear oscillators and the symmetries of animal gaits," *J. Nonlin. Sci.*, vol. 349, pp. 349–392, 1999.
84. M. Golubitsky, I. Stewart, P.-L. Buono, and J. Collins, "Symmetry in locomotor central pattern generators and animal gaits," *Nature*, vol. 401, pp. 693–695, 1999.
85. P. Buono and M. Golubitsky, "Models of central pattern generators for quadruped locomotion," *J. Math. Biol.*, vol. 42, no. 4, pp. 291–326, 2001.
86. I. Stewart, "Network opportunities," *Nature*, vol. 427, pp. 601–604, 2004.
87. F. Theunissen, J. Roddey, S. Stufflebeam, H. Clague, and J. Miller, "Information theoretic analysis of dynamical encoding by four primary sensory interneurons in the cricket cercal system," *J. Neurophysiol.*, vol. 75, pp. 1345–1359, 1996.
88. O. David and K. Friston, "A neural mass model for meg/eeg coupling and neuronal dynamics," *NeuroImage*, vol. 20, pp. 102–109, 2003.
89. E. Haskell, D. Q. Nykamp, and D. Tranchina, "A population density method for large-scale modeling of neuronal networks with realistic synaptic kinetics," *Neurocomputing*, vol. 38-40, pp. 627 – 632, 2001.
90. M. Breakspear and C. Stam, "Dynamics of a neural system with a multiscale architecture," *Phil. Trans. R. Soc. B*, no. 360, pp. 1051–1074, 2005.
91. W. Freeman and W. Schneider, "Changes in spatial patterns of rabbit olfactory eeg with conditioning to odors," *Psychophysiology*, vol. 19, no. 1, pp. 44–56, 1982.
92. W. Freeman, *Neurodynamics: An Exploration In Mesoscopic Brain Dynamics.* Springer-Verlag, 2008.
93. M. Griniasty, M. Tsodyks, and D. Amit, *Conversion of temporal correlations between stimuli to spatial correlations between attractors.* MIT Press, 1999.
94. R. Cossart, D. Aromov, and R. Yuste, "Attractor dynamics of network up states in the neocortex," *Nature*, vol. 423, no. 5723, pp. 283–288, 2003.
95. D. Hebb, *The organization of behavior.* NY: Wiley, 1949.
96. T. J. Wills, C. Lever, F. Cacucci, N. Burgess, and J. O'Keefe, "Attractor Dynamics in the Hippocampal Representation of the Local Environment," *Science*, vol. 308, no. 5723, pp. 873–876, 2005.
97. F. L. da Silva and e. a. , "Epilepsies as dynamical diseases of brain systems: basic models of the transition between normal and epileptic activity," *Epilepsia*, vol. 44, no. 12, pp. 72–83, 2003.
98. A. Babloyantz and A. Destexhe, "Low-dimensional chaos in an instance of epilepsy," *Babloyantz*, vol. 83, no. 12, pp. 3513–3517, 1986.
99. M. S. Titcombe and e. a. , "Dynamics of parkinsonian tremor during deep brain stimulation," *Chaos*, vol. 11, pp. 766–773, 2001.
100. A. Prinz, A. Bucher, and D. Marder, "Similar network activity from disparate circuit parameters," *Nat. Neuroscience*, vol. 7, pp. 1345–1352, 2000.
101. J. Green and A. Arduini, "Hippocampal activity in arousal," *J. Neurophysiol*, vol. 17, no. 6, pp. 533–57, 1954.

102. G. Buzsaki, “Theta oscillations in the hippocampus,” *Neuron*, vol. 33, no. 3, pp. 325–40, 2002.
103. G. Buzsaki, “Theta rhythm of navigation: link between path integration and landmark navigation, episodic and semantic memory,” *Hippocampus*, vol. 15, no. 7, pp. 827–40, 2005.
104. E. Mac Cormac and M. Stamenov, *Fractals of brain, fractals of mind: in search of a symmetry bond.* John Benjamins B.V., 1996.
105. C. Nunn, *From Neurons to Notions: Brains, Mind and Meaning.* Floris Books, 2008.
106. H. Haken, *Synergetics: An Introduction.* Berlin: Springer, 1983.
107. H. Haken, “Synergetics,” *Scholarpedia*, vol. 2, no. 1, p. 1400, 2007.
108. H. Haken, “Synergetics of brain function,” *International Journal of Psychophysiology*, vol. 60, no. 2, pp. 110–124, 2006.
109. A. Pellionisz and R. Llinás, “Tensor network theory of the metaorganization of functional geometries in the cns,” *Neuroscience*, vol. 16, pp. 245–273, 1985.
110. V. Makarenko and R. Llinás, “Experimentally determined chaotic phase synchronization in a neuronal system,” *Proc. Natl. Acad. Sci. U.S.A.*, vol. 95, pp. 15747–15752, 1998.
111. E. Leznik, V. Makarenko, and R. Llinás, “Electrotonically mediated oscillatory patterns in neuronal ensembles: an in vitro voltage-dependent dye imaging study in the inferior olive,” *J. Neurosci.*, vol. 22, no. 7, pp. 2804–2815, 2002.
112. S. Roy and R. Llinás, *Dynamic geometry, brain function modeling, and consciousness.* Elsevier B.V., 2008.
113. K. Kaneko, *Simulating Physics with Coupled Map Lattices.* World sci., 1990.
114. K. Kaneko and I. Tsuda, *Complex Systems: Chaos and Beyond, A Constructive Approach with Applications in Life Sciences.* Springer Verlag, 2000.
115. P. Erdos and A. Renyi, “On the evolution of random graphs,” *Publ. Math. Inst. Hung. Acad. Sci.*, vol. 5, pp. 17–61, 1960.
116. R. Kozma, “Neuropercolation,” *Scholarpedia*, vol. 2, no. 8, p. 1360, 2007.
117. J. M. Beggs, “Neuronal avalanche,” *Scholarpedia*, vol. 2, no. 1, p. 1344, 2007.
118. J. Beggs and D. Plenz, “Neuronal Avalanches in Neocortical Circuits,” *J. Neurosci.*, vol. 23, no. 35, pp. 11167–11177, 2003.
119. H. M. Breskin and e. a. , “Percolation in living neural networks,” *Physical Review Letters*, vol. 97, no. 18, 2006.
120. M. Boden, *What Computers Can’t Do: The Limits of Artificial Intelligence.* New York: Harper Row, 1972.
121. M. Boden, *Mind as machine.* Oxford University Press, 2006.
122. J. W. Lichtman, J. Livet, and J. R. Sanes, “A technicolour approach to the connectome,” *Nature Rev. Neurosci.*, vol. 9, pp. 417–422, 2008.
123. J. Lehrer, “Neuroscience: Making connections,” *Nature*, vol. 457, pp. 524–527, 2009.
124. T. Gowers, J. Barrow-Green, and I. Leader, *The Princeton Companion to Mathematics.* Princeton University Press, 2008.
125. M. Barr and C. Wells, *Category Theory for Computing Science.* Montreal: CRM Publications, 1999.
126. S. Mac Lane, “Concepts and categories in perspective,” *Duren, P. (Ed.)*, vol. 21, pp. 353–365, 1988.
127. B. Pierce, *Arrows, structures, and functors: the categorical imperative.* The MIT Press, 1991.
128. S. Awodey, *Category Theory.* Oxford University Press, 2006.
129. S. Mac Lane, *Mathematics Form and Function.* Springer, 1986.
130. R. Geroch, *Mathematical physics (Lectures in Physics).* University of Chicago Press, 1985.
131. S. Mac Lane, *Categories for the Working Mathematician.* Springer, 1998.
132. F. Lawvere and S. Schanuel, *Conceptual Mathematics: A First Introduction to Categories.* Cambridge University Press, 1995.

133. M. Arbib and E. Manes, *Arrows, structures, and functors: the categorical imperative.* New York: Academic Press, 1975.
134. A. Schalk and H. Simmons, *Introduction to Category theory in for easy movements.* www.cs.man.ac.uk/mfg, 2005.
135. J. Macnamara, *The Logical Foundations of Cognition.* Oxford University Press, 1994.
136. W. C. Hoffman, "The dialectics of mind," *The Journal of Mind Theory*, vol. in press, 2010.
137. S. Philips, W. Wilson, and G. Hadford, "What do transitive inference and class inclusion have in common? categorical (co)products and cognitive development," *PLoS Comput Biol.*, vol. 5, no. 12, 2009.
138. R. Brown and T. Porter, "Category theory and higher dimensional algebra: potential descriptive tools in neuroscience," in *Proceedings of the International Conference on Theoretical Neurobiology, edited by Nandini Singh, National Brain Research Centre, Conference Proceedings 1 (2003) 80-92*, (Delhi, India), February 2003.
139. E. C. Zeeman, *The Topology of the Brain and Visual Perception.* New York: Prentice-Hall, 1962.
140. J. L. Fidaeiro, *Categories for Software Engineering.* Springer, 2004.
141. M. Kary, "(math, science, ?)," *Axiomathes*, 2009.
142. R. Rosen, "Autobiographical reminiscences of rober rosen," *Axiomathes*, vol. 16, pp. 1–23, 2006.
143. J.-C. Letelier, J. Soto-Andrade, F. G. Abarzúa, A. Cornish-Bowden, and M. L. Cárdenas, "Organizational invariance and metabolic closure: Analysis in terms of (m,r) systems," *Journal of Theoretical Biology*, vol. 238, no. 4, pp. 949 – 961, 2006.
144. Z. Arzi-Gonczarowski and D. J. Lehmann, "From environments to representations: A mathematical theory of artificial perceptions," *Artificial Intelligence*, pp. 187–247, 1998.
145. M. Healy, "Category theory applied to neural modeling and graphical representations," *Proceedings of IJCNN 2000, IEEE*, 2000.
146. Y. Neuman and O. Nave, "A mathematical theory of sign-mediated concept formation," *Applied Mathematics and Computation*, 2008.
147. C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, no. 27, pp. 370–423, 1948.
148. R. H. Tamarin, *Principles of genetics.* New York: McGraw-Hill, 2002.
149. J. Gómez and R. Sanz, "To cognize is to categorize revisited: Category theory is where mathematics meets biology," *Proceedings of AAAI 2009*, 2010.
150. E. Rosch and C. B. Mervis, "Family resemblances: studies in the internal structure of categories," *Cognitive Psychology*, vol. 7, pp. 573–605, 1975.
151. E. Rosch, *Cognition and Categorization.* Erlbaum, 1978.
152. J. Gomez et al., "Naturalized epistemology for autonomous systems," in *Naturalised Epistemology Workshop*, (Kazimierz, Poland), September 2007.
153. S. Philips and W. H. Wilson, "Categorial compositionality: A category theory explanation for the systematicity of human cognition," *PLoS Comput Biol.*, vol. 6, no. 7, p. 14, 2010.
154. R. M. Shiffrin, "Perspectives on modeling in cognitive science," *Topics in Cognitive Science*, vol. 2, no. 4, pp. 736–750, 2010.
155. N. Koblitz, "A tale of three equations; or the emperors have no clothes," *The Mathematical Intelligencer*, vol. 10, no. 1, pp. 4–10, 1988.
156. S. Huntington, *Political Order in Changing Societies.* Yale University Press, 1968.
157. J. Brown, *Philosophy of Mathematics: an Introduction to the World of Proofs and Pictures.* London and New York: Routledge, 1999.
158. N. Koblitz, *Mathematics as Propaganda.* Springer Verlag, 1981.
159. H. Simon, "Some trivial but useful mathematics," *unpublished paper*, 1987.
160. R. Wille and U. Wille, "On the controversy over huntington's equations : When are such equations meaningful?," *Mathematical Social Sciences*, vol. 25, no. 2, pp. 173 – 180, 1993.

161. J.-P. Changeaux and A. Connes, *Matière à pensée.* Odile Jacob, 2000.
162. E. Nagel, *The Structure of Science: Problems in the Logic of Scientific Explanation.* Hacket Publishing Company, 1979.
163. M. Bunge, *La Investigación Científica.* Ariel España, 1969.
164. M. Bunge, *Emergence and Convergence: Qualitative Novelty and the Unity of Knowledge.* University of Toronto Press, 2003.
165. M. Bunge, "Some topical problems in biophilosophy," *Journal of Social and Biological Systems*, vol. 2, no. 2, pp. 155 – 172, 1979.
166. R. Courant and H. Robbins, *What Is Mathematics? An Elementary Approach to Ideas and Methods.* Oxford University Press, 1996.
167. D. Bohm, *The Essential David Bohm.* London: Routledge, 2002.
168. E. Dehaene, "A few steps toward a science of mental life," *Mind, Brain, and Education*, vol. 1, no. 1, pp. 28–47, 2007.
169. R. Luce, "Four tensions concerning mathematical modeling in psychology," *Annual Reviews of Psychology*, no. 46, pp. 1–26, 1995.
170. L. Smolin, *The Trouble With Physics.* Houghton Mifflin Harcourt, 2006.
171. D. van Dalen, "How connected is the intuitionistic continuum?," *The Journal of Symbolic Logic*, vol. 62, no. 4, pp. 1147–1150, 1997.
172. A. Connes, "A view of mathematics," http://www.alainconnes.org/docs/maths.pdf.
173. J. Fodor, *The Modularity of Mind.* Cambridge, Massachusetts: MIT Press, 1983.
174. P. Wang, "Three fundamental misconceptions of artificial intelligence," *Journal of Experimental Theoretical Artificial Intelligence*, vol. 19, no. 3, pp. 249–268, 2007.
175. G. Longo, "Mathematical intelligence, infinity and machines: Beyond the godelitis.," *Journal of Consciousness Studies*, no. 6, pp. 191–214, 1997.
176. L. Dive, *An Epistemic structuralist account of mathematical knowledge.* PhD thesis, 2003.
177. J. Petitot et al., *Naturalizing Phenomenology: Issues in Contemporary Phenomenology and Cognitive Science.* Stanford University Press, 1999.
178. N. Wiener, *I am a mathematician: The later life of a prodigy.* Garden City NY: Doubleday Company, 1956.
179. S. Salthe, "The spontaneous origin of new levels in a scalar hierarchy," *Entropy*, vol. 6, pp. 327–343, 2004.
180. S. Salthe, "A hierarchical framework for levels of reality: Understanding through representation," *Axiomathes*, no. 19, pp. 87–99, 2009.
181. N. Baas and C. Emmeche, "On emergence and explanation," *Intellectica*, no. 25, pp. 67–83, 1997.
182. H. Putnam, *Reason, Truth and History.* Cambridge: Cambridge University Press, 1981.
183. M. Dummett, *Thought and Reality.* Oxford University Press, 2006.
184. J. Holland, *Emergence: From Chaos To Order.* Basic Books, 1999.
185. M. Shanahan, *Solving the Frame Problem: A Mathematical Investigation of the Common Sense Law of Inertia.* The MIT Press, 1997.
186. I. C. Baianu and e. a. , "Categorical ontology of complex spacetime structures: The emergence of life and human consciousness," *Axiomathes*, no. 17, pp. 223–352, 2007.
187. C. Pawlowitsch, "Why evolution does not always lead to an optimal signaling system," *Games Econ. Behav.*, no. 63, pp. 203–226, 2008.
188. E. O. Wilson, *Consilience.* New York: Knopf, 1998.
189. G. Tononi, O. Sporns, and G. Edelman, "Measures of degeneracy and redundancy in biological networks," *PNAS*, vol. 96, no. 6, pp. 3257–3262, 1999.
190. J. Gómez et al., "Cognitive ontologies: Mapping structure and function of the brain from a systemic view," *proceedings in AAAI*, 2008.
191. J. Derrida and E. Roudinesco, *For what tomorrow: a dialogue.* 2004.
192. E. Livingston, "Cultures of proving," *Social Studies of Science*, vol. 29, no. 6, pp. 867–888, 1999.

193. E. Landry, "Shared structure need not be shared set-structure," *Synthese*, vol. 158, pp. 1–17, 2007.
194. J. Franklin, "Philosophy mathematics and structure," *Philosopher*, vol. 1, no. 2, pp. 867–888, 1995.
195. A. Zee, *Fearful Symmetry: The Search for Beauty in Modern Physics.* Princeton Science Library, 2007.
196. G. A. Miller, "Charge densities of the neutron and proton," *Phys. Rev. Lett.*, vol. 99, p. 112001, Sep 2007.
197. W. Elasser, *The physical fundation of biology.* Pergamon Press, 1958.
198. M. Katz, *Templets and complet patterns.* Cambridge University Press, 1986.
199. M. A. Arbib and M. B. Hesse, *The construction of reality.* Cambridge University Press, 1986.
200. S. Mac Lane, "The genesis of mathematical structures," *Cahiers Topol. Geom.*, vol. 21, pp. 353–365, 1980.
201. S. Shapiro, "Space, number and structure: A tale of two debates," *Philosophia Mathematica*, vol. 4, no. 3, pp. 148–173, 1996.
202. L. Skyttner, *General Systems Theory: Problems, Perspectives, Practice.* World Scientific Publishing Company, World Scientific Publishing Company.
203. E. Landry, "Category theory: The language of mathematics," *Philosophy of Science*, vol. 66, pp. 14–27, 1998.
204. S. Shapiro, "Structure in mathematics and logic: A categorical perspective," *Philosophia Mathematica*, vol. 4, no. 4, pp. 209–237, 1996.
205. E. Landry, "Category theory as a framework for an in re interpretation of mathematical structuralism," in *The Age of Alternative Logics* (J. Benthem, G. Heinzmann, M. Rebuschi, and H. Visser, eds.), pp. 163–179, Springer Netherlands, 2006. 10.1007/978-1-4020-5012-$7_1$2.
206. R. Goodrich, "An agent-based conception of models and scientific representation," *The British Journal of Aesthetics*, vol. 28, no. 1, pp. 48–58, 1988.
207. P. Teller, "Twilight of the perfect model model," *Erkenntnis*, vol. 55, no. 3, 2001.
208. M. Suarez, "Scientific representation: Against similarity and isomorphism," *International Studies in the Philosophy of Science*, vol. 17, no. 3, pp. 225–244, 2003.
209. C. Carver, "When mechanistic models explain," *Synthese*, 2006.
210. J. Woodward, *Making Things Happen: A Theory of Causal Explanation.* Oxford University Press, 2005.
211. "Einstein year." `http://www.einsteinyear.org/facts/brownian_motion/`.
212. G. Galilei, *Dialogues concerning two new sciences.* Dover Publications, 1954.
213. R. J. Matthews, *The Measure of Mind Propositional Attitudes and Their Attribution.* Oxford University Press, 2007.
214. L. Narens, *Introduction to the Theories of Measurement and Meaningfulness and the Use of Invariance in Science.* Lawrence Erlbaum Associates, 2007.
215. R. N. Giere, *Explaining science: a cognitive approach.* University of Chicago Press, 1990.
216. R. N. Giere, *Science without laws.* University of Chicago Press, 1999.
217. R. Giere, "An agent-based conception of models and scientific representation," *Synthese*, 2009.
218. S. Shapiro, *Thinking about Mathematics.* Oxford University Press, 2000.
219. S. Shapiro, *Oxford Handbook of the Philosophy of Mathematics and Logic.* Oxford University Press, 2005.
220. G. Birkhoff and S. Mac Lane, *A Survey of Modern Algebra.* The Macmillan Co., 1946.
221. F. Bonsignorio, "On the stochastic stability and observability of controlled serial kinematic chains," in *Proceedings of the ASME 2010 10th Biennial Conference on Engineering Systems Design and Analysis ESDA2010*, (Turkey), July 2010.
222. H. Poincaré, *Science and Hypothesis.* New York, NY, USA: Dover Publications, 1952.
223. E. Kandel, *Principles of Neural Science.* New York, NY, USA: McGraw-Hill, 2000.
224. K. Lashley and G. Clark, "The cytoarchitecture of the cerebral cortex of ateles: a critical examination of the architectoic studies," *J. Comp. Neurol.*, no. 85, pp. 223–305, 1946.

225. C. Wernicke, *The symptom-complex of aphasia.* New York, NY, USA: Appleton and Co., 1908.
226. J. Fuster, "The module: crisis of a paradigm book review, "the new cognitive neurosciences" 2nd edition, m.s. gazzaniga, editor-in-chief, mit press," *Neuron*, no. 26, pp. 51–53, 2000.
227. D. Gabbay and F. Gunthner, *Handbook of Philosophical Logic 2nd edn, Vol. 13.* New York, NY, USA: Kluwer, 2005.
228. W. Wimsatt, *Aggregate, composed, and evolved systems: Reductionistic heuristics as means to more holistic theories.* Springer, 2006.
229. A. Clark, *Being There: Putting Brain, Body and World Together Again.* MIT Press, Bradford Books, 1997.
230. W. Bechtel, "The compatibility of complex systems and reduction: A case analysis of memory research. minds and machines," *Minds and Machines*, no. 11, pp. 83–502., 2001.
231. *Gray's Anatomy: The Anatomical Basis of Clinical Practice, 40th edition.* Churchill-Livingstone, Elsevier, 2008.
232. T. Haruna, *Algebraic Theory of Biological Organization.* PhD thesis, 2008.
233. D. MacKay, "Mind-like behaviour in artefacts," *J Philos Sci*, vol. III, pp. 352–353, 1953.
234. D. MacKay, "On comparing the brain with machines," *part of a symposium on "Cybernetics" at the Sept. 1953; see a report by E. C. Cherry in Nature, 1953, 172, 648-49; reprinted in the American Scientist, 1954, 42, 261-268. ; reprinted in the Annual Report of the Smithsonian Institute, 1954, 231-240*, pp. 231–240, 1953.
235. D. MacKay, "Towards an information-flow model of cerebral organization," *Advancement of Science, part of a symposium on "Cerebral Activity" at the Sept. 1955 British Association meetings in Bristol; see a report by M. Piercy in Nature, 1955, 176, 576-78*, vol. 12, no. 42, pp. 392–395, 1956.
236. T. J. Sejnowski et al., "Computational neuroscience," *Science*, vol. 9, pp. 1299–1306, 1988.
237. I. Thomas, "Tracing the brain's connections," 2010.
238. G. Mack, "Universal dynamics, a unified theory of complex systems: Emergence, life and death," *Commun.Math.Phys.*, no. 219, pp. 141–178, 2001.
239. D. Heeger and D. Ress, "What does fmri tell us about neuronal activity?," *Nature Reviews Neuroscience*, vol. 3, no. 2, p. 142, 2002.
240. N. Logothetis, "What we can do and what we cannot do with fmri," *Nature*, vol. 453, no. 7197, p. 869, 2008.
241. A. M. de Rezende, "Le point de départ dans la philosophie de merleau-ponty," *Revue Philosophique de Louvain*, vol. 73, no. 19, pp. 451–480, 1975.
242. B. Bakker, "The concept of circular causality should be discarded. commentary on marc d. lewis: Bridging emotion theory and neurobiology through dynamic system modeling," *Behavioral and Brain Sciences*, vol. 28, pp. 195–196, 2005.
243. A. K. Seth, "Causal networks in simulated neural systems.," *Cognitive neurodynamics*, vol. 2, pp. 49–64, March 2008.
244. B. Palsson, *Systems Biology: Properties of Reconstructed Networks.* Cambridge University Press, 2006.
245. J. G. White and e. a. , "Factors that determine connectivity in the nervous system of caenorhabditis elegans," *Cold Spring Harb Symp Quant Biol 48 Pt 2*, vol. 23, no. 48 Pt 2, pp. 633–40, 1983.
246. J. Smith, K. C. R. D. Ellenberger, H.H. C. Ballanyi, and J. Feldman, "Pre-botzinger complex: a brainstem region that may generate respiratory rhythm in mammals," *Science*, vol. 254, no. 5032, pp. 726–729, 1991.
247. V. Jirsa, "Connectivity and dynamics of neural information processing," *Neuroinformatics*, vol. 2, pp. 183–204, 2004.
248. P. Churchland, *Neurophilosophy Toward a Unified Science of the Mind-Brain.* The MIT Press, 1986.
249. W. Wiltschko and R. Wiltschko, "Collective dynamics of 'small-world' networks," *Journal of Experimental Biology*, vol. 199, no. 1, pp. 29–38, 1996.
250. B. Horwitz, "The elusive concept of brain connectivity," *NeuroImage*, vol. 19, pp. 466–470, June 2003.

251. P. Rodriguez, “Neural decoding of goal locations in spatial navigation in humans with fmri,” *Human Brain Mapping*, vol. 31, no. 3, pp. 3391–397, 2009.
252. D. Sharp, M. Awad, J. Warren, R. Wise, G. Vigliocco, and S. Scott, “The neural response to changing semantic and perceptual complexity during language processing,” *Human Brain Mapping*, vol. 31, no. 3, pp. 365–377, 2009.
253. C. Summerfield, T. Egner, J. Mangels, and J. Hirsch, “Mistaking a house for a face: Neural correlates of misperception in healthy humans,” *Cerebral Cortex*, vol. 16, pp. 500–508, 2006.
254. V. Marian, Y. Shildkrot, H. Blumenfeld, M. Kaushanskaya, Y. Faroqi-Shah, and J. Hirsch, “Cortical activation during word processing in late bilinguals: Similarities and differences as revealed by functional magnetic resonance imaging,” *Journal of Clinical and Experimental Neuropsychology*, vol. 29, no. 3, pp. 247–265, 1996.
255. F. Crick and C. Asanuma, “Certain aspects of the anatomy and physiology of the cerebral cortex,” pp. 333–371, 1986.
256. E. Bullmore and O. Sporns, “Complex brain networks: graph theoretical analysis of structural and functional systems,” *Nature Reviews Neuroscience*, vol. 10, pp. 186–198, 2009.
257. V. Braitenberg and A. Schütz, *Cortex: Statistics and Geometry of Neuronal Connectivity*. Berlin: Springer-Verlag, 1998.
258. B. Hellwig, “A quantitative analysis of the local connectivity between pyramidal neurons in layers 2/3 of the rat visual cortex,” *Biol. Cybern.*, vol. 82, pp. 111–121, 2000.
259. B. B. Averbeck and M. Seo, “The statistical neuroanatomy of frontal networks in the macaque,” *PLoS Comput. Biol.*, vol. 4, no. 4, 2008.
260. M. Abeles, *Corticonics: Neural Circuits of the Cerebral Cortex*. Cambridge University Press, 1991.
261. D. Watts and S. Strogatz, “Collective dynamics of ’small-world’ networks,” *Nature*, vol. 393, p. 2440442, 1998.
262. B. Ermentrout, “Neural networks as spatio-temporal patternforming systems,” *Rep. Prog. Phys.*, vol. 61, pp. 353–430, 2004.
263. B. Bollobas, *Random Graphs. Cambridge Studies in Advanced Mathematics 2nd Ed.* Cambridge UK: Camb Unridge University Press, 2001.
264. S. Strogatz, “Exploring complex networks,” *Nature*, vol. 410, pp. 268–276, 2001.
265. D. Watts, *Small Worlds*. Princeton, NJ.: Princeton University Press, 1999.
266. O. Sporns, *Graph theory methods for the analysis of neural connectivity patterns*. Boston, MA: Klwer, 2002.
267. X. Wang and G. Chen, “Complex networks: small-world, scale-free and beyond,” *IEEE Trans Circuits Syst*, vol. 31, no. 6-20, 2003.
268. O. Sporns and J. Zwi, “The small world of the cerebral cortex,” *NeuroInformatics*, vol. 4, pp. 145–162, 2004.
269. A.-L. Barabsi and R. Albert, “Emergence of scaling in random networks,” *Science*, vol. 286, no. 5439, pp. 509–512, 1999.
270. S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. Hwang, “Complex networks: Structure and dynamics,” *Physics Reports*, vol. 424, pp. 175–308, February 2006.
271. J. A. S. Kelso, J. P. Scholz, and G. Schöner, “Nonequilibrium phase transitions in coordinated biological motion: critical fluctuations,” *Physics Letters A*, vol. 118, no. 6, pp. 279 – 284, 1986.
272. J. Kelso, *Fluctuations in the coordination dynamics of brain and behavior*. Singapore: World Scientific, 2000.
273. S. Bressler and E. Tognoli, “Operational principles of neurocognitive networks,” *Int. J. Psychophysiol.*, vol. 60, pp. 139–148, 2006.
274. O. Sporns, “Connectome,” *Scholarpedia*, vol. 5, no. 2, p. 5584, 2010.
275. V. M. Eguíluz, D. Chialvo, G. Cecchi, M. Baliki, and A. Apkarian, “Scale-free brain functional networks,” *Phys. Rev. Lett.*, vol. 94, p. 018102, Jan 2005.
276. O. Sporns, “Small-world connectivity, motif composition, and complexity of fractal neuronal connections,” *Biosystems*, vol. 85, no. 1, pp. 55 – 64, 2006. Dedicated to the Memory of Ray Paton.

277. K. Oshio et al., *Database of Synaptic Connectivity of C. elegans for Computation*, 2003.
278. J. Szentagothai and P. Erdi, "Self-organization in the nervous system," *J. Social Biol. Struct.*, vol. 12, pp. 367–384, 1989.
279. J. Fuster, *Cortex and Mind Unifying Cognition.* Oxford University Press, 2005.
280. C. U. Smith, "Does history repeat itself? cortical columns: 3. a cortex of columns," *Cortex*, vol. In Press, Corrected Proof, pp. –, 2008.
281. V. Mountcastle, "The columnar organization of the cerebral cortex," *Brain*, vol. 120, pp. 701–722, 1997.
282. M. F. Casanova, *Neocortical Modularity and the Cell Minicolumn.* New York: Nova Biomedical Books, 2005.
283. J. Horton and D. Adams, "The cortical column: a structure without a function," *Phil. Trans. R. Soc. B*, vol. 360, no. 1456, pp. 837–862, 2005.
284. G. Blasdel and G. Salama, "Voltage-sensitive dyes reveal a modular organization in monkey striate cortex," *Nature*, vol. 321, pp. 579–585, 1986.
285. P. Rakic, "Radial versus tangential migration of neuronal clones in the developing cerebral cortex," *Proc. Natl Acad. USA Sci.*, vol. 92, no. 11, pp. 323–327, 1995.
286. U. Alon, "Network motifs: theory and experimental approaches," *Nat Rev Genet*, vol. 8, no. 6, pp. 450–61, 2007.
287. V. B. Mountcastle, "The neural mechanisms of cognitive functions can now be studied directly," *Trends in Neurosciences*, vol. 9, pp. 505 – 508, 1986.
288. I. Prigogine, *Structure, Dissipation and Life.Theoretical Physics and Biology, edited by P. Marois.* Amsterdam: Noth-Holland Publishing, 1969.
289. A. Turing, "Between genotype and phenotype: protein chaperones and evolvability," *Philosophical Transactions of The Royal Society of London, series B*, vol. 237, p. 3772, 1952.
290. S. Dehaene and L. Naccache, "Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework," *Cognition*, vol. 79, pp. 1–37, 2001.
291. G. Werner, "Brain dynamics across levels of organization," *Journal of Physiology (Paris)*, vol. 101, pp. 273–279, 2007.
292. W. Freeman, "The wave packet: an action potential for the 21st century," *Integr. Neurosci.*, vol. 2, no. 1, pp. 3–30, 2007.
293. W. Freeman and B. van Dijk, "Spatial patterns of visual cortical fast eeg during conditioned reflex in a rhesus monkey," *Brain Research*, vol. 422, pp. 267–276, 1987.
294. J. Barrie and et al., "Spatiotemporal analysis of prepyriform, visual, auditory, and somesthetic surface eegs in trained rabbits," *Journal of Neurophysiology*, vol. 76, no. 1, pp. 520–539, 1996.
295. Y. Oka et al., "Odorant receptor map in the mouse olfactory bulb: In vivo sensitivity and specificity of receptor-defined glomeruli," *Neuron*, vol. 52, no. 5, pp. 857–869, 2006.
296. S. Bressler, *The formation of global neurocognitive state.* New York: Springer, 2007.
297. T. Koenig et al., "Decreased eeg synchronization in alzheimer's disease and mild cognitive impairment," *Neurobiology of Aging*, vol. 26, no. 2, pp. 165–171, 2005.
298. J. P. Welsh et al., "Is autism due to brain desynchronization?," *Int J Dev Neurosci.*, vol. 23, no. 2-3, pp. 253–63, 2005.
299. T. Huntsberger and R. Tunstel, E. Kozma, *Onboard learning strategies for planetary surface rovers.* San Antonio, TX: TCI Press, 2006.
300. R. Kozma and W. Freeman, "A possible mechanism for intermittent oscillations in the kiii model of dynamic memories - the case study of olfaction," 1999.
301. R. Kozma and W. Freeman, "Chaotic resonance: Methods and applications for robust classification of noisy and variable patterns," *Int. J. Bifurc. Chaos*, vol. 11, no. 6, pp. 2307–2322, 2001.
302. R. Kozma, W. Freeman, and P. Erdi, "The kiv model nonlinear spatio-temporal dynamics of the primordial vertebrate forebrain," *Neurocomputing*, vol. 52, pp. 819–826, 2003.
303. S. Muthu, R. Kozma, and W. Freeman, "Applying kiv dynamic neural network model for real time navigation by mobile robot emma," in *Proceedings on IEEE International Joint Conference on Neural Networks*, (Delhi, India), 2003.

304. D. Harter and R. Kozma, “Chaotic neurodynamics for autonomous agents,” *IEEE Trans Neural Netw*, vol. 16, no. 3, pp. 565–579, 2005.
305. E. Kant, *The Critique of Pure Reason.* 1781.
306. T. Polger, *Philosophy of Mathematics and Natural Science.* Princeton: Princeton University Press, 1949.
307. C. Hull, *A behavior system.* Yale University Press, 1952.
308. J. O’Keefe and L. Nadel, *The Hippocampus as a Cognitive Map.* Oxford University Press, 1978.
309. M. J. Milford, *Robot Navigation from Nature.* Springer, 2008.
310. E. Tolman, “Cognitive maps in rats and men,” *Psychological Review*, vol. 55, no. 4, pp. 189–208, 1948.
311. A. Redish, “The hippocampal debate: Are we asking the right questions?,” *Behavioral Brain Research*, vol. 127, no. 935, 2001.
312. R. Muller, J. Kubie, and J. Ranck, “Spatial firing patterns of hippocampal complex-spike cells in a fixed environment,” *The Journal of Neuroscience*, vol. 7, pp. 1935–1950, 1987.
313. M. Fyhn, S. Molden, M. Witter, E. Moser, and M.-B. Moser, “Spatial representation in the entorhinal cortex,” *Science*, vol. 305, pp. 1258–64, 2004.
314. T. Hafting, M. Fyhn, S. Molden, M.-B. Moser, and E. Moser, “Microstructure of a spatial map in the entorhinal cortex,” *Nature*, vol. 436, pp. 801–6, 2005.
315. E. Moser, E. Kropff, and M. Moser, “Place cells, grid cells, and the brain’s spatial representation system,” *Annual review of neuroscience*, vol. 31, pp. 69–89, 2008.
316. T. Solstad, E. Moser, and G. Einevoll, “From grid cells to place cells: A mathematical model,” *Hippocampus*, vol. 16, pp. 1026–1031, 2006.
317. H. Eichenbaum, *The Cognitive Neuroscience of Memory: An Introduction.* Oxford University Press, 2002.
318. J. Adámek, H. P. Gumm, and V. Trnková, “Presentation of Set Functors: A Coalgebraic Perspective,” *Journal of Logic and Computation*, 2009.
319. D. Touretzky and A. Redish, “Theory of rodent navigation based on interacting representations of space,” *Hippocampus*, vol. 6, no. 3, pp. 247–70, 1996.
320. N. Burgess, M. Recce, and J. O’Keefe, “A model of hippocampal function. neural networks,” *Neural networks*, vol. 7, pp. 1065–1081, 1994.
321. A. Arleo and W. Gerstner, “Spatial cognition and neuro-mimetic navigation: a model of hippocampal place cell activity,” *Biological Cybernetics*, vol. 83, pp. 287–299, 2000.
322. S. Jung, MW.and Wiener and B. McNaughton, “Comparison of spatial firing characteristics of units in dorsal and ventral hippocampus of the rat,” *J. Neurosci.*, vol. 14, pp. 7347–56, 1994.
323. K. B. Kjelstrup et al., “Very large place fields at the ventral pole of the hippocampal ca3 area,” *oc. Neurosci. Abstr.*, vol. 33, no. 93, p. 1, 2007.
324. R. A. Wilson and F. C. Keil, eds., *The MIT Encyclopedia of the Cognitive Sciences.* Cambridge, Massachusetts: The MIT Press, 1999.
325. W. Scoville and B. Milner, “Loss of recent memory after bilateral hippocampal lesions,” *J. Neurol. Neurosurg. Psychiatry*, vol. 20, pp. 11–21, 1957.