Philosophical issues in psychiatry
Copenhagen, May 25-26, 2006

Entoptic vision and physicalist emergentism

Jean Petitot
CREA, Ecole Polytechnique, Paris
petitot@poly.polytechnique.fr

INTRODUCTION

We will be concerned with neurally implemented mathematical models of visual hallucinations. They raise the philosophical issue of neural reductionism in cognitive sciences.

Classically, a reductionist thesis posits that complex high level phenomena, structures and processes can be reduced, as far as their scientific explanation is concerned, to underlying lower level phenomena, structures and processes. The most paradigmatic and best-investigated example is the reduction of macroscopic thermodynamics to microscopic molecular and atomic movements (temperature = mean kinetic energy per degree of freedom, etc.). Let us begin with some conceptual precisions.

1. Reductionism can be a particular scientific thesis concerning a specific scientific theory: it is precisely the case with the reduction of macro thermodynamics to micro statistical mechanics. But it can be also a general metaphysical claim on the ultimate nature of reality. That's the case with different forms of monism. Idealist monism posits the universal reducibility of reality to mind while materialist monism posits the universal causal reducibility of reality to matter and energy. In this paper we will be concerned only with scientific reductionism.

2. Scientific reductionism can be *objective* or *methodological*. It is objective when it concerns explanations in terms of primitive objects (atoms, neurons, etc.) and methodological when it concerns deflationist nominalist explanations (Occam's razor). To day, a very important debate in cognitive science has to do with the eliminability of "mental" or "conscious" concepts and their reduction to neural concepts (see e.g. the Dennett/Chalmers controversy). In this paper we will be concerned with "objective" reductionism.

3. In our narrow, scientific, empirical, and objective sense, reductionism concerns mainly *complex systems* possessing at least two levels of reality: a micro underlying level where a great number of elementary units are in interaction and a macro emergent one where *macro self-organized structures* emerge. In such a perspective, reductionism is inseparable from converse concepts such as "emergence", "supervenience" or "functionalism". Functionalism means that

macro structures having a functional role can exist only if they are materially implemented in an underlying material substrate, but are at the same time, as functionally meaningful structures, largely *independent* of the fine grained physical properties of the substrate they are implemented in. The paradigmatic example is the opposition software/hardware in computer sciences (see philosophers like Putnam, Fodor, Pylyshyn, etc.) but functionalism also applies in natural sciences where it is an aspect of emergence.

4. There is a general agreement on the fact that in complex systems having different levels of reality at different scales, there exist collective behaviors ruled by laws that are not the laws of the micro underlying level. It is the case for critical phenomena, percolation, self-organized criticality, reaction-diffusion equations, dissipative structures, turbulence, cellular automata, neural networks, ant colonies, swarms, stock markets, etc. According to one's conception of laws, one can develop different conceptions of this empirical fact.

(i) *Eliminativism and epiphenomenalism*: laws being only empirical regularities lacking any objective (and a fortiori ontological) content (Hume's empiricist thesis), emerging structures are purely epiphenomenal and can be scientifically eliminated "salva veritate".

(ii) *Holistic realism* (it is the converse position): laws being real in the ontological sense, the emerging level possesses an ontological reality and cannot therefore be reduced.

(iii) *Causal reductionism and objective emergentism*: laws being objective, that is at the same time empirically grounded and mathematically formalized, the emerging level has no ontological content but is nevertheless much more than a simple empirical regularity. It is *causally* reducible to complex interactions at the micro underlying level but it shares nevertheless some empirical and theoretical *autonomy*.

We will be concerned here with this third type of reductionism.

5. The main difficulty that has to be tackled in such a perspective is the relation between causal reduction and theoretical autonomy. *Mathematics* play here the fundamental role. Indeed, the formal equivalent to causal reduction is *mathematical deduction*. But deducibility is a syntactic property and doesn't entail any evident *conceptual* derivation (it is for that very reason that mathematics constitute an authentically "synthetic" knowledge even if their proofs are "analytical"). Therefore the fact that the structures and properties of the macro level can be mathematically deduced from the micro one doesn't mean that the representational content of its conceptual description can be reduced to the representational content of the micro level. Causal reduction paralleling mathematical deduction is not a *conceptual* reduction. Emerging critical phenomena have some measure of autonomy and belong to autonomous levels of reality.

6. In cognitive sciences, reductionism concerns in particular the reducibility of *mental* states, representations, contents, structures, events and processes to *neural* activity. The most difficult challenge is that of the reduction of *consciousness*. For monist materialists such as Dan Dennett, mind is epiphenomenal and without any causal efficiency, and must be reduced to neurobiology. Conversely, for dualist philosophers such as David Chalmers, phenomenal consciousness is irreducible and there will always remain an "explanatory gap".

Let us recall very briefly the classical opposition introduced by Ned Block between access consciousness (object oriented awareness, intentionality, representational contents) and phenomenal consciousness (first person experience, subjective evidence, qualia). It is phenomenal consciousness that raises the main problem for reductionism and many philosophers consider indeed that it yields refutations of physicalism.

As the problem is extremely subtle, we will present a very specific example, which can be treated in details.

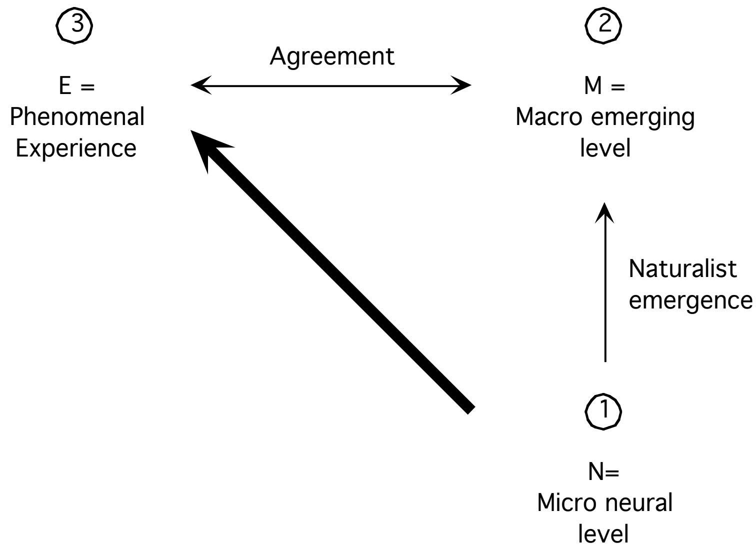
1. PHENOMENAL EXPERIENCE AND MATHEMATICAL REDUCTIONISM

One of the main difficulties encountered in clarifying the concepts of emergence and reduction in cognitive science is that they entangle two completely different problems:

1. A problem of emergence of macro-structures out of physiological mechanisms processed in an underlying micro neural substrate *N*.
2. A problem of matching or agreement between some objective emergent macro-structures and some subjective phenomenal experience *E*.[1]

If we seek to correlate directly *E* with *N* we will run into great hurdles since the emergence problem is entangled with the opposition "third person objectivity" / "first person subjectivity". We will argue that a *naturalist emergentist* answer to the problem of existence of "Neural Correlates of Consciousness" (NCC) can be philosophically clarified only if the phenomenal emergence $N \to E$ is "factorized" (as mathematicians would say) through a macro natural level *M* according to the following diagram:

[1] See for instance Noë's and Thompson's target paper against Chalmers « Are there Neural Correlates of Consciousness? » in *Journal of Conciousness Studies*, 11, 1, 2004.



The factorization disentangles the problem:

1. The "vertical" emergence **1** $\rightarrow$ **2** belongs to the naturalist domain. The problem it raises is not a problem of naturalization but a problem of *deducing* **2** from **1**. As we saw in the introduction, it is a problem of mathematical reductionism, functionalism, ontological dependency, etc.

2. The "horizontal" agreement **1** $\leftrightarrow$ **2** raises a completely different problem of matching representational contents, of naturalization, etc. but not a problem of emergence.

As we will focus on problems of visual perception concerning the phenomenal experience of spatial forms, we will call the macro naturalist level **2** a *morphological* level.

The factorization of **1** $\rightarrow$ **3** through **2**: **1** $\rightarrow$ **2** $\rightarrow$ **3** reveals the methodological obstruction to any clarification of the problem:

1. the micro neurophysical level **1** is well known thanks to findings in neurobiology and psychophysics;

2. the phenomenal level **3** is well described thanks to phenomenological descriptive eidetics;

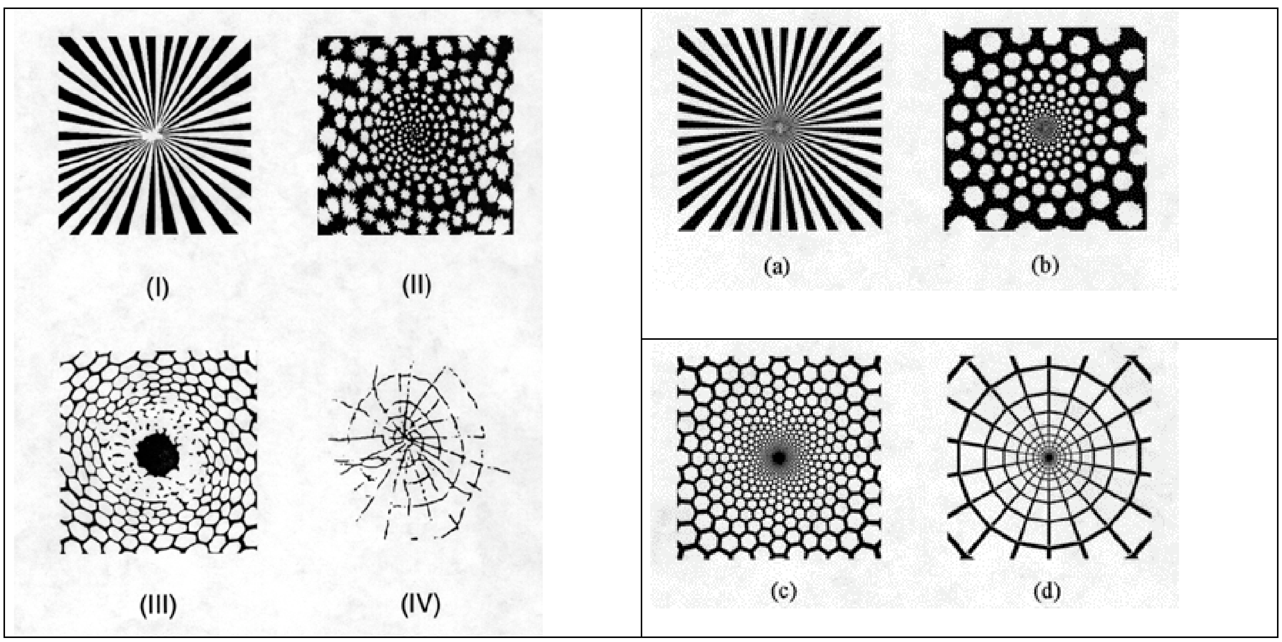
3. but the macro morphological level **2** is largely unknown: *it cannot be conceptually derived directly from* **1**. It can only be *mediately mathematically* deduced.

Therefore, it depends entirely on deep mathematical achievements. In the absence of **2**, arguments concerning a supposed emergence **1** $\rightarrow$ **3** become metaphysical and scientifically undecidable. But in presence of **2** they depend hic et nunc on the contingent progress of mathematics.

It is the main problem of any anti-reductionist thesis positing that "it will be never possible to reduce this to that". For physicalism must be a mathematical physicalism and the claim that something cannot be proved makes sense only for internal limitations of mathematics and cannot therefore be relevant for physical modeling.

What is called the *explanatory gap* between **1** (neurophysiology) and **3** (phenomenal experience) depends itself entirely on **2**. Indeed, the emergence **1** $\rightarrow$ **2** is of the very same type as the one that is well known in statistical physics (thermodynamics, phase transitions and critical phenomena). The macro morphological structures at level **2** are nomologically deducible from the micro interactions at level **1**. If **2** is lacking, as there cannot be a *direct content matching* between **1** and **3** for obvious reasons, there exists an explanatory gap. But the stronger the mathematical deduction **1** $\rightarrow$ **2**, the weaker the explanatory gap. And we see no principled reasons to reject the regulative idea that, at the limit, **2** would become sufficiently strong to fill the gap.

We will focus on an example and comment on its foundational issues. It is a very limited but very striking example that concerns some recent works on *entoptic vision* and visual hallucinations. It can be summarized by the following comparison:



These morphologies represent spontaneously hallucinated geometric visual patterns:
<u>Left</u>: we have an example of **3** (phenomenological report) since these are drawings of visual hallucinations *E* experienced by subjects.
<u>Right</u>: we have an extraordinary example of mathematical models of morphological macro structures *M* since these are models very recently constructed by mathematicians.

I will try to present in not too technical a way the key result that these morphological structures *M are mathematically deducible from the encoding of a neural functional architecture* *in the synaptic weights of a neural net*. They provide therefore a striking example of mathematical deduction **2**.

This non-intuitive and non-trivial result has to be proved and cannot, by any means, be anticipated purely conceptually.

2. ENTOPTIC VISION AND PHENOMENAL CONSCIOUSNESS

2.1. Entoptic vision and hallucinations

Entoptic vision concerns some geometrical patterns who are perceived after exposures to a violent flickering light, absorptions of substances such as mescaline, LSD, psilocybin, ketamin, some alkaloids (peyote), or near-death experiences.

I have no personal experience of the matter but it is known that subjects see spontaneously and vividly typical forms: tunnels and funnels, spirals, lattices (honeycombs, triangles), cobwebs, etc. These typical forms can operate on the drawings of any type of objects as we find it in the paintings from Indian Mexican tribes.

> « Such visual imagery is dynamic and the illusory contours usually explode from the center of gaze to the periphery, appearing initially in black and white before bright colors take over, and eventually pulsate and rotate in time as the experience progresses » (Yves Frégnac, 2003).

These illusory forms were already classified a long time ago (1928) by the great neurophysiologist Heinrich Klüver (1897-1979) who provided many clinical reports on them. Klüver was a student of Max Wertheimer and introduced Gestalt psychology in the United States. He called the most typical forms *planforms*.

2.2. The status of the example

The example is particularly interesting because it yields a case of pure phenomenal consciousness.

1. It is clearly a qualitative and phenomenal experience.
2. It is clearly subjective.
3. It is neither conceptual nor linguistic but is the object of graphical reports.
4. It is not a mode of access consciousness.
5. It is without any intentional, representational and semantic content.
6. It is not embedded in a complex "narratively" organized stream of consciousness.
7. It is not involved in sensory-motor loops.

It is therefore particularly adapted for probing *a possible matching between a phenomenal content and a neural dynamics*.

2.3. The NIMH program

Before coming to the mathematical deduction **1** $\rightarrow$ **2**, let us mention programs at the National Institute of Mental Health (NIMH) aiming at probing the neuroreceptors with varied substances. In the receptor space each substance shifts the balance of activity of the brain away from the origin, by a vector representing the profile of binding affinities at different receptors.

> « In a brain-centered reference frame, the origin is based on absolute levels of activity at each receptor population. The state of the brain is constantly on the move. We can think of it as a complex dynamical system, in which the trajectory follows high-dimensional orbits, and switches among many "attractors". »

> « In this dynamic reference frame, drugs will create a perturbation along the binding vector, thereby pushing the system into a new attractor.» (Thomas Ray, 2004)

3. SPONTANEOUS EMERGENCE OF GEOMETRIC VISUAL PATTERNS

3.1. Reference papers

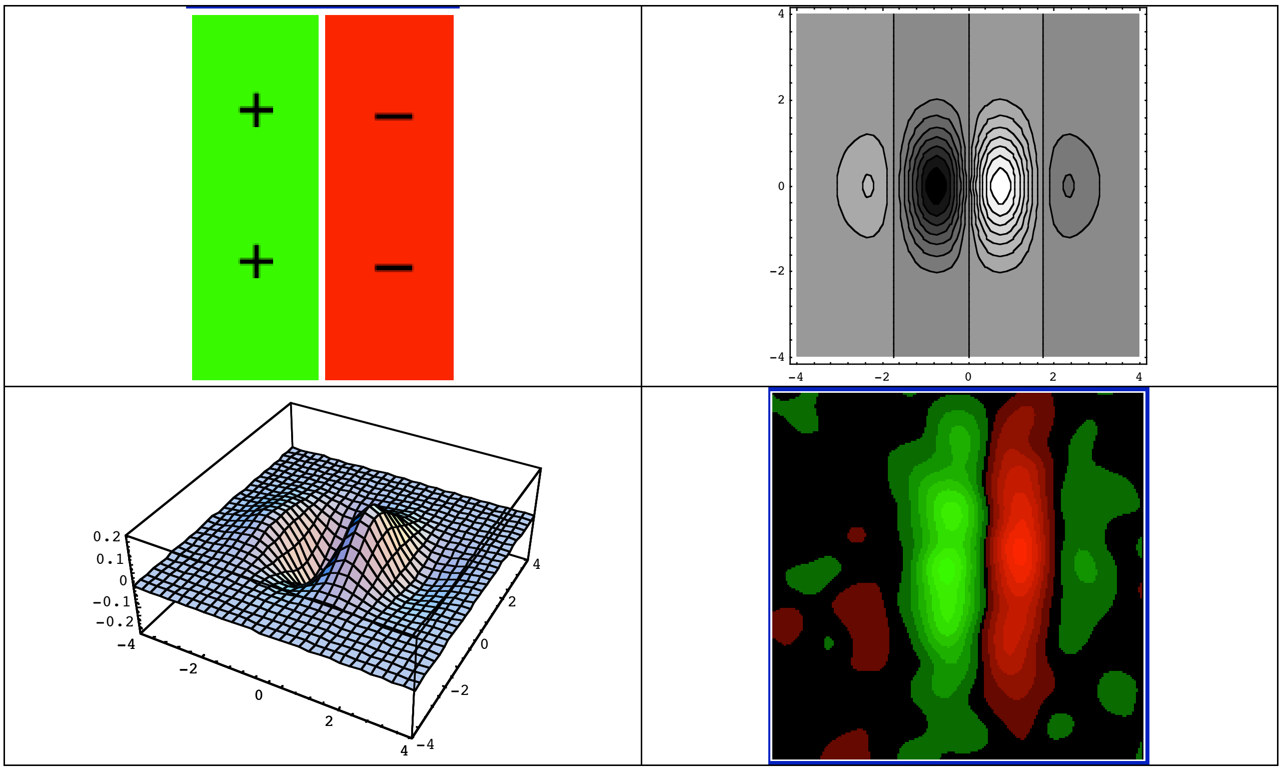
The first paper on the subject was the one by Ermentrout and Cowan (1979): « A mathematical theory of visual hallucinations ». Bard Ermentrout is one of the major specialists of synchronization in networks of oscillators. Jack Cowan is a specialist of neural networks (Wilson-Cowan equations).

Recently, the subject was completely revisited by Paul Bressloff, Jack Cowan, Martin Golubitsky, Peter Thomas, and Matthew Wiener: « Geometric visual hallucinations, Euclidean symmetry and the functional architecture of striate cortex » (2001). Paul Bressloff is a specialist of models of vision and Martin Golubitsky one of the major specialists of bifurcations in dynamical systems. See also the very recent Bressloff's & Cowan's paper: « The functional geometry of local and horizontal connections in a model of V1 » (2003).

3.2. The simple neurons of V1

The receptive field (RF) of a visual neuron is the domain of the retina to which it is connected through the neural connections of the retino-geniculo-cortical pathways. It is decomposed into ON zones (lighting induces a depolarization and a stimulation of the rest spiking response) and OFF zones (lighting induces an hyper-polarization and an inhibition of the rest spiking response).

Sophisticated techniques enable to record the *level curves* of the receptive profiles (G. De Angelis, Berkeley). For the receptive profiles of the simple cells of V1 the level curves can be modeled by third order derivatives of Gaussians (similar to Gabor functions).



Due to the structure of its RP, each simple cell of V1 (which operates as a filter by convolution on the optical signal) detects a preferential orientation, which can be assimilated with its RF-content. *We define therefore a first (low) level of abstraction* by saying that these neurons detect, at a certain scale, pairs (a, p) of a spatial (retinal) position a and of a local orientation p at a. Pairs (a, p) are called *contact elements* in differential geometry.

3.3. Neural nets and Hopfield equations

Now, we must do a little bit of mathematics if we want to understand the argumentation. I hope that readers are not too phobic against mathematics but it is impossible to discuss reductionism and emergenge without looking precisely at mathematical models.

The authors work in the product (called a fibration)

$$\pi : V = R \times \mathbf{S}^1 \rightarrow R$$

where the base space R is the plane of the retina and the fiber $\mathbf{S}^1$ the space of orientations p. We choose in the total space V modeling V1 coordinates (a, φ) with a = retinal position and φ = the angle of the orientation p. Let $z(a, \varphi, t)$ be the activity of V1 at the neuron (a, φ) at time t. If we know the real-valued function $z(a, \varphi, t)$, we know the evolution of the activity of V1. According to the physicalist scheme, we look for the PDE (partial differential equation) governing the evolution of z. This equation must express the temporal derivative of z as a function of z.

We start with standard Hopfield equations for a discrete neural net. Let u_i , $i = 1, \ldots, N$ be formal neurons with activity $z_i(t)$. If time and space are discrete, standard Hopfield equations are (*local rules* of interaction) enabling to compute the activity $z_i(t)$ of the neuron u_i from its interactions with other neurons u_j and from the input. The interactions are governed by what are called "synaptic weights" that is weights of the connection between neural units. We get

$$z_i(t+1) = \sum_{j=1}^{j=N} w_{ij}\sigma\left(z_j(t)\right) + h_i(t)$$

where σ is a *non-linear* gain function (with $\sigma(0) = 0$), w_{ij} the weight of u_i the connection between the neural units u_i and u_j, and h the external input. We see that the activity of the unit u_i at time $t+1$ is essentially a non-linear weighted sum of the activity at time t of the units to which it is connected.

To find the continuous limit when the temporal step Δt evolves from $\Delta t = (t+1) - t = 1$ to an infinitesimal step dt, one can put $z_i(t+1) = F(\ldots)$ into the form

$$z_i(t+1) - z_i(t) \,/\, ((t+1) - t) = -z_i(t) + F(\ldots)$$

So, at the limit, when time is continuous and space discrete, we get a system of N ordinary differential equations:

$$\frac{d z_i(t)}{d t} = -z_i(t) + \sum_{j=1}^{j=N} w_{ij}\sigma\left(z_j(t)\right) + h_i(t)$$

On the other hand, if space is also continuous, the indexes i become spatial variables v and we get a partial differential equation:

$$\frac{\partial z(v,t)}{\partial t} = -z(v,t) + \int_V w\langle v \,|\, v'\rangle \sigma\left(z(v',t)\right) dv' + h(v,t)$$

$w(v \mid v')$ being the weight of the connection between the neuron v and the neuron v'.

The authors use the following Hopfield equation:

$$\frac{\partial z(a,\varphi,t)}{\partial t} =$$
$$-\alpha z(a,\varphi,t)+\frac{\mu}{\pi}\int_0^{\pi}\int_R w\langle a,\varphi | a',\varphi'\rangle \sigma(z(a',\varphi',t))da'd\varphi' + h(a,\varphi,t)$$

where $w(a, \varphi \mid a', \varphi')$ is the weight of the connection between the neuron $v = (a, \varphi)$ and the neuron $v' = (a', \varphi')$, α a parameter of decay and μ a parameter of excitability of V1. The μ parameter is the simplest example of localization of a substance in a receptor space and of the fact that the substance shifts the balance of activity of the brain away from the origin.

We want now to deduce some emergent global properties of such a typically complex system.

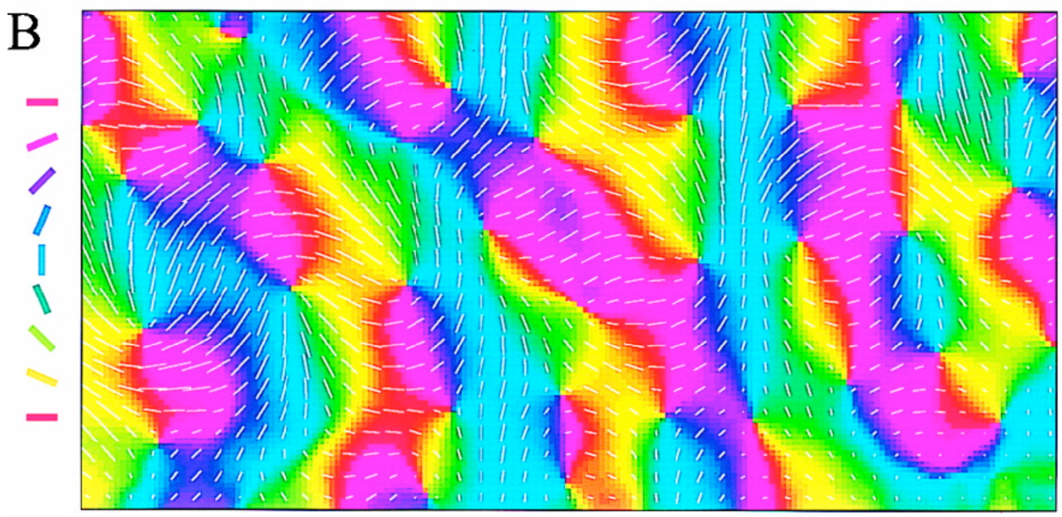
To explain how the morphological macro structures M (**2**) can emerge from the dynamics of the neural substrate N = V1 (**1**), the key point is that the *functional architecture of V1 can be encoded in the synaptic weights* $w(a, \varphi \mid a', \varphi')$.

3.4. The functional architecture of V1

Many experimental results are now available concerning the functional architecture (FA) of V1. They show that V1 is organized through a *double* structure:

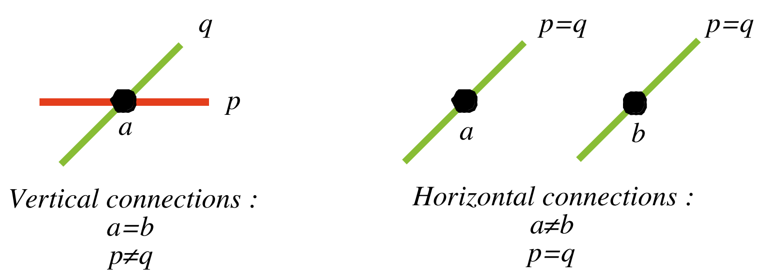
1. *A "vertical" structure* given by the retino-geniculo-cortical retinotopy. The *hypercolumns* (Hubel and Wiesel) are micro neural modules which associate retinotopically to each position a of the retina R a full exemplar P_a of the space $P = \mathbf{S}^1$ of the orientations p at a. This means that the fribration structure of V = V1 is neurally implemented.
2. *An "horizontal" structure* given by the horizontal cortico-cortical connections which connect neurons in different hypercolumns.

This double structure gives rise to geometrically well organized structures called *pinwheels*. V1 is reticulated by a network of singular points which are the centers of pinwheels. Locally, around these singular points all the orientations are represented by the rays of a "wheel" and the local wheels are glued together in a global structure. The relation between pinwheels rays (colors) and preferred orientations (white strokes) is shown in the following figure taken from Shmuel and Grinvald 2000 and representing the FA of the area V1 of a cat.



The fundamental feature of this functional architecture is that cortico-cortical connections connect neurons of the *same orientation* in different hypercolumns. This means that the system implements what is called in differential geometry a *parallel transport* and is able to know, for b different from a, if the orientation q at b is *the same* as the orientation p at a.

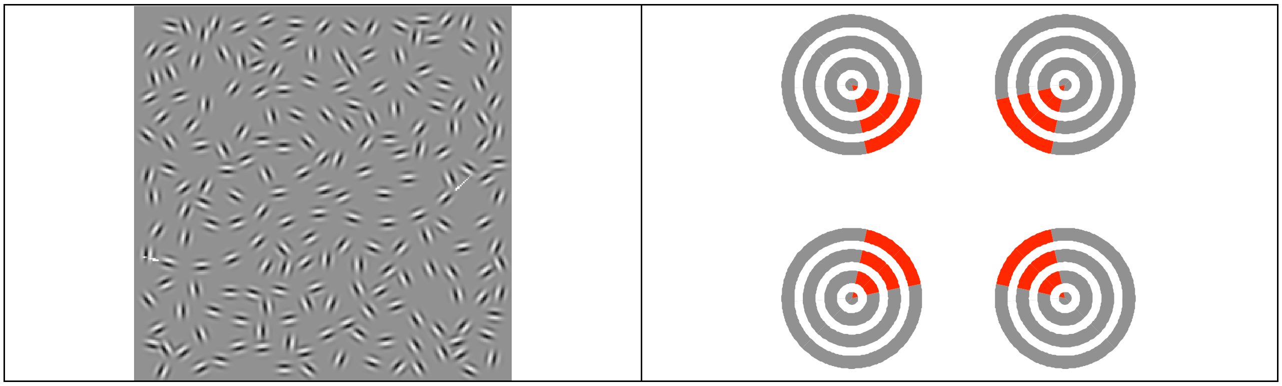
1. The retino-geniculo-cortical "vertical" connections give an internal meaning for the relations between (a, p) and (a, q) (different orientations p and q at the same point a).
2. The "horizontal" cortico-cortical connections give an internal meaning for the relations between (a, p) and (b, p) (same orientation p at different points a and b).



Moreover cortico-cortical connections connect neurons coding pairs (a, p) and (b, p) *such that p is the orientation of the axis ab*. As was summarized by one of the main specialist of the field:

> « The system of long-range horizontal connections can be summarized as preferentially linking neurons with co-oriented, co-axially aligned receptive fields ». (William Bosking)

This functional architecture of V1 is basic for visual perception. It explains the Gestalt principle of *good continuation* and the striking capacities of integration of contours and boundaries, even of illusory contours as in Kanizsa's well known experiments (figure below, right). For instance, it explains the pop-out of a curve of local contact elements (a_i, p_i) against a background of random elements (the concept of *association field* of Field, Hayes & Hess experiments, figure below, left)



3.5. The encoding of the functional architecture into the synaptic weights

Let us now encode this functional architecture in the weights $w(a, \varphi \mid a', \varphi')$ of the neural net V modeling V1. We must translate the geometrical properties of FA into formulae concerning $w(a, \varphi \mid a', \varphi')$.

♦ The local vertical connections inside a single hypercolumn yield a term:

$$w\langle a,\varphi|a',\varphi'\rangle = w_{ver}(\varphi-\varphi')\delta(a-a')$$

where δ is what is called a Dirac function and imposes $a = a'$ ($\delta(a - a') = 0$ if $a - a' \neq 0$, $\delta(a - a') = \infty$ if $a - a' = 0$, and the integral of δ is 1). This formula means that, in a single hypercolumn (condition $a = a'$), the connection between a neuron detecting the orientation φ and a neuron detecting the orientation φ' is weighted by a function $w_{ver}(\varphi - \varphi')$ depending only upon the difference of phases $\varphi - \varphi'$.

♦ The lateral horizontal connections between different hypercolumns yield a term:

$$w\langle a,\varphi|a',\varphi'\rangle = w_{hor}(a-a',\varphi)\delta(\varphi-\varphi')$$

where the factor $\delta(\varphi - \varphi')$ imposes $\varphi = \varphi'$ and expresses the fact that the horizontal cortico-cortical connections connect pairs (a, p) and (b, q) with $p = q$ and that their weight is only a function $w_{hor}(a - a', \varphi)$ of the vector $a - a'$ and the common orientation φ of p and q.

♦ Moreover, the fact that $p = q = ab$, that is the property of coaxiality, is expressed by the fact that the direction of the vector $a - a'$ is φ and that the function $w_{hor}(a - a', \varphi)$ depends only on the measure s of $a - a'$ with respect to the unit vector e_φ in the direction φ:

$$w_{hor}(a-a',\varphi) = w_{hor}(s)\delta(a-a'-se_\varphi)$$

We must emphasize the fact that the synaptic weights w are $E(2)$-*invariant* under the roto-translation group of motions of the plane $E(2)$ and that the PDE satisfies therefore an $E(2)$-symmetry ($E(2)$-equivariance).

3.6. Dynamically emerging morphologies and bifurcations

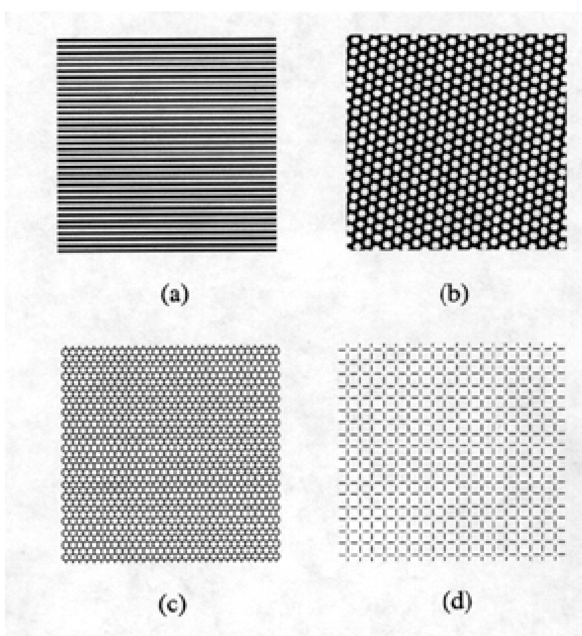
We are now ready to explain mathematically *why macro morphologies can spontaneously emerge in such a geometrically structured neural network*. We suppose that there exist no external input, that is $h = 0$. For $\mu = 0$, the rest state $z \equiv 0$ is trivially the state of the network and it is stable (you see nothing).

Now, the analysis of the PDE shows that, as the parameter μ increases, this initial activation state $z \equiv 0$ can become *unstable* and *bifurcate* spontaneously for some *critical values* μ_c of μ. The increasing of μ models an increasing of the excitability of V1 due to the action of some substances on the nuclei (locus coerulus, raphé) that produce neurotransmitters (serotonin, noradrenalin).

The new stable activation states present spatial patterns generated by an $E(2)$ *symmetry breaking*. The bifurcations can be analyzed using classical (but sophisticated) methods:

1. Linearization of the PDE near the solution $z \equiv 0$ and the critical value μ_c.
2. Spectral analysis of the linearized equation.
3. Computation of its eigenvectors (eigenmodes).

Here are some examples of eigenmodes in V.

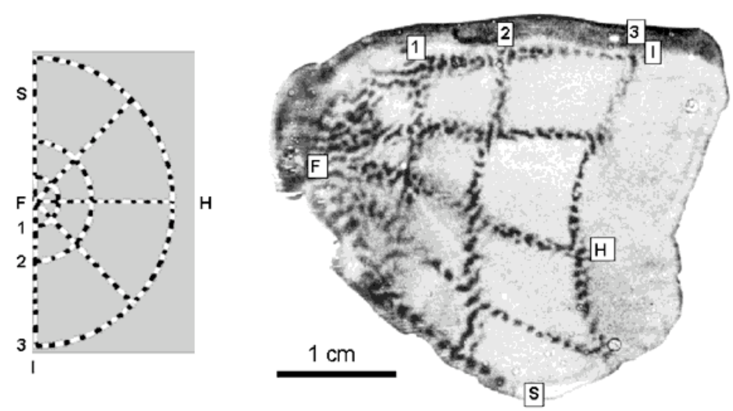


What is here of utmost importance is that the eigenmodes can be mathematically deduced from the Hopfield equations while, at the same time, they share some autonomy because they essentially depend on very abstract symmetry breakings independent of the fine structure of the network. In that sense mathematical deducibility is essentially different from theoretical (conceptual) physical reducibility.

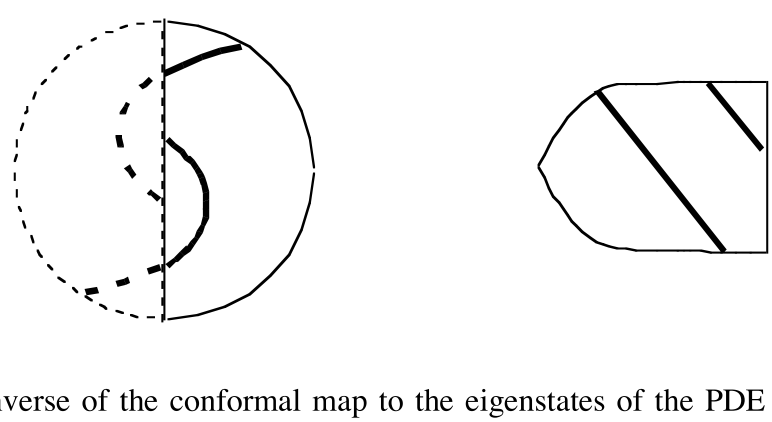
As we have already emphasized it, the situation is quite analogue to that found in statistical physics where critical behaviors can be classified, via the renormalization group, in universal classes independent of the specific fine-grained physical structure of the substrate.

3.7. The perception of emerging patterns as virtual retinal images

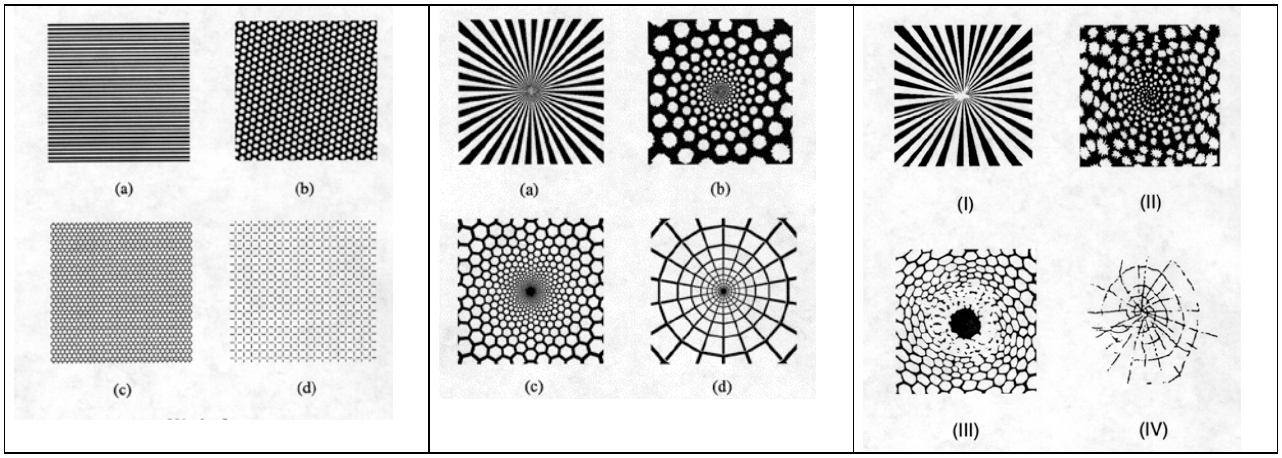
The last step is to reconstruct from the eigenmodes in V1, the corresponding virtual retinal images that would be hallucinated. For that, we must take into account the retinotopic *conformal_map* mapping the retina on V1. A good dipole model is for instance Log[(z+0.333)/(z+6.66)].



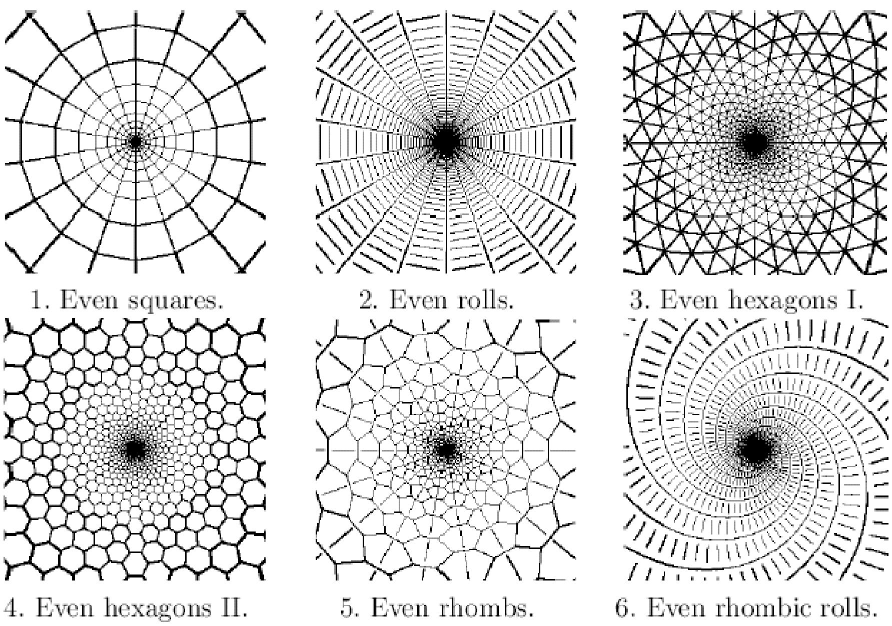
Lines in V1 correspond to spiral on the retina.

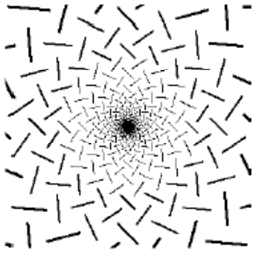


If we apply the inverse of the conformal map to the eigenstates of the PDE we get quite exact models of Klüver's planforms. *Klüver's planforms are isomorphic to eigenmodes of the* *bifurcated solutions of the neural network in the synaptic weights of which the functional architecture of V1 has been encoded.*

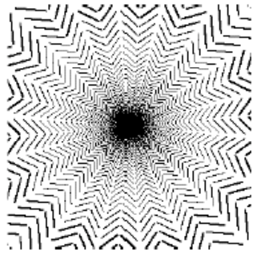


Many other eigenmodes can be deduced from this bifurcation scheme. Here are some examples:

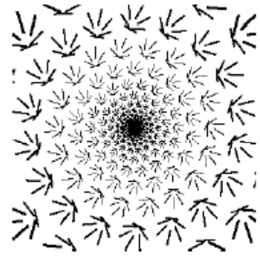




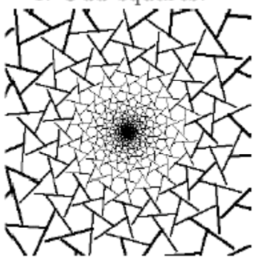
1. Odd squares.



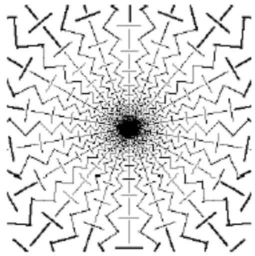
2. Odd rolls.



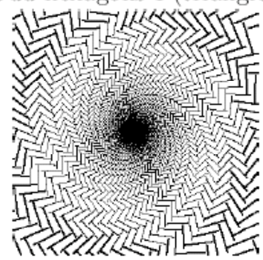
3. Odd hexagons I (triangles).



4. Odd hexagons II.



5. Odd rhombs.



6. Odd rhombic rolls.

<u>Annex for mathematicians</u>: The spectral analysis of the PDE is rather technical.

♦ After having linearized the PDE around the trivial solution $z \equiv 0$ we look for solutions of the form:

$$z(a, \varphi, t) = e^{\lambda t} z(a, \varphi).$$

When $\mu = 0$, $\lambda = -\alpha$ and the solution is $z(a, \varphi, t) = e^{-\alpha t} z(a, \varphi)$.

♦ The solutions are stationary only if $\lambda = 0$. Otherwise they decay ($\lambda < 0$) or diverge ($\lambda > 0$) exponentially. In that case saturation imposed by the non-linear gain function stabilizes them.

♦ We get an equation (for the eigenvalues λ) of the form:

$$\lambda z(a, \varphi) = -\alpha z(a, \varphi) + \sigma'(0)\mu \left[\int_0^\pi w_{ver}(\varphi - \varphi') z(a, \varphi') \frac{d\varphi'}{\pi} + \beta \int_R w_{hor}(a - a', \varphi) z(a', \varphi) da' \right]$$

where β is a constant measuring the relative strength of the vertical and horizontal connections.

♦ Using Fourier series of z, w_{ver} and w_{hor} for the periodic variable φ and Fourier transforms for the spatial variable a, and identifying the coefficients of the terms in the two sides of the equation we get *dispersion relations* of the form

$$\lambda = -\alpha + \mu F$$

where F is a (complicated) function of the Fourier coefficients.

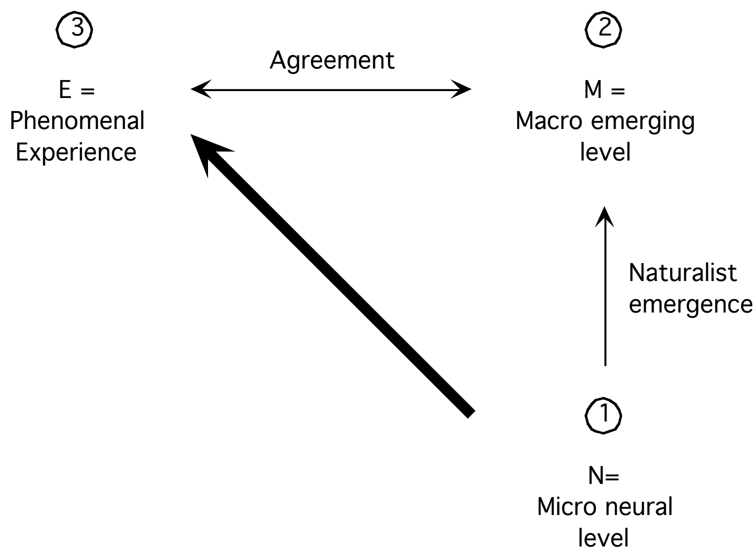
♦ For $\mu = 0$, $\lambda = -\alpha$ and the state decays exponentially to 0. When μ increases, λ will vanish for a certain $F(w^F{}_{loc,n}, w^F{}_{lat}(q))$ (where $w^F{}_{loc,n}$ and $w^F{}_{lat}(q)$ are Fourier coefficients) and a certain critical value μ_c.

♦ The bifurcation activates the corresponding terms in the Fourier series and transforms, which elicitates certain eigenmodes.

♦ The key point is that the *symmetries* imply strong constraints. It is the most technical part of the deduction.

4. PHILOSOPHICAL DISCUSSION

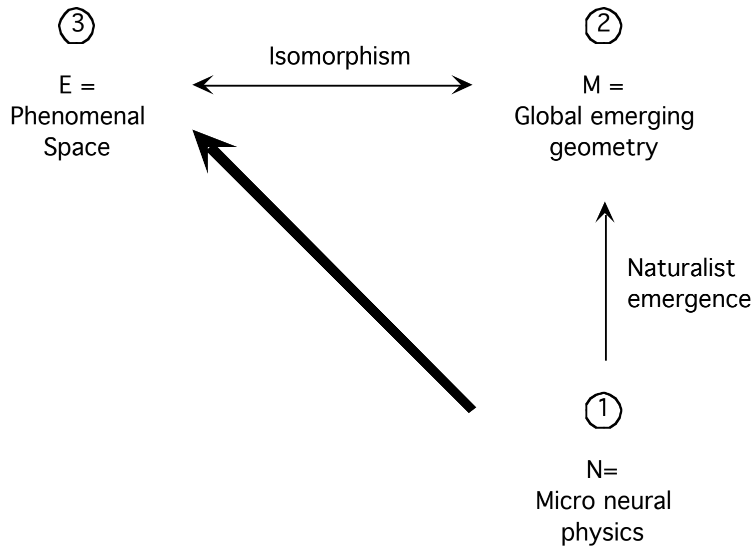
We have now at hand sufficient data for testing some philosophical issues. Let us first recall our initial factorization **1** → **2** → **3**:



4.1. The agreement 2 ↔ 3

In what concerns the agreement between the macro emergent level M (**2**) and the phenomenal experience E (**3**), we see that it is extremely strong, very much stronger than a mere correlation. It is in fact the strongest form of content matching since it is, at the limit, an *isomorphism.*

But we have to be very cautious here: contents are not *conceptual* contents but *geometric* non-conceptual contents. The isomorphism is a geometrical one that obtains between the geometrical content of an experienced intuitive space and the geometrical content of a mathematical formal space. We can therefore make the factorization **1** → **2** → **3** much more precise.



We see that the reductionism-emergentism game is not played between holistic phenomenal experience and local neurophysics but between global phenomenal experience, local physics, and emerging global *geometry*. The isomorphism is a matching *between morphologies that are geometrical contents*. It is not problematic as such. What is problematic is the matching of the *two types of spaces* where these morphological contents are embedded and framed.

We meet here what is for me the true hard problem of consciousness, from Kant's transcendental aesthetics to Poincaré and Husserl: what is the link between space as a phenomenal « form of intuition » and space as a geometrical construct.

Space is a *multisensorial* common format constitutive of *phenomenal* consciousness. Spatial representations are non-conceptual and non-propositional phenomenal forms. Space is the form of the phenomenality itself. It is *internal* and *subjective*, rooted in the internal temporality of first person experience, but at the same time it is the universal form of third person *external objectivity*. Moreover it is not only a phenomenal "form of intuition" but also a geometric entity which can be mathematized.

If we accept the identity of these two kinds of spatiality, then the matching **2 ↔ 3** becomes a very strong non conceptual *type-type identity* content matching: "eigenmodes" ↔ "planforms". If on the contrary we don't accept this identity, then the matching becomes an isomorphism between morphologies embedded in two different (phenomenal and geometric) ambient spaces (which is already a noteworthy result).

4.2. The emergence 1 → 2 as neural correlate

The interest of the mathematical model is to explain by means of a long mathematical deduction the morphological emergent level *M* from the physical activity of the underlying neural substrate N. In that case there is no explanatory gap (the problem is rather the phenomenality of space as we have just seen).

We meet here a probing example of *non-reductionist physicalism* satisfying the three constraints called by Jean-Michel Roy "attributive", "ontological", and "explanatory".

- ♦ M (**2**) belongs to N (**1**) due to the natural properties of N that are necessary and sufficient for its instantiation.
- ♦ **1** $\rightarrow$ **2** is stronger than supervenience but eventually weaker than type-type identity if the two modes of space are not identical.
- ♦ **2** is causally and mathematically reducible but manifests some measure of autonomy since its *geometry* is an emergent structure.

Moreover, in that case V1 provides a NCC (neural correlate of consciousness) in Chalmers' sense for the visual experience E. Our mathematicians have *proved* that N = V1 (**1**) is what is called a *bridge locus* for M (**2**), and through M, for E (**3**).

> « A NCC for content is a minimal neural representational system N such that the representation of a content in N is sufficient, under conditions C, for the representation of that content in consciousness. » (Chalmers 2000, p. 31)

Of course, it is well known that V1 doesn't correlate well with high-level perceptual structures possessing an intentional and semantic content (perception of objects). Higher areas (e.g. IT, inferior temporal) correlate much better. But it is not a problem here because V1 correlates nevertheless extremely well with the geometric morphologies we want to explain. And it is truly *minimal* since it implements minimally its functional architecture.

Now, the main difficulty is to define correctly what we consider to be the *representational content* of N = V1. It is here that the key concept of *functional architecture* becomes really unavoidable. We know the local physical representational content (RC) of the simple individual neurons of V1: receptive field and orientation preference. Alva Noë and Evan Thompson are clearly right when they claim that:

> « RF-content is too thin to sustain a match with perceptual experience. » (Noë-Thompson, 2004, p. 90)

> « It is difficult to see how it [a structural coherence] could be built up (...) out of RF-contents atoms. » (*Ibid.*, p. 14)

But we have also to take into account the functional architecture of V1 where "vertical" retinotopic connections are completed with "horizontal" cortico-cortical connections. Through its FA, V1 does acquire an emerging global geometric content according to the following table:

Local neural RC	≡	Receptive fields, orientation preference,
Functional architecture	≡	Retinotopy, horizontal connections
Global geometrical RC	≡	Eigenmodes

♦ *If we accept* the deduced eigenmodes as a part of the representational content of the neural system N = V1 as a whole, then Bressloff's example show that a *NCC does exist*.

♦ *If we don't accept* that contents are logically transitive (in other words that a content C_2 mathematically deducible from a content C_1 is a part of C_1), then we have to define what type of representational content these geometrical emergent contents can be.

In any case, the fundamental limit of the critiques raised against the NCC and the MCD (matching content doctrine) is in general to underestimate the crucial role of neural functional architectures.

We will call *deduced RC* (RC_2) a structure mathematically deduced from the basic RC (RC_1) of a neural system N. According to one's conception – analytic or not – of mathematics one will accept or not RC_2 as part of the RC of N. But in any case the problem is that RC_2 *cannot be conceptually inferred* (without mathematics) from RC_1 and *depends hic et nunc on the contingent progress of mathematics*.

4.3. The philosophical status of the example

To conclude this part, let us specify a little bit more the philosophical status of Bressloff's example.

1. It is neither *eliminativist reductionism* nor *epiphenomenalism*: the eigenmodes modeling Klüver's planforms exist as geometrical idealities independently of their fine grained neurophysiological implementation; their necessary and sufficient condition is the functional architecture.
2. It is not *dualism*: the correlation $N \leftrightarrow M$ is not a relation between two heterogeneous ontological domains. The emerging morphologies are geometrical idealities and there exists no specific ontology of mathematical idealities (no Platonism).

3. It is not *ontological emergentism* in the sense that the emerging structures would be not only autonomous but also irreducible. Here, they are at the same time autonomous and deduced.

4. It is not *functionalism* in the classical sense of a mere relation of implementation of the upper macro level **2** into the lower micro level **1**. The structures of **2** emerge from a *dynamical bifurcation process* namely *a critical phenomenon of symmetry breaking*.

5. It is not *representationalism* that eliminates the non-representational properties of mental contents: space is not a conceptual and a representational content but the general condition of possibility of intentional and semantic contents.

6. It is *dynamical emergentism*: the global geometry M can be deduced neither from the behavior of isolated individual neurons nor from the collective behavior of a *differently organized* neural net (but it is multi-realizable as long as the functional architecture obtains). Nevertheless it is law-like and can be nomologically deduced from the functional architecture of the NC system $N = V1$.

The most appropriate philosophical approach seems to be a variant of the *double-aspect theory* of Thomas Metzinger:

> « Scientifically describing N and phenomenally experiencing E are just two different ways of accessing one and the same underlying reality. » (Metzinger, 2000, p.4)

« The same underlying reality » is here a *neural dynamics*. It can give rise to a geometric ideal emerging reality (geometric RC) or to a phenomenal experienced one (phenomenal RC).

If we look carefully at the factorization $\mathbf{1} \to \mathbf{2} \to \mathbf{3}$ we can conclude:

1. There is no really hard problem for the emergence $\mathbf{1} \to \mathbf{2}$ of global geometry out of local physics.

2. There is no really hard problem for the content matching $\mathbf{2} \leftrightarrow \mathbf{3}$ (morphological isomorphism).

3. But *there does exist a hard problem* concerning the matching between geometrical space and phenomenal space. But this is not a matching of conceptual contents. It is a matching between pure intuitions and mathematical idealities.

BIBLIOGRAPHY

Bosking, W., Zhang, Y., Schofield, B., Fitzpatrick, D., "Orientation Selectivity and the Arrangement of Horizontal Connections in Tree Shrew Striate Cortex", *Journal of Neuroscience*, 17-6, (1997), 2112-2127.

Bressloff, P., Cowan, J., Golubitsky, M., Thomas, P., Wiener, M., "Geometric visual hallucinations, Euclidean symmetry and the functional architecture of striate cortex", *Phil. Trans. R. Soc. Lond. B*, (2001), 356, 299-330.

Bressloff, P., Cowan, J., "The functional geometry of local and horizontal connections in a model of V1", *Neurogeometry and Visual Perception* (J. Petitot, J. Lorenceau, eds.), *J. Physiology (Paris)*, 97, 2-3, (2003), 221-236.

Chalmers, D., "What is a Neural Correlate of Consciousness?", *Neural Correlates of Consciousness: Empirical and Conceptual Questions* (T. Metzinger ed.), MIT Press, 2000.

Das, A., Gilbert, C.D., "Long range horizontal connections and their role in cortical reorganization revealed by optical recording of cat primary visual cortex", *Nature*, 375, (1995), 780-784.

De Angelis, G.C, Ohzawa, I., Freeman, R.D, "Receptive field dynamics in the central visual pathways", *Trends in Neuroscience*, 18, (1995), 451-458.

Ermentrout, B., Cowan, J., "A mathematical theory of visual hallucinations", *Biol. Cybernetics*, 34, (1979), 137-150.

Field, D.J., Hayes, A., Hess, R.F., "Contour integration by the human visual system: evidence for a local 'association field'", *Vision Research*, 33, 2, (1993), 173-193.

Frégnac, Y. Introduction to *Neurogeometry and Visual Perception* (J. Petitot, J. Lorenceau, eds.), *J. Physiology (Paris)*, 97, 2-3, (2003).

Hayek, F., *The Sensory Order: An inquiry into the Foundations of Theoretical Psychology,* Chicago: U. of Chicago Press, 1952.

Kanizsa, G., *Grammatica del Vedere*, Bologna, il Mulino. English ed. *Organization in Vision: Essays on Visual Perception*, Praeger, 1979.

Mallat, S., *A Wavelet Tour of Signal Processing*, New York, Academic Press, 1998.

Metzinger, T., (ed), *Neural Correlates of Consciousness: Empirical and Conceptual Questions*, MIT Press, 2000.

Noë, A., Thompson, E., "Are there Neural Correlates of Consciousness?", *Journal of Consciousness Studies*, 11, 1, 2004.

Petitot, J., "Morphological Eidetics for Phenomenology of Perception", *Naturalizing Phenomenology: Issues in Contemporary Phenomenology and Cognitive Science*, (J. Petitot, F. J. Varela, J.-M. Roy, B. Pachoud, eds), Stanford University Press, Stanford, 330-371, 1999.

Petitot, J., "The Neurogeometry of pinwheels as a sub-Riemannian contact structure", *Neurogeometry and Visual Perception* (J. Petitot, J. Lorenceau, eds.), *J. Physiology (Paris)*, 97, 2-3, (2003).

Ray, T., in *Toward a Science of Consciousness VI*, Tucson, April 7-11, 2004.

Shmuel, A., Grinvald, A., "Coexistence of linear zones and pinwheels within orientation maps in cat visual cortex", *Proc. Natl. Acad. Sci. USA*, 97, 10, (2000), 5568-5573.
Thom, R., *Stabilité Structurelle et Morphogenèse*, New York, Benjamin, Paris, Édiscience, 1972.