

A formal window on phenomenal objectness

Silvère Gangloff

October 7, 2020

1 Introduction

Along with the development of cognitive sciences that followed the important neuroimaging progress of the years 1990, some hope has arisen to gain some insight into the intuitive notion of consciousness (the fact that one has some phenomenal experience resulting from the physical action on oneself of something that exists "out-there"), by connecting philosophical intuition and direct introspection to the description of cerebral mechanisms in order to find in them, following the model of physical science, the origin of this phenomenon. Let us note that the importance of this notion comes not only from practical applications (for instance detecting consciousness in a person in coma state) but also from the understanding of human species, as consciousness is (at least apparently) specific to this species.

This led to a blooming of theories, each of which correlates (and sometimes identifies) consciousness with some cerebral organisation principle (see for instance the predictive coding [F09], or the global neuronal workspace [DCN11]), and often relies on a *constancy hypothesis*, which is to say that there is a one-to-one correspondance between signals propagating in the brain (or more generally elements of *information*) and elements of perception. Although very attractive, this kind of hypothesis is each time intuitive only on a very restricted domain of experience, and for this reason is rejected by philosophers (notably M.Merleau-Ponty). In particular, these theories fail to explain the origin of the existence and aspect of phenomenal experience in a physical way (the so-called *hard problem*).

As a matter of fact, it is still debatable whether any theorisation of phenomenal experience as such can properly be considered as reliable *knowledge*. This difficulty comes from a conflict between the fidelity to the investigation method and the fidelity to the object of investigation - for in order to follow the method one has to distort the object, while keeping the object intact makes it difficult to follow the method - which is incarnated in oppositions between theorists.

For instance, D.Dennett [D91] argued that since direct introspection and intuition can sometimes fail, one should not rely on them for the constitution of reliable knowledge. A scientific theory of *mental events* (which constitute phenomenal experience) should thus be constructed "from a third-person point of view, since all science is constructed from that perspective" [[D91],p.71], granting priority to the method over the object. This led D.Dennett to defend the philosophical position according to which artificial intelligence describes the organisation of the brain accurately enough so that one can consider them to be equivalent (what J.Searle called *strong AI*), to the point that the existence of consciousness is denied (the intuition of it is just an illusion).

On the other hand the preoccupation of J. Searle, who stood in a strong opposition to D.Dennett, was to preserve the intuition over the scientific method, since even if what is given by the intuition is only appearance it is precisely what is to be understood and any theory of consciousness should account for its existence and aspect. In particular for J.Searle strong AI is inadequate, as it is possible (the Chinese room argument) to conceive an experience for which the conclusions of this theory contradict the evidence (that in this experience, I don't *understand* Chinese language). Therefore if the scientific method is not compatible with the object of investigation, the method has to be changed, not the object:

"If we have a definition of science that forbids us from investigating this part of the world, it is the definition that has to be changed, and not the world". - J.Searle, The mystery of consciousness, p. 114.

Although some theorists take less radical positions than D.Dennett in accepting "the reality of experience and mental life while keeping the methods and ideas within the known framework of empirical science" [V96], this kind of positions is still not acceptable since the object is simply kept outside of scope of investigation.

In a parallel way, J.Petitot et al. [PPRV00] argued in multiple publications that it is possible to transcript the discourse of phenomenology - which, following F.Brentano's project of scientification of philosophy undertaken notably by Husserl and later Merleau-Ponty, is the part of philosophy which is the closest to natural sciences - on the experience into mathematical formalism. Here mathematical language is seen as a medium through which the first-person perspective can be related to the neurophysiological description of brain mechanisms. In particular, J.Petitot [P93] observed that the most suitable way to represent phenomenal space is to use differentiable surfaces, leading him to a transcription of some of Husserl's analyses of perception into notions built over differentiable surfaces in the field of differential geometry (in particular vector bundles and their sections). As a consequence he formulated the more general statement that the discourse of phenomenology can be formalised in geometrical terms.

At this point it is natural to remember Husserl's opposition to the formalisation of phenomenology. The response of J.Petitot to this recall is that in his reflexion Husserl was influenced by the limitations of the science of his time (the movement of axiomatisation of mathematics in particular, and the limits in terms of available mathematical tools). Let us note that this reasoning is also used by experimentalists such as L.Albertazzi [A18], for whom experimental phenomenology methods can also overcome the obstacles pointed out by Husserl. In persisting furthermore into a *formalisation* or a *naturalisation* of phenomenology, these authors are undermining the meaning of Husserl's warning. D.Zahavi [Z14] notably pointed out that his opposition to the naturalisation have philosophical reasons more than scientific ones, recalling that "for Husserl, the problem of consciousness should not be addressed on the background of an unquestioned objectivism but in connection with overarching transcendental considerations". This objectivism that Zahavi finds in J.Petitot et al. is the philosophical position according to which the reality can be described in terms of objects (in particular mathematical ones) - which could be defined here as parts of the Experience that can be described with both complete accuracy and finite cognitive (or language) means. The focalisation on this type of experiences and the distortion of reality in which it results comes from an orientation towards the method.

In fact the problem with the transcription into mathematical objects of phenomenal experience is not the unavailability of accurate enough tools and the orientation of science: it lies in the distortion of what is to be understood along with its objectivist identification to objects that one finds to be close enough to it in the intuition. D.Zahavi goes a bit further than this opposition to objectivism in noticing that, in the fidelity to the scientific method and in their naturalisation project, J.Petitot et al. have to abandon the transcendental aspect of phenomenology (in particular the study of consciousness as the condition of possibility of meaning, truth and appearances) to turn towards phenomenological psychology, a local descriptive approach. Moreover there is an additional epistemic difficulty in the path undertaken by J.Petitot et al. in the absence of a natural correspondance between the psychological descriptions of phenomenology (transcribed in mathematics) and the data of natural science, which is an evidence for the non-sufficiency of the bare use of mathematical language as a medium in the constitution of a relation between the physical world and phenomenal experience. As a matter of fact, the need for the creation of meaning in-between the disciplines of phenomenology (and philosophy in general) and natural science (in particular mathematics) requires also the preservation of both disciplines' principles of investigation, along with a possible inflexion in the form of these disciplines.

In this article I attempt to respond to Searle's call for a redefinition of *science*, or more precisely of the restrictions that (should) apply to some form of knowledge in order for it to be considered as

reliable. This kind of definition used to take historically, in particular in the texts of Kant, Schopenhauer, Wittgenstein (verificationism), R.Carnap (tolerance principle), or K.Popper (falsifiability criterion), the form of a demarcation between science and philosophy in general, which was aimed at preventing any form of dogmatism in institutionalised knowledge. By their separative nature, these criteria do not represent steps towards an understanding of the illusory 'essence' of these disciplines: they contribute to what they actually are by creating what philosophy and science are and should be, in themselves and respectively to each other. This history has an impact on the actual possibility of these disciplines to cooperate on the complex matter of consciousness. In fact some of these criteria (as K.Popper noticed for verificationism in the beginning of *The logic of scientific discovery*) also prevent any science to exist when they are applied in a strict way. Despite this inaccuracy, it is not acceptable to simply ignore the demarcation history:

"What needs to be emphasized in closing, however, is that there is a decisive difference between claiming that philosophy and empirical science should cooperate, and denying their very difference."
- D.Zahavi [Z14].

As a matter of fact, simply denying this difference is exposing oneself to dogmatism again. One should think of the necessary cooperation between philosophy and empirical science as a dialogue between subjects who, despite allowing the other to act on oneself (through representations), keep their identity along the dialogue. As the irregularity of consciousness as an object of study results in the dilution of scientific discovery, this dialogue should be focused on the creation of meaning and the extension of our intuition's current bounds. I believe that an archaeology (in M.Foucault's sense) of the historical limit between these two disciplines can lead to a minimal (and abstract) instantiation of this limit that allows this dialogue to be possible. In the present text, I formulate such a limit under the form of a combination criterion that deals with an aspect of the discourse (statical-dynamical), in which the actual differentiation between the philosophical and mathematical (and thus scientific) discourses is rooted.

The remainder of this text is devoted to a practical application of this combination criterion. The recent *integrated information theory* of consciousness, which has received some attention in the past decade, begins with the observation that some characteristics are shared by all the experiences that one can imagine having. The consequence is that these characteristics are specific to the phenomenon of experience itself, and not to any particular experience. One may then search for mechanisms underlying these characteristics, *explaining* phenomenal experience as such. In a context in which brain mechanisms are usually represented by probabilistic dynamical systems, the formalisation of these characteristics (based on a notion of cause-effect structure) serves as a meaning-creation device between experience and mechanisms. It has been deployed in particular to draw a correspondance between grid-like neural networks, that are present in various areas of the brain associated with spatiality (such as the visual cortex), and specific aspects of spatial experience. Although natural, this approach brings several epistemic difficulties. While most of its critics (in particular the one of S.Aaronson) point out counter-intuitive philosophical consequences, this theory also suffers from the intractability of the mathematical notions that it defines (put into evidence in [MMM] for the notion of integration, although for a particular interpretation of it), which leads to question its ability to extend our intuition beyond its current bounds. Let us note that despite counter-critics which 'recall' the initial counter-intuitive character of some scientific predictions, the critics based on counter-intuitive consequences still stand since in these cases the intuition has the reflex to go beyond its domain of application, while for consciousness situation is fundamentally different (for the reason that intuition is inherent to the object of study). Moreover the formalisation process that is used comes along, by its nature, with some meaning loss, reflected in the case made of spatiality. This case makes manifest the difficulty to develop further the correspondance made between the structure of Experience and the formalism beyond its formalisation domain, in particular in the use of ad hoc argumentation. As some of the ideas presented in integrated information theory are original and of interest, I would like, following Kant:

"Now it may seem natural that as soon as one has abandoned the territory of experience, one would not immediately erect an edifice with cognitions that one possesses without knowing whence, and on the credit of principles whose origin one does not know, without having first assured oneself of its foundation through careful investigations, thus that one would have long since raised the question how the understanding could come to all these cognitions a priori and what domain, validity and value they might have." - **Kant, Critique of pure reason, p.128,**

to analyse further the formalisation process of integrated information theory instead of building upon the formalism that it produced, in particular under the form of an excessive defense of its current form, in the way described and criticised as follows by K.Popper:

"A system such as classical mechanics may be 'scientific' to any degree you like; but those who uphold it dogmatically—believing, perhaps, that it is their business to defend such a successful system against criticism as long as it is not conclusively disproved—are adopting the very reverse of that critical attitude which in my view is the proper one for the scientist." - **K. Popper, The logic of scientific discovery, p. 28.**

I analyse the theory both in its foundations and its application to phenomenal spatiality, relating the epistemic difficulties of the theory to the form of its discourse. After providing some intuitions on how to 'unknot' this theory I introduce a formal framework for the study of a particular aspect of consciousness which consists in the ability of the subject to distinguish 'objects' in its Experience (objectness, or the character of what integrated information theory calls a *distinction*). An understanding of this phenomenon, chosen for the ability to introspect it, would represent an important progression towards consciousness itself. The presented framework, which is specific to the study of objectness (in particular there is no reason to extend it beyond), is meant to obey to the combination criterion, and thus allow the dialogue between mathematics and phenomenology, as well as to solve the problems of tractability of the formalism and its ability to represent the reality.

2 On the statical and dynamical phases of the discourse

In the following, I will think of knowledge, and thus the discourse (through which knowledge is constructed and transmitted) in terms of some elementary acts of language that I call designations. Here a designation is the simple act of pointing to an area of the Experience (the totality of my experiences); it can be an act upon myself or another person, using any mean (in particular words). I also call a designation any collection of designations that are identified in the intuition; an element of this collection is then called an instance of the designation. A designation is said to be **static** when all its instances point towards the same area of Experience. For instance, the designation of "the triangle" (that acts on any experience by selecting the triangle it contains if it does contain a unique one), together with all its derivatives in other languages, is a static designation (in other words, there is no loss of meaning through translation). Otherwise, a designation is said to be **dynamic**. One can think for instance to the designation of "life", which contains in it instances pointing to various areas of Experience, sometimes including certain forms such as viruses, sometimes not.

Let us notice that one could find a natural objection to the static character of the designation of "triangle", since some of its instances select 'triangles' drawn on manifolds (for instance bidimensional surfaces such as a sphere or a torus) and not only plane ones. However although these designations use the same word, they are not identified in the intuition (and in fact can be easily differentiated). On the other hand two instances of the designation "life" are thought as two possible objective interpretations of the same 'intuition'.

Moreover one can notice that the static character of a designation is correlated with the interpretation of its origin (the area of Experience towards it points) as existing "out-there" as an object, whose current state of existence does not depend on my will (what happens 'in-there'),

while the dynamic character of a designation is correlated with the influence of my will on its form. In a sense the opposition static-dynamic is a projection on the discourse and its morphology of the Schopenhaur's opposition between the representation and the will.

2.1 The dynamical in the collective discourse

2.1.1 On dogmatism

In the following, I will also see the totality of the discourse as organised in a series of strata, whose distinction derives from the practical use of the discourse: the individual **inner discourse**, meaning the inner activity of the subject who projects pre-constructed concepts onto his or her own experience, the **intersubjective discourse**, which consists in the reciprocal action of designations between two subjects or more (including the progressive synchronisation on the meaning of designations, through designations of designations), and the **collective discourse** which consists in the elements of discourse which are asserted intersubjectively to have collective value.

In other words, the collective stratum of the discourse is a support of knowledge, on which anyone can rely to support his or his own actions. From the value of this discourse derives the value of its writers. As a consequence the integrity of this stratum of the discourse is threatened in many ways by the will of these writers: for instance the will for academic (or spiritual) prestige, or the possibility to act on people's mind at the collective level, in order to induce in them a particular behavior. For the execution of this will, one could be tempted to enforce in the collective stratum the inscription of a discourse that originates in a intersubjective agreement which, despite the fact that it is not representative of the collective, derives its force from an illusion of representativeness which comes from the number of subjects together with the formation of an intersubjective will. This is what is usually called dogmatism.

In order to prevent this phenomenon, a natural way to do is to make use of a practical criterion for granting a discourse an access to the collective stratum. In particular one usual sufficient condition for a designation to get this access is to have statical character: indeed from a cognitive point of view this staticity is interpreted as a sign for the existence of an object "out-there" in the world. This mode of existence makes this object independent, in its current state at least, from any subject and permanent in an ideal way in the phenomenal world.

As a consequence dogmatism takes most of the time the form of a confusion (whether it is conscious or not) about the statical or dynamical character of the designations involved in a discourse, for instance by an artificial staticisation, sometimes by the apparent but artificial beauty of symmetrically structured formula and systems (see for instance the critics of Kant and Hegel by Schopenhaur, in *The World as Will and Representation*).

2.1.2 On the demarcation *problem*

The strategy of access granting to the collective stratum took multiple forms in the past. For instance Humean empirical science restricted the access to close and methodical observations and descriptions of a meticulously controlled domain of experience of the world of sense. The problem with this criterion is that it excludes from the collective stratum some part of the discourse which constitutes manifestly reliable knowledge: mathematics. As a reaction to strict empiricism and in order to include mathematics and more generally the theoretical aspect of knowledge while excluding manifest dogmatism (for instance in Leibniz metaphysics), Kant elaborated a demarcation between natural science and philosophy based on functional differentiation and complementarity. After Kant this line of thoughts was followed by attempts to formulate an accurate demarcation line between these two disciplines (the **demarcation problem**).

Following M.Foucault's *Archaeology of knowledge* I analysed (partially) the mode interaction between mathematics and philosophy and its history, in particular through the search of demarcation criteria, in order to understand how this interaction gave form to the identity and distinction of these disciplines respectively to each other. I arrived at an interpretation of the evolution of the demarcation problem in terms of statical and dynamical uses of the language which consists in

a philosophical **orientation**, which means that the existence of continuum from a radical philosophical position, incarnated in particular in logical positivism's verificationism, of exclusion of the dynamical use of language from the collective stratum to a position of conditional inclusion, incarnated in particular by phenomenology.

All along this orientation there is a constant **separation of two *phases*** - where "phase" is a way to designate a group of designations that is distinguishable beyond particular texts - one made of all statical designations and the other one of all dynamic designations. The criterion that I propose for reliable knowledge is this invariant, understood in an abstract way, together with its explicitation in the discourse. The reduction of the actual difference between philosophy and natural science to this abstract and minimal separation shall allow them to interact in a closer way.

2.1.3 The self-exclusion of philosophy based on its dynamical character

The most radical position that I analysed in this text is the one of *verificationism* (adopted notably by logical positivism) whose principle is captured in the following formula:

"The right method of philosophy would be this. To say nothing except what can be said, i.e. the propositions of natural science, i.e. something that has nothing to do with philosophy: and then always, when someone else wished to say something metaphysical, to demonstrate to him that he had given no meaning to certain signs in his propositions. [...] Whereof one cannot speak, thereof one must be silent." - **L. Wittgenstein** in the **Tractatus logico-philosophicus**.

Analysis: According to this position, language can be thought of as consisting in a set of propositions, where a proposition is a construction made of signs assembled with logical connectors. With this identification in mind, one should be careful about that all the signs in a proposition have a meaning, such as in natural science, for it to potentially constitute knowledge. In other words only these propositions *can be said*. Although in metaphysics the attention is directed to certain things whose existence is not denied straightforwardly here, one should not talk about these things because of our inability to make use of the language to do so. As a consequence, one should restrict philosophy to a domain which has nothing to do with what it actually is, in a sense one should exclude philosophy from the collective stratum of the discourse.

Interpretation:

1. **What does meaning mean:** I think of a sign as simply a synonym for a designation, which means that it has a meaning (in the collective stratum) when it designates something that exists "out-there", and because of that it is possible to use a word to communicate this designation to another subject. In particular, this means that the designation can be communicated without loss of meaning, that is to say that what is pointed at by me through the designation is the same area of experience as what is pointed at by the other (this can be thus considered as a collective designation). On the other hand, a designation does not have a meaning when a significant loss of meaning is unavoidable in its communication. In particular when I try to communicate it, I can not be sure that I and the other are talking about the same thing. Furthermore there can not be a control from the other on that what my designation points at is an illusion or not. All that is a source of instability.
2. **On the properties of natural science's language:**

In the formula *natural science*, one should attribute to the term *natural science* a broad meaning in this context, not only referring to the discourse on the Nature as understood nowadays or the βίος but to the discourse about *Nature* understood as all that is "out-there" (including also physics for instance). The particularity of natural science is that in its language all terms have meaning (they refer to something "out-there", by definition of natural

science as the discourse on all that is "out-there"). In order to understand the viability of a discourse in general, one should think about this discourse in an abstract framework where the discourse is thought as a correspondance between words and parts of the Experience one has. In order to understand what makes the discourse of natural science meaningful, I would like to avoid tautology and should then suspend the intuition of the nature of the "out-there" and try to understand how meaningfulness derives from properties of the discourse itself. My claim here is that in the language of natural science, the terms are ensured to have a meaning because any term can be derived by construction from *simpler* terms (in the sense that these terms are easier to transmit intersubjectively as designations) which have clear meaning, in a way such that the construction rules ensure that any constructed term has a meaning.

For clarification, these properties can be seen for instance in Euclidian geometry, where designations can be considered as sequences of visual experiences extracted from the Experience (for instance the progressive construction of a triangle). These designations can be constructed from the simplest designations of point, straight line and circle - which consist intuitively in a strict correspondance between a word and a visual object, or a class of visual objects that can be identified by simple cognitive transformations on these objects of experience; for instance, a rotation of a straight line onto another one - and constructions of sequences of experiences by combination of these terms (for instance the arbitrary addition of an object, or the distinction of a pair of objects's intersection). These transformations are also designated by words and a designation which belongs to Euclidian geometry's discourse consists in the association of a series of words with a series of visual experiences. One can convince oneself that the transformations used in these constructions always derive an experience from another one, and therefore any designation points to something that can be thought as "out-there".

3. Strictly metaphysical terms lack meaning:

On the other hand some terms that are specific to the metaphysical discourse do lack meaning in the sense that they do not point to something "out-there". This is the case for instance for the term of *an absolutely necessary being* (referring to God), used by Leibniz in his proof of God's existence, which was criticized by Kant in the *Critique of Pure Reason*. Although this term may correspond to "something out-there", the term is not pointing at it in a way that it can be the object of a collective agreement ; for this reason, it lacks meaning. Bergson, in *The two sources of morality and religion*, gives a simpler example (although more abstract) of the *square circle* (a circle that has the property to be a square), for which it is clearly not possible to find anything in the Experience that it points to: this term is a pure construction of language, that is possible only by abstracting words from their meaning. Comparatively to Wittgenstein, Bergson goes further than excluding metaphysical terms for their lack of meaning and explain the reason why they don't: one can find for instance an analysis of the term *freedom* in *Time and Free Will: An Essay on the Immediate Data of Consciousness*, pp. 131-137.

The danger of using terms that have no meaning is that when combining them with other terms, one can not be ensured that the constructed terms or propositions correspond to anything "out-there". Moreover in more complex cases than the square circle, and despite the absence of meaning, such terms or propositions can give a temporary illusion of meaning. This can be illustrated with Kant's antinomies of pure reason, which are examples of contradictory propositions that can both be *proven* (by pure reason) to be true. For instance, the atomism antinomy: *every composite substance in the world is made up of simple parts, and nothing anywhere exists save the simple or what is composed of the simple.* (thesis) ; *no composite thing in the world is made up of simple parts, and there nowhere exists in the world anything simple* (antithesis). However since these two propositions contradict each other, it is not possible that both are true. As a consequence, as well as for the square circle, the pair of

'proofs' of these propositions should include terms that lack meaning: in the interpretation of the proofs these ambiguous terms are interpreted in distinct ways, keeping the coherence of the whole.

This problem has a counterpart in modern computer science, where it is solved with the notion of objects *type* (for instance integer or matrix) attached to the objects themselves: this type attachment permits to ensure that operations on objects can be executed only when the objects involved are of a certain type in order to preserve meaning through any of these operations.

4. That which should not be said:

In order to preserve the integrity of the collective stratum one should thus restrict this stratum to propositions whose terms have a clear meaning (only these propositions should be said), excluding metaphysics and more generally the whole discourse of philosophy (including the study about consciousness): in the text of Wittgenstein philosophy excludes itself from the collective stratum because of its dynamical character. However this criterion is too restrictive for the problem of consciousness physical origin to be considered as a definitive solution to the demarcation problem. Moreover as noticed by K. Popper, Wittgenstein's position excludes manifestly reliable knowledge such as mathematics.

In terms of statical and dynamical phases:

Along this analysis, it becomes clear that one should differentiate abstract examples of meaningless designations such as the square circle, which do not correspond to anything in the intuition, from examples that are closer to the intuition - such as the term freedom - which does correspond to a region of Experience (although complex), in particular the set of situations in which I feel free. Designations of the last kind, despite the absence of meaning in a full sense, can be seen as a step in a series of designations identified into a dynamical one, progressively enclosing a thing that exists "out-there" (while this is not possible for empty designations such as the square circle). In this direction of doubt, one should consider a priori any dynamical designation of this kind as possibly leading (experientially and not deductively) to a statical designation and more definitive knowledge. This constitutes a first step towards an inclusion of the dynamical phase in the collective stratum.

2.1.4 Kantian functional separation

One of the main points of Kant in the *Critique of pure reason* was (in short) to understand how to account for the apodictic character of mathematics (whose truth lies beyond any individual discourse and is thus should belong to the collective stratum) whose paragon is Euclidian geometry, while preserving the dogmatism exclusion of Hume's empiricism, in particular of the use of pure reason in Leibnizian metaphysics. His position on this use of pure reason can be abstracted from the following formula:

"Thus the principle of reason is only a rule [...] Thus it is not a principle [...] of the empirical cognition of objects of sense, hence not a principle of the understanding, for every experience is enclosed within its boundaries (conforming to the intuition in which it is given); nor is it a constitutive principle of reason for extending the concept of the world of sense beyond all possible experience; rather it is a principle of the greatest possible continuation and extension of experience, in accordance with which no empirical boundary would hold as an absolute boundary; [...]. Hence I call it a regulative principle of reason [...]" - I. Kant, Critique of pure reason, p. 520-521.

Analysis: In this short paragraph, Kant distinguishes the *principle of reason*, which refers to the Leibnizian sufficient reason principle (which states that everything must have a reason or a cause),

from the *principles of understanding* and the *constitutive principles of reason*, which are actual examples of principles. The principle of reason is not properly a principle but a *only* a rule, and if a principle of reason, it is a *regulative principle*.

Interpretation:

1. **Principles:** Here a principle is a statement that is inherent to a functional mode of the relation between a subject and his or her Experience and participates to the relation with Experience as such (and not to particular experiences). Kant mentions two of these functions.
2. **On the understanding function:** The understanding is the faculty of human cognition that is directed in the Experience to things that are already expressed in the collective language but not fully expressed at the subject's individual stratum. Understanding the articulation and the function of a mechanism, for instance, consists in a sequence of cognitive operations: the creation of an empty concept of this mechanism; the focus on certain parts of the experience (the pieces of the mechanism and their articulation, as well as the effect of the functioning of the mechanism); their collection and assembling in the empty concept.
3. **The function of constitution:** On the other hand the cognitive function of constitution consists in the creation of concepts at the individual and intersubjective levels which are not present in the collective stratum, enriching the conceptual set (that Kant names the *the concept of the world (of sense)*) of the subject or group of subjects with concepts that one can use to designate some parts of the Experience. This creation can be decomposed in: the pure appearance of an area of Experience; the observation that no concept in the current conceptual set encloses accurately enough this area of Experience; the projection of articulations of some of these concepts onto it; the selection of the closest articulation and its stabilisation into a concept of the area in question. In the transition from the individual stratum to the intersubjective one, the subject can make use of his or her current conceptual set in relation to the collective conceptual set to direct the other's attention to the some part of his or her Experience and make it appear in the other's Experience and conceptualised in a catalysed process.
4. **In relation with the functions statical part:**
Although the precise implementation of these cognitive functions (of understanding and constitution) and their description are not entirely statical designations, this is the case for the statement of their existence and their independence from the current set of possible experiences. A principle of such a function is a statement about it that participates in the statical part of its designation.
5. **A rule of experiences:**
By opposition to a principle a rule is a statement that is true of every possible experience, where an experience is possible when I can imagine having it. As well I interpret the "possible continuation and extension of experience" as the set of possible experiences. The empirical boundaries of this set should not be considered as absolute ones: in other words, a subject can conceive the possibility of having other experiences than the ones that currently considered as possible. Over a strict extension of this set of possible experiences this rule could not be longer true. As a designation, it is thus dynamical. On the other hand, a principle does not hold this dependency. In that sense, the principle of reason is not an actual principle but *only a rule*.
6. **A regulative action on the relation with Experience:** Despite that Kant considers the principle of reason as a rule of experiences and not a principle (of cognition), he still attributes it a functional role in the construction of the collective stratum, in the sense that it is a regulative principle, meaning that it regulates the relation between the subject and his or her Experience. In particular it acts by orientation on the 'look' to make it search, under any

circumstance, the cause for the phenomena that the subject observes, without however any guarantee to find such a cause. Although this kind of rules does not participate knowledge, they are of practical use in its constitution, in the extension of collective Experience. This can be extrapolated to the attribution of a regulative role of the whole philosophical discourse over the progression of the scientific one.

A consequence of the attribution of functions to the philosophical and scientific discourses in Kant's position is the cognitive (and then material) separation of these discourses, rooted in their different use of the language (dynamical versus statical). In particular, this separation implies (implicitly) the exclusion of the philosophical discourse from the collective stratum, even if it has an action on its form.

2.1.5 The falsifiable class and potential staticisation

The (falsifiability) criterion proposed by K. Popper goes a bit further, including in principle some class dynamical designations (in the form of statements) in the collective stratum. These dynamical statements are the ones that allow the possibility to falsify them. This falsification is in fact a particular case of staticisation - which means rendering dynamic designation definitively static. In other words the use of this criterion represents a change in the notion of reliability of the knowledge which constitutes the collective stratum, from the simple statical character of the designations constituting the collective stratum to the possibility of an easy regulation of its content by selective exclusion (through the action of falsification). This shift allows in particular the possibility of a collective exchange of ideas (and thus not only bare facts).

In a strict sense, the criterion of falsifiability is not sufficient in itself to ensure the reliability of the collective stratum: for instance L. Laudan noticed in *The Demise of the Demarcation Problem* [Laudan], this criterion allows in principle some theories collectively accepted as pseudo-sciences - which are such for they make retention of false statements by suspending the falsification action in isolation from further intersubjective dialogue. However the possibility is left to complete the criterion with a mechanism of collective enforcement of caution falsification use. However L. Laudan followed another strategy, which consists in the redefinition of the distinction between science and non-science in terms of discourse reliability. In the same direction P. Thagard [Thagard] defined a pseudo-science as a stagnating theory (in other words a theory which exists by staticisation of the dynamical phase, whether this staticisation is active or passive). The combination criterion (of the phases separation) that I propose here is a generalisation of these lines of thoughts. The interest of this more abstract formulation is of practical use not only in the characterization of a theory after its construction and existence, but in its orientation during the constitution itself of the theory.

2.1.6 The dynamical in the collective stratum by explicitation

In his doctoral dissertation, *Der Raum. Ein Beitrag zur Wissenschaftslehre* (Space, a contribution to the theory of science), R. Carnap made a comparative analysis of different types of spaces that he identified and their respective conceptualisations, out of a comparison between modern and abstract axiomatisations of Euclidian geometry (notably by D. Hilbert, A. Tarski, and G. D. Birkhoff) and ancient versions. These types of spaces are the following: the intuitive space, the abstract space, and the physical one. An important point of this analysis is that these conceptualisations can coexist for the reason that they are not speaking about the same domains of Experience (in other words, they do not constitute a series of evolutions of a dynamical designation but distinct designations).

In my interpretation of Carnap's work, this kind of thoughts led him to formulate the following *principle of tolerance*:

"In logic there are no morals. Everyone may construct his logic, i.e., his form of language, as he wishes. He must only, if he wishes to discuss it with us, clearly indicate how he wishes to do it,

and give syntactical rules instead of philosophical considerations." - **R.Carnap, 1934, Logical syntax of language.**

In this formula, the "syntactical rules" refer implicitly to the definition of an axiomatic system. As long as it is clearly defined, any of these systems is of interest, in particular towards a better understanding of existing systems by comparison. Although Carnap's aim is not to the possibility of an inclusion of the dynamical phase in the collective stratum, one can interpret this principle by replacing Carnap's syntactical rules with transcendental operations (such as Husserl's ἐποχή). This leads as an application - despite the fact that this philosophical movement appeared in the different context of F.Brentano's project of scientification of philosophy - to (Husserlian in particular) phenomenology, characterised by E.Husserl and M.Merleau-Ponty as follows:

"The method of phenomenology is to go back to things themselves" - **E.Husserl, 1967.**

"To return to the things themselves is to return to that world which precedes knowledge, of which knowledge always speaks and in relation to which every scientific schematization is an abstract and derivative sign language, as the discipline of geography would be in relation to a forest, a prairie, a river in the countryside we knew beforehand" - **M.Merleau-ponty, 1945.**

Interpretation:

1. A shade of the language on Experience:

In this quote M.Merleau-Ponty makes a clear distinction between the system of signs out of which the language is made of and the world that it is speaking about, the purpose of language being to communicate designations (through the signs) towards parts of this world that are intersubjectively experienced and more generally about the intersubjective Experience. This distinction is motivated by the phenomenon, derived from nature language itself, of the tendency to consider a pre-constructed (at the collective level) conceptualisation of things instead for the things themselves, despite their difference (in the example of Merleau-Ponty the representations of geographic maps are clearly different from what they represent: forests, prairies and rivers for instance), unless this identification is strikingly denied by the experience itself, including communication with others. In other words, signs can occult things themselves, reducing the phenomenal field. Although M.Merleau-Ponty takes as an example the discipline of geography, I prefer the example of History, for which the extent and the complexity of this phenomenon is wider, and closer to quotidian experience (although the distinction between geographic representations and the things themselves being easier to conceive). For instance we tend, when interpreting historical documents, to project on them the culture of our time or our mode of thinking, distorting the actual information that this document contains. In quotidian Experience this phenomenon can be seen as a particular case of D.Khaneman's *fast brain* [Kahneman] automatic functioning (the functioning mode which, in trade-off between celerity and precision, tips towards celerity) in the sense that, despite the difference between the projection and the thing itself, the approximation is good enough for the current use of this representation.

2. Opening the world: the operation of ἐποχή:

The point of Phenomenology is precisely to counter this phenomenon, which can be devastating when some representations are taken for the things they represent (as designations they are considered as static despite their dynamical character) and transmitted to the collective stratum, where designations are abstracted from their context and origin (which is thus forgotten and thus not retroactable). This can lead on one hand to dogmatism, on the other

hand to the absurd - for instance Russel's paradox (which constitutes one of Husserl's historical motivations for phenomenology). In order to destaticise these designations, in other words to go back to things themselves, Husserl's phenomenology relies on the transcendental operation of $\epsilon\pi\omicron\chi\eta$ (or bracketing), meaning the suspension of judgement. To be more precise, I interpret this operation as the inhibition of the projection reflex, of already formed associations between articulated conceptual structures and parts of experience, onto the opening of the Experience (this part that only appears). The effect of this suspension is to allow other articulated conceptual structures to be projected on this opening, potentially enclosing it more accurately than the current projection. In other words, the operation of $\epsilon\pi\omicron\chi\eta$ opens locally the possible associations between articulations of concepts and the opening of experience. For instance, by the $\epsilon\pi\omicron\chi\eta$, one can 'solve' Russel's paradox by re-cognising the importance of the construction process of a (mathematical) set in its definition (which gives it a meaning). This operation is particularly important to counter dogmatism's confusion between static and dynamic designations in the transition to the collective stratum.

3. An algebra of transcendental operations:

In my interpretation of it, the conceptual void left by the $\epsilon\pi\omicron\chi\eta$ is filled through an automatic process of projection during which the competition between conceptual articulations is not biased anymore towards the pre- $\epsilon\pi\omicron\chi\eta$ projected concept. This happens beyond the domain of observability in the intuition (this part precisely is unavoidable in the process of the construction of knowledge). This bias cancellation can be seen to be positively in favour of all the (previously unfavored) conceptual articulations. This is what in M.Merleau-Ponty's terminology is called the position of the position (where the second occurrence of position is the natural projection of conceptual articulations onto the opening of experience, and the first one is the result of the $\epsilon\pi\omicron\chi\eta$). After some stabilisation on another conceptualisation, this second one is likely to be more accurate than the first one. Along with the position operation, there is another transcendental operation, called the negation in M.Merleau-Ponty's terminology, which consists in the negation of the current projected articulated conceptual structure. With this operation of negation, the definitive conceptualisation is made static (and thus objective in the transition to the intersubjective stratum) by the result of a dynamical process (of construction of knowledge) which combines in series the operations of position of the position and negation of the negation (where the first occurrence of negation comes from the world itself, negating the negation of the association between the concept and the part of Experience).

4. The inclusion of the dynamical phase:

Beyond F.Brentano's initial project of scientification of philosophy, and despite its various theorisations by the authors, Phenomenology could be characterized transversally as the movement of inclusion of the dynamical phase in the collective stratum of the discourse. This inclusion is done by making explicit its dynamical character - interpreting dynamical designations as pieces of a dynamical knowledge process - and making explicit as well the structure of this process: the combination in series of transcendental operations in an indefinite refinement progression towards some static designation. In this inclusion there is a separation of the dynamical and static phases: the former one lies in the direction of the designations, the later one in the 'transcendental algorithm' in which the discourse is rooted. In order to make this more tangible (in particular for mathematicians) one can see an analogy between the shift operated in the reliability notion by Phenomenology and the shift operated in mathematics from a notion of computation characterised as series of algebraic operations on rational numbers to one characterised as the dynamics of computing machines such as Turing's ones. In the first sense a number is thought as computable when it can be obtained after a series of algebraic operations, and in the second sense when it can be approximated with arbitrary precision by the outputs of an algorithm. In this last sense computable numbers can not be 'grasped' cognitively but they can be algorithmically enclosed indefinitely with more and more precision.

2.1.7 Pre-statical designations

Beyond the definition of a transcendental method, an important aspect of Phenomenology is its restriction onto a domain where dynamical designations actually enclose something that we know to exist "out-there" (the transcendental analog of a computable number), while it is possible to conceive such dynamical designation that reveals themselves to be empty after some time (such as the square circle), despite the lack of knowledge on this thing. The certainty of existence of a thing can be reduced to the simple negation of the transcendental operation of suppression (a particular case of the transcendental negation), which consists in conceiving an experience in which the part of the experience designated is suppressed: if I can not conceive such an experience, therefore I am sure that the designation actually points at something that exists (in the same way as I can not conceive an experience in which there is a square circle, which therefore can not exist). For the sake of unicity one can consider this thing to be the whole of all that exists in this part of Experience. I would call such a designation **pre-statical**. The most direct example is the one of the designation 'I', the class of which designations I will denote \mathcal{I} in order to differentiate it from the individual one 'I', as well as the world (that which is outside of 'I') and consciousness as the form of the relation between \mathcal{I} and the world (that Heidegger designated with the one and structured concept of being-to-the-world (*dasein*)). One can find that this pre-statical character (for instance, the \mathcal{I}) is often correlated with a relation with a static designation that is contained in it (for instance, the body): the pre-statical designation is therefore the intuitive extension of this static designation (this differentiates the \mathcal{I} from God for instance, which is completely internal). This explains the pre-statical character but also the presence of the designation in the intersubjective stratum (which reflects the nature of the designation), since it can be communicated through this static designation and intuitive association with its extension and difference with the static designation. However, there is a relative loss of information in this communication, since this intuitive extension differs from a subject to another. The "anchor" of the static designation is the reason why a pre-statical designation is static in its form, while dynamic in its content.

2.1.8 Conclusive remarks

In this section I have shown how, from the progression along a philosophical orientation which ends with the formulation of the phenomenological method and domain, can be abstracted the constant separation of a statical and a dynamical phases of the discourse. The phenomenological inclusion of the dynamical phase can be interpreted as a shift of the limit between the two phases from the periphery to the inside of the collective stratum. Although one can not identify philosophy with the dynamical phase itself, one can state that it is the part of the collective stratum that chooses the dynamical as a mode of ex-pression (in coherence with Kant's functional demarcation). Furthermore, this analysis makes clearer that the project of naturalising phenomenology is meaningless, since it is essential to phenomenology itself as a part of philosophy to allow knowledge to be dynamical when it should be. The only - unacceptable - way to implement this project would be to force the dynamical phase into staticity, which is a form of dogmatism. On the other hand another (more acceptable) way to progress towards staticity is to keep the method of Phenomenology while changing its scope: this is the path followed by micro-phenomenology, which aims at a better control on the dynamical process of knowledge and an acceleration of its convergence by a focus on restricted areas of Experience, and by controlling the transmission of information from the first person perspective to the third person one [see for instance [Z19]]. While micro-phenomenology follows the experimental method, I propose a theoretical counterpart in the second part of this text, which uses the combination criterion of phases separation. The use of this criterion allows to dissolve the separation between mathematics (and natural science) and philosophy - this separation coming with a conceptual architecture which prevents the development of understanding in the direction of consciousness (in particular the opposition between subjective and objective and the association between science and objectivity) - while preserving the restrictions that motivated their separation.

On a higher level of interpretation, verificationism naturally leads to a form of objectivism

when the attention lies in the objective domain of the Experience, assimilating in belief the world to this objective domain. It is thus fundamental in order to avoid objective dogmatism (about the nature of the world) to better understand how the language and the practice of mathematics in their abstract objectivity are related to the more general relation with the Experience in terms of conceptualisation (production of concepts). In this direction, one could see that logical and more generally mathematical propositions are designations conditioned on transcendental algorithms (ordered sequences of transcendental operations): for example, the equivalence of two segments with respect to their length is itself equivalent to the experience of undistinction (of two separate things) conditioned on the rigid displacement of one segment towards the other (using translation and rotation) ; the transcendental algorithms of mathematics are concerned with particular experiential contents which are constructions made out of multiple displacements in space and time (Carnap's abstract space) out of the undividable (the point in the visual space, the element in the attention space, etc). This undividable, as well as parts of the Experience that phenomenology focuses on, exists "out-there" in the same sense, since there is something that exists in my intuition and because of the undividability, it is this thing that exists at the end of my consideration. All the constructions made out of it then exist in the same way. In this way, phenomenology and mathematics are similar in their strategy for reliable knowledge; in a sense, one can see mathematics as a particular form of phenomenology (following Varela's project of phenomenologising science), directed towards formal objects, and generalise them into what I would call **transcendental algorithmics**. In the following section, my aim will be to analyse, through a partition of mathematics into three types (based on the type of experiential areas they are directed to), the implementation of the phases separation strategy in which the form of mathematics - as a discipline - is rooted. This strategy is mainly based on a paradiscursive separation of the dynamical phase and a progression based on a phenomenology of the consideration of already constructed formal objects; by basing itself on already constructed objects, mathematics never leave the ground of experience (in agreement with Kantian philosophy).

2.2 The phases separation strategy in mathematics

2.2.1 Localisation principle and objectivity

As I mentioned above the specificity of mathematics as a phenomenology is its focus on formal objects. An important part of the archaeology of the limit between the authors' phenomenology (and philosophy in general) and mathematics is to analyse the mode of existence of these objects compared to more general areas of Experience that one can designate. Common intuition on the nature of their objectivity refers usually to the notion of *universal observer*: an area of Experience is an object when it is independent from the \mathcal{I} who perceives it. In other words any observer is an instance of this universal observer with respect to this area of Experience. For common intuition science is defined as the form of discourse whose domain is (by focus) the objective part of Experience. However the objectivity notion involved in this definition - in other words the limit between the objective domain and the subjective one - does not refer itself to anything objective, for the reason that the invariance from the subject is not observable in the Experience.

On the other hand it is possible to draw some rules for the constitution of objects. In fact this constitution is inherent to the intersubjective constitution of the discourse, during which subjects establish agreements on the identification of some experiences across observers (or subjects) [following here K. Popper], without which no intersubjectivity would be possible. As a consequence as pointed out by Husserl in his critics of Frege's foundationalism, one can not constitute mathematics without recourse to the Experience (in other words the constitution of mathematics is intrinsically intersubjective): what matters is that their character of reliable knowledge, which comes from the intersubjective stability of objective designations. Furthermore objects acquire their externality - which is specific to them compared to more general areas of Experience - through this operation of identification. As well similar agreements are established (implicitly) on (transcendental) operations that can be done on experiences. The collection of all the possible experiences constructed out of the action of a series of these transcendental operations on objects form the most general

horizon of a universal intersubjective discourse.

Despite this, because of the non-objectivity of the objectivity notion itself, the transition from the intersubjective stratum to the collective one has to involve, in a sort epistemic social contract, the delimitation objectively and explicitly a subdomain of the whole objective domain: this is what I would call the **localisation principle**. The various possibilities of subdomains and classes of subdomains correspond to the disciplines and subdisciplines of the collective discourse.

For instance the domain of (plane) Euclidian geometry consists in a set of experiences obtained by extraction. These experiences can be characterized as visual experiences of a colorless bounded plane surface scripted with a finite collective of geometrical objects. These objects are the ones constructed out of points, lines and circles, displaced arbitrarily by the (motor) operations of rotation of the surface, its translation and other operations of this type. As a consequences these experiences lie in the set of possible experiences enriched with transcendental operations (including geometrical ones). For another instance, an import part of experimental physics consists in an agreement on its domain, according to a criterion of extractability from experiences ensuring certain control conditions (on parameters involved in an experiment for instance), as well as on the content of these experiences.

2.2.2 Consciousness and the limits of the scientific domain

Before continuing the analysis of mathematics as a phenomenology, I would like to notice that by this localisation principle, objectivity in the mathematical sense can not include consciousness in its domain, for the reason that consciousness, by nature, transcends every possible object. As a consequence, a science (in the current sense of an objective study) of consciousness is not possible. However one can conceive the possibility of a reliable knowledge about it. It is possible for instance by projecting onto the phenomenon and explore in depth each of the objective progressions made out of this projection by an explicit agreement on each of the considered objective enclosures of the phenomenon, until it is possible to progress towards another more precise enclosure, and so on. Considering consciousness as the relation between some subject \mathcal{I} and his or her Experience, it seems reasonable to restrict the possible subjects to the class, which I shall denote \mathcal{J} , of the ones who are able to act on one's own Experience, model it into a series of experiences in a possible set of experiences enriched with transcendental operations and report its exact content in an intersubjective dialogue with any other subject of this class (in particular they can possibly agree on them). In fact these conditions are implicitly assumed by K.Popper's falsifiability criterion, since a proposition can be falsified only if anyone has the possibility to point out an experience in which the direction in this experience towards which a proposition points to is not present in it - in other words the proposition is not true for this experience.

This clarification is an important matter: despite the fact that the current objective domain - in other words the union of all the localised domains - is included in but different from the universal objective domain, these two domains are often identified by subjects whose attention is focuses on the current objective domain in such a way that they do not perceive anything outside of it, in particular as a consequence of forgetting the contractual (based on agreements) foundation of objectivity. This objectivist distortion of reality leads itself to a form of dogmatism in which an intersubjective will acts by artefact and by the authority its intersubjective unterm (without agreements) phenomenological progression, upon the representations of other subjects.

A similar phenomenon can be observed in the transition from the individual level to the intersubjective one. In this context the staticity of a designation at the intersubjective level is based on an agreement which results from dynamics of co-negation and co-position of subjects onto the static character of individual designations. As a consequence an individual designation has potential for staticity at the intersubjective level when it is static at the individual level, and this can be corrupted by an internal form of dogmatism (unless the subject is self attentive to the possibility of self-occlusion of this sort).

The potential of an individual designation to be static at the intersubjective level depends on its staticity at the individual level, which can be corrupted by an internal dogmatism (unless the subject is self-attentive to the possibility of self-occlusion by the internal language). In the

transition from the individual to the intersubjective level dogmatism often consists in the transmission of internally dynamical designations as static ones, preventing co-negation or co-position from the other on these representations and dynamicising by artifice the other's representations - this instead of engaging in an intersubjective dialogue of co-action on respective representations.

Considering consciousness, such a corrupted process has no reason to produce knowledge about consciousness as such but only an intersubjectively synchronised partial representation of the current form of one individual's consciousness (which has no collective interest). On the contrary the formal framework proposed in the last section of this text does follow this observation by including intradiscursively the dynamical process of intersubjective agreement, in a formal way.

2.2.3 On dualism

The focalisation of attention on the objective world can manifest itself in some other forms, including dualism, understood as the differentiation in essence of the objective and subjective parts of Experience. In other words dualism consists the staticisation of this partition, which prevents in any growth of the objective domain and in particular any progression of science towards consciousness. In this essentialisation the subjective part is often identified with the more general notion of phenomenal experience, forgetting that the objective part also does participate in the experience of the subject. This observation leaves open the possibility to render consciousness as such objectifiable, in a sense to be determined. In this project one has to go beyond dualism as well as forms of anti-dualism which simply deny the existence of the limit, in other words the existence of anything objective or subjective against the intuition: in fact this limit exists but its conception is dynamical.

While natural kind approach of consciousness (for instance in the approach proposed by T.Bayne) relies on ways to correlate objective part of the Experience it to subjective ones in their current meaning [see for instance J.Searle's notion of the existence of ontologically subjective and epistemically objective parts of Experience], building on an artificial partition of the Experience, I propose to understand better what objectivity means by abstracting the properties of areas of Experience that are specific to objects. One could use eneralisations of these properties as well as an abstraction the ways this domain grows to make it grow in the direction that one chooses. In this sense I follow Kant:

"A great part, perhaps the greatest part of business of our reason consists in analyses of the concepts that we already have as objects." - **I.Kant, Critique of pure reason, p.129.**,

and analyse how mathematics historically did extend their domain. I make this analysis through a partition of mathematics into three following types I identified: problematic, analogical and reflexive mathematics. Although implemented in different ways in each of these types, the phases separation strategy follow a common line. This abstraction will be useful in the following in order to elaborate a strategy for the growth of the objective domain in direction of consciousness.

2.2.4 Analogical mathematics

Amongst all types of mathematics, analogical mathematics are the closest in the common intuition of "discourse about Experience": in order to explain a phenomenon P_1 by its relation with another one P_2 , the method consists in finding in the Experience other experiences that appears in the intuition to be close enough respectively to P_1 and P_2 (compared to other experiences that \mathcal{J} can imagine having) which can be described mathematically (in other words, they statical designations). After that one relates these two experiences by a sequence of objective operations, producing the second out of the first one. Because of all these closeness relations, this sequence of objective operations does convince that there is a causal relations between the two phenomena P_1 and P_2 , and this is convincing enough to serve, even at the collective level, as a guiding predictive principle (and not a proof) towards an objective experience of the relation between P_1 and P_2 . This is in this sense that one can interpret J.Searle's words:

"Science preserves the appearance while giving us a deeper insight in the reality behind the appearances." - J.Searle, **The Mystery of consciousness**, p 111.

The appearance is preserved in the sense that in the intuition, the experiences that we compare are close enough to be identified (at least for a beginning). Let us consider some examples for two types of analogical identifications I identified: the analogies by morphological similarity, and by concept tokening.

2.2.4.1 The physical basis celestial bodies movement The theory of universal gravitation, which was developed by the mathematician I.Newton, consists in a mathematical system describing the dynamical behavior of a massive body, such as a planet, in relation with the forces applied on it by other massive bodies. This theory was recognised in particular after the discovery of the planet Neptune using knowledge about Uranus movement. Despite the singularity of this revolution and its impact, it relied heavily on the earlier discovery of J.Kepler related to the form of planets movement (whereas the universal gravitation theory provides an explanation of this movement which relies on the concept of force) around the sun: they follow an elliptic path. This discovery was the conclusion of overcoming the psychological barrier of the association - already present in Plato's Timaeus - of astronomy with the function of *representing the visible motion of celestial bodies by a combination of uniform rotations on circles*.

Remark 1. *I believe that the omnipresence, nowadays, of this form of physics is responsible for forgetting that the fundamental discovery of I.Newton was based on J.Kepler's one. Together they form a theoretical progression that is decomposed into works of distinct natures - the first one is physical and the second one mathematical. There is no reason to expect that an understanding of consciousness as such would manifest itself through a similar progression, especially because of the dilution of scientific time, that I mentioned above: even if such an understanding should probably be found in relation with brain architecture, there is no guarantee that one can rely on the current knowledge (in particular even the accurate identification of some neural correlates of consciousness would not imply an understanding of how phenomenal experience is generated). Because of time dilution, one should focus on a reliable way to explain the structure of phenomenal experience as such with adapted mathematical tools, before any attempt of explanation in physical terms.*

The approach followed by J.Kepler relied on an analogy by morphological similarity between an experience which consists in the experiential measure of the sun's relative positions, earth and mars, and another one which consists in a geometrical model of this measure (the positions are assimilated with points). At this point, the problem is mathematical in the sense of finding a geometrical figure (abstracting oneself from the combinations of circles) which is the closest (with an intuitive notion of closeness which was formalised later) to the experience which consists in the set of position data (this is another analogy by morphological similarity). The statement of planets movement following an elliptic path is convincing because it provides an experience that one can compare with the one that comes from the measure of reality and identify them in one's intuition. It is constitutive (in Kantian sense) for it completes the actual experience towards the opening of the Experience (allowing the prediction of Neptune's presence for instance). The phases separation strategy is found here in the reasoning structure which separates (implicitly) mathematics of data and geometrical figures relation and the morphological analogy, which is dynamical (the closeness in the intuition depends on the experiential context and the development of knowledge on the subject).

2.2.4.2 Other examples Another example of morphological analogy use can be found in A.Turing's work *The Chemical basis of morphogenesis* (1952): here the analogy is made between patterns observable in the Nature - such as the Giant pufferfish's ones - and some idealised version of these, now called *Turing patterns*. A.Turing proved that these idealised patterns can be produced by a reaction-diffusion algorithm initialised on a uniform configuration of a set of different types elements (representing chemical ones). As a consequence he provided a hypothetical mechanism

through which the natural patterns could be produced out of chemical interactions. In this case one can also observe a paradiscursive separation of phases: the text consists in the static part itself while the analogical part is left to the periphery of the discourse (the intuition of the reader).

One can find other examples of morphological analogy use in mathematics in more recent examples and other application fields, in the modelisation of economical situation structure using game theory (developped by J.Nash) or the reproduction of phenomenal aspect of physical reality (such as the existence of phases for certain materials) in statistical physics.

The other type of analogy I mentioned (concept tokening) consists in the association of two experiences based on the intuition that they belong to a same class of experiences (for instance a tree belongs to the class "tree" which groups all the experiences of a tree). The use of this type of analogy can be found for instance in J. Von Neumann's work *The theory of self-reproducing automata*: he makes an association between the dynamics of a theoretical machine which reproduces itself (its dynamical behavior leads unavoidably to the construction of a perfectly identical machine) that one can think as a series of ideal experiences and the experience of living beings self-reproduction. Each of these series of experiences belong to the intuitive class of "self-reproduction": it tokens this class in the sense that the concept of this class becomes present to the mind when an occurrence appears.

2.2.4.3 Misuses of analogical mathematics A fundamental aspect analogical mathematics way to *formalise* reality is that its formalisation - based on an analogy between an actual experience and an ideal one - is able to point out of an operations sequence on the mathematical model to an extension of the Experience beyond the model itself (through another analogy). In other words such an extension of the Experience makes valuable the formalism, used or developped in the theoretical progression towards the reality of some particular experiences, despite the analogy's dynamical nature. On the other hand (although they are often considered as meaningful in themselves) isolated analogies (without any derived extension of the Experience), or even a coherent group of parallel analogies have no more meaning than the one they have as dynamical designations (as a step in the phenomenological progression towards a potential object). This is an important matter in a context in which the multiple theoretical attempts of a correspondance between phenomenal experience and architecture of the brain rely on such analogies.

2.2.5 Problematic mathematics

For mathematicians the development of mathematics often results from the solution of mathematical *problems*. A problem takes in general the form of an experiential statement of multiple possible types: for instance, is it possible to *construct* such or such object which satisfies such and such informal properties ? or can one find an exact (or simpler) expression for such or such formula ? or is such or such property true of an object ? or simply is such statement true ? In the same way, the origin of a problem can be of various types: it can be the extraction of an already solved problem from its original context and its transposition to another one or the simple intuition the statement truth, the possibility of a construction etc. In fact a problem is the product from the consideration and manipulation of already existing (statical) mathematical objects and concepts, in the intuition and at the periphery of this objective domain.

For a historical example one can consider the solvability of polynomial equations by the radicals method: while the possibility of this method's application was known when the involved polynomial has small degree (up until four) through distinct discoveries, no one was able to find possible a systematic application of this method when the polynomial's degree is greater than five. Out of this observation the mathematical problem came out to be formulated as follows: 'is it possible at all to apply the method by radicals to find solutions of a polynomial equations when the degree of the polynomial is greater than five ?'

While the majority of mathematical problems are formulated only in the individual or intersubjective strata (internal dialogue or informal discussions), only some of them access to the collective stratum, most of the time under the form of a conjecture, which consists in a bet about some proposition's truth. A conjecture is an equivalent of a pre-statical designation - in the vocabulary

introduced above - in the sense that one can not conceive any of the conjecture's definitive directions as a designation to be anything else than "true" or "not true". For a conjecture to access the collective stratum its dynamical character has to be stable (it is difficult to find a definitive answer): the reason is that a solution for such a conjecture should involve an extension of the mathematical domain, through the invention of non-existing mathematical architectures or the connection of a domain to another one (for instance), and imply solutions of other minor problems and the development of new mathematics.

In problematic mathematics the text itself - the part of its discourse which lies in the collective stratum - consists mostly in the statical phase (problems solutions): except for conjectures (explicitly stated as such in the text), as well as the (dynamical) relation with conjectures or informal problems (although they are not the object itself of the text) which motivate the study and are explicitly separated in the introductory or conclusive parts of the text, the dynamical phase is kept outside of the text. Despite that the dynamical phase is fundamental for the practice of mathematics and their evolution, in particular through regulative principles such as computability (one should follow directions of research where non trivial statements or computations are possible), but also in the search of a solution itself. Although never systematised as such, the progression towards a solution is phenomenological in its structure, in the projection (analysed before) onto the intuition of the articulation of already existing mathematical concepts, whether they are collectively valued or only in the restricted individual or intersubjective domain of research. In particular, the discourse's dynamical phase acts during the search for a solution by restriction (in a short term process) of the set of concepts which could be useful in this search. In a sense, the mathematical domain grows in the consideration of mathematical objects, and the object which underlies the intuition coming out of this consideration lies in the consideration. However this extension is of different nature from the transcendental act of consideration (as statical) and thus can not be identified with it.

In the case of the problem of solvability by radicals for polynomial equations, the solution came out as a consequence of E.Galois's theory, which consists in the development of notions such as the symmetry group for a polynomial equation's set of solutions and statements related to these notions and solvability by radicals. Here the most important output of the theory, despited motivated by the problem, is the creation of concepts along the theory's phenomenological progression.

The practice of mathematics reveals another way to see that the mathematical text is of statical nature in the sense that I determined above: in fact the proposition of a solution to a problem, stated as the proof of a theorem, is considered to be as an actual solution (and scripted in the collective stratum) when one can not detect in it any error, in particular in the logical architecture of the proof. In other words the truth statement for this proposition, equivalent to the negation of error's presence through time, is statical as a designation.

2.2.6 Reflexive mathematics

Sometimes the extension of the domain of mathematics results from the formalisation of mathematical practice itself (as particular cognitive dynamics). For instance the conception of a mathematical proof as an articulation of logical relations (Frege's logics), of a general mathematical object as a set of elements (Cantor's set theory), the notions of computation (A.Turing's computing machines), and object construction (Category theory) participate closely to this practice at the cognitive level. Historically the extension of the mathematical domain in this direction was involved in the search for an understanding of the limit of mathematics progression (in particular in the existence of undecidable problems such A.Turing's halt problem) but also in the solution of some problems (for instance Hilbert's Xth problem (Entscheidungsproblem), solved also by A.Turing) questioning the algorithmic solvability of earlier problems. The concepts constructed this way come out of the mathematician's self-observation in his or her manipulation of already defined mathematical objects.

For instance, the notion of computing machine introduced by A.Turing was conceived as capturing the general notion of computation - meaning the production of a mathematical object out of another one using a sequence of mathematical operations (for instance the simplification of a

formula into another one or equivalently the evaluation of a function on a particular parameter). The definition of these computing machines is a simplification of cognition's fundamental structure in a first person. Interestingly enough it is also a simplification of the computation process from an exterior perspective: in fact it is probable that A.Turing's intuition in which this definition is rooted consisted in the abstraction of the behavioral pattern of human *computers* (who were employed at this time to execute computations, determined by scientists, on a long paper tape).

The collective value of this formalisation comes however from the solution of Hilbert's Xth problem (which implies semantically the formalisation of the notion of computation) and afterwards from the identification of the general notion of computation with the process of a computing machine. This identification (formulated as the Church-Turing thesis) is derived from the projection of these two concepts (an purely intuitive one and a mathematical one) onto the current domain of mathematics: for a mathematical object, one projects the intuitive notion of computation in order to determine if this object consists in a computation and then projects the notion of computing machine on the verification that this algorithm can be simulated by a particular computing machine. One can then observe that the two concepts coincide on some subdomain of the current mathematical domain. As this subdomain grows with the number of tests for this coincidence, one gets convinced that they are equivalent. However it is clear that the nature of the concepts involved in this reasoning prevent this equivalence from the possibility to be proved mathematically. Instead the intuition of the equivalence constitutes here a definition of the purely intuition notion by the mathematically defined one (definition which is commonly accepted and used in practice, in other words it has accessed the collective stratum).

Reflexive mathematics expand in general by means of such a double projection, the equivalence of whose terms has epistemic force correlated to the width of its validity domain. The equivalence statement is statical when this domain grows indefinitely. Interestingly enough one can interpret the acceptance of this equivalence as the recognition of its character as reliable knowledge, despite its non-mathematical character - which illustrates the interest of the notion of statical statement. In this type of mathematics the phases separation strategy is manifested by the presence in the text of the formal notion under the form of a definition while the coincidence with the intuitive notion is left outside of the text's core.

Reflexive mathematics are fundamentally different from analogical ones: despite the fact that they both consist in the constitution of relation between intuitions and mathematical objects, the nature of these relations are of different natures. In analogical mathematics these relations are direct and based on phenomenological closeness, while they are indirect (by double projection) in reflexive mathematics. However they are similar in the existence of recurrent misuses, for instance when the domain of the equivalence is extended without rigour (although on a rigorous base domain). In order to contribute to the progression of reflexive mathematics, one needs to understand the nature of these mathematics and prefer the rigorous application of their principle to another domain of Experience to an artificial extension of its current domain: in my opinion there should be a shift from the consideration of 'active' cognition to 'passive' one - in other words perception, or how are unconsciously constructed the object that the subject perceives as such in his or her Experience. This shift implies a change in the method, according to the principles.

2.2.7 Conclusive remarks

The purpose of the mathematical typology I displayed in this section was to observe that across their difference in the mode of extension of the mathematical domain, they follow a common abstract line which is similar to the one of the authors' Phenomenology: the difference lies in the restriction of mathematics to the formal. In particular the phases separation strategy is a constant of this typology, and the limit is paradiscursive, despite the importance of the dynamical phase in the practice of mathematics, as well as their development and meaning. In the progression towards perception and objectness (and thus consciousness as such) one needs to shift this limit intradiscursively, while keeping the phases separation.

3 Around phenomenal objectness

In this section I will get to the practice of consciousness theorisation. The main goal is to build a framework in order to study first-person perspective objectness (meaning the distinction in an a priori undifferentiated experiential background of particular areas of Experience), which is a fundamental aspect of human consciousness (\mathcal{J} am conscious if \mathcal{J} can distinguish things in my Experience), in a systematic and reliable way. For this purpose I rely on the history of consciousness studies in two ways. I shall first analyse a recent theory about consciousness mentioned in the introduction, namely the integrated information theory, in particular its foundations and some of its developments which attempts to originate phenomenal spatiality in the properties of some brain mechanisms. The purpose of this analysis is to notice that this theory suffers from important epistemic difficulties: while successful analogical theories meet a trade-off between closeness of the formalism to direct introspection (in other words the ability of the formalism to describe reality) and the tractability of this formalism (the possibility of progress of mathematical understanding on the formalism itself) this one does not meet it: in other words it is epistemically sterile. As this kind of difficulties can be expressed using the vocabulary introduced above (in particular the possibility of combining phenomenology and mathematics), one can see this section as an application of the general considerations of this first part. More precisely the current form of integrated information theory can be seen as a 'knot' of statical and dynamical designations which, despite this, is upheld by defenders who, "dogmatically-believing", think that it is their "business to defend such a successful system against criticism as long as it is not conclusively disproved" (K.Popper) [in particular, and significantly, through the use of coercive persuasion]. After exposing some intuitions on the way to 'unknot' the dynamical and statical phases of this discourse, using in particular the combination criterion defined in the first part and some systematisation of the theoretical orientation defined by D.Dennett in [Dennett], I expose the framework mentioned above, in which the statical and dynamical phases are explicitly separated and which meets a trade-off between closeness and tractability (as it allows a theoretical progression in its direction).

3.1 A knot of statical and dynamical designations

3.1.1 Statisation of the structure of Experience into an *axiomatic* system

A central aspect of the integrated information theory is the construction of a fundamental system which is meant to be roughly, on the model of mathematical axiomatic systems, a set of reliable discourse elements (called 'axioms' here) out of which the theory's discourse is constructed (whether these elements do belong to this discourse or are tools for its construction). The axiomatic character of this system has been recently questioned, in particular by the philosopher T.Bayne. After I expose this fundamental system in the present section, I review T.Bayne's critics and some of the counter-critics. In fact both of them rely on unquestioned conceptions about axiomatic systems, are not accurate enough for the eye of a mathematician. I propose instead to think about the elements of the theory's fundamental system in the abstract terms of their function in the discourse, as well as to use the terms introduced in the first part: in this terminology one can see the denomination of the fundamental system as an 'axiomatic' one is a first example of artificial staticisation in this theory.

3.1.1.1 Presentation: For integrated information theory, an axiom of phenomenal experience (or consciousness) is a determined as "self-evident truth about consciousness" ([p.2, l.104). Altogether they are "meant to capture the essential properties of experience (phenomenal existence)". A remarkable aspect of integrated information theory's 'axioms' is that although their number and identity (by naming) are stable across publications, their formulation changes with time. I reproduce below a list of the axioms and for each of them provide two formulations: the first one appeared in [?] (2014) and the second one in [?] (2017):

1. (a) **Existence:** *Consciousness exists – it is an undeniable aspect of reality. Paraphrasing Descartes, "I experience therefore I am."*

- (b) **Intrinsèque existence:** *Consciousness exists **intrinsically**: my experience is real – indeed, that my experience here and now exists – for example, I see my bedroom – is the only fact I can be immediately and absolutely sure of; moreover, it exists for me, from my intrinsic, subjective perspective.*
2. (a) **Composition:** *Consciousness is compositional (structured): each experience consists of multiple aspects in various combinations. Within the same experience, one can see, for example, left and right, red and blue, a triangle and a square, a red triangle on the left, a blue square on the right, and so on.*
 - (b) **Composition:** *Consciousness is structured: each experience has internal structure, being composed of phenomenal distinctions, bound together in various ways, which exist within it. Thus, within the same experience, I can see different locations in visual space, different colors, different objects, objects of particular colors at particular locations and so on.*
 3. (a) **Information:** *Consciousness is informative: each experience differs in its particular way from other possible experiences. Thus, an experience of pure darkness is what it is by differing, in its particular way, from an immense number of other possible experiences. A small subset of these possible experiences includes, for example, all the frames of all possible movies.*
 - (b) **Information:** *Consciousness is specific: each experience has a specific form – a particular composition of particular phenomenal distinctions, bound together in various ways, and thereby differing in its specific way from other experiences. Thus, my experience of the bedroom here and now is just what it is, containing a bed, a body on it, a bookcase, a blue book on the left shelf of the bookcase, and so on. Moreover, being that way, my current experience necessarily differs from other experiences, containing other objects and colors, sounds and smells, or an experience of pure darkness and silence and so on.*
 4. (a) **Integration:** *Consciousness is integrated: each experience is (strongly) irreducible to non-interdependent components. Thus, experiencing the word “SONO” written in the middle of a blank page is irreducible to an experience of the word “SO” at the right border of a half-page, plus an experience of the word “NO” on the left border of another half page – the experience is whole. Similarly, seeing a red triangle is irreducible to seeing a triangle but no red color, plus a red patch but no triangle.*
 - (b) **Integration:** *Consciousness is unitary: each experience is irreducible to non- interdependent components. So are the phenomenal distinctions within an experience, as well as their relations. Thus, I experience a whole visual scene, not the left side of the visual field independent of the right side (and vice versa). Furthermore, I see the object lying on the table as a book, not as a set of features. Finally, seeing the book as a blue book is irreducible to seeing the color blue without the book, plus the book without the color blue.*
 5. (a) **Exclusion:** *Consciousness is exclusive: each experience excludes all others – at any given time there is only one experience having its full content, rather than a superposition of multiple partial experiences; each experience has definite borders – certain things can be experienced and others cannot; each experience has a particular spatial and temporal grain – it flows at a particular speed, and it has a certain resolution such that some distinctions are possible and finer or coarser distinctions are not.*
 - (b) **Exclusion:** *Consciousness is definite, in content and spatio-temporal grain. Thus, my experience has just the content it has, neither less – say, excluding colors, nor more – say, including awareness of blood pressure. Similarly, my experience flows at a particular speed – each experience encompassing a hundred milliseconds or so – neither faster nor slower.*

3.1.1.2 Critics and counter critics of integrated information theory's *axiomatic* system

In his article [Bayne], T. Bayne provides a detailed and specific critic of each of the elements of the above system. His claims is the following: *"I argue that none of the five alleged axioms is able to play the role that is required of it, either because it fails to qualify as axiomatic or because it fails to impose a substantive constraint on a theory of consciousness."* This claim is based, roughly, on the fact that to each axiom correspond multiple possible interpretations. While some of these interpretations are implausible since they are evidently too 'constraining', others involve terms (such as *phenomenological distinctions*) that are too loosely defined. The interpretation of these terms (for instance, concerning their nature, for phenomenological distinctions) make the statements tautological, in other words not 'constraining' at all. Other interpretations of the statements are constraining but fail to be axiomatic for the reason that there is no general acceptance of these interpretations by specialists (which consists here in a definition of what 'axiomatic' means according to T.Bayne).

It is important to note that the interpretations proposed by T.Bayne are only ones of all the possible ones: as a consequence it is possible to conceive the existence of some interpretations at the same time axiomatic and constraining. The terms that are loosely defined in the statements have at least two types of possible interpretations: a constative one (which means that it consists in the only observation of existence of the phenomenon) and a hypothetical one, which posits a possible explanation of the phenomenon or a mechanistic characterisation. Formulated in the constative version the statements may appear as tautological, unless we differentiate in it, as a designation, the experiential part (that which in the Experience is pointed at) and the concept which corresponds to it. As a matter of fact we tend to think about language on the model of technical languages such as mathematics (in particular Euclidian geometry) for which there is a one-to-one categorial mapping between areas of Experience and their concept, leading to their assimilation.

In fact to the concept of consciousness corresponds an area of intuition which is so *large* that everything that (right now) can be said to be true or false about it is trivial and as a consequence can not constrain any theory. Anything else is therefore a hypothesis and thus is certain to divide specialists. As a matter of fact because of the nature of the phenomenon of consciousness and the state of our knowledge about it, anything is *apodictically* true if and only if it can not impose a constraint, and this equivalence can be derived analytically (withtout referring to a theory in particular) and T.Bayne's critic of integrated information is merely an illustration of this equivalence.

On the other hand some of the counter critics appeal to the idea that Euclidian geometry's axioms (paragon for axiomatic systems), was not stable through history and replaced for instance by modern axiomatic systems in which the parallel postulate does not appear to be axiomatic (development of non-Euclidian geometries during the XIXth century, Hilbert's axiomatic system for Euclidian geometry). For these counter critics the consequence is that axiomatic systems can be - in the vocabulary introduced before - composed with dynamical designations, just as integrated information theory's system is, neutralising a priori critics which based on the non-evidence of the truth of this system. However these counter critics do not stand under close examination, for multiple reasons: first, that non-Euclidian axiomatic systems are concerned with another experiential context - precisely geometrical objects that are different from the ones embedded in the Euclidian plane - whose construction was motivated by the question of the existence of non-Euclidian geometries (where in particular the parallel postulate is not verified). The consequence of these constructions was the possibility to think about physical space in other terms than only Euclidian ones. Second, that the modern axiomatic systems (such as Hilbert's, Tarski's and Birkhoff's ones) do not invalidate Euclidian system in its truth. In fact their developement can not be considered as a refinement of Euclidian system meant to replace it. For instance the purpose of Hilbert's system was to optimise the axiomatisation of Euclidian geometry by grouping the axioms in types

and understanding the way they are involved in the proof of theorems in order to prove the system's completeness - every true proposition derives logically from only axioms. The actual proof of Hilbert's system completeness was made later by Tarski in the book *The completeness of elementary algebra and geometry*. After that Tarski constructed his own system of axioms in order to write Euclidian geometry in terms of quantifiers, allowing him to prove that this geometry is decidable (every proposition can be proven to be true or not). Regarding Birkhoff's axioms, they were meant for pedagogical purpose, in particular the possibility to check the truth of a proposition with students' tools (protractor and ruler).

Furthermore these modern axiomatic systems came in the context of a foundationalism, whose purpose was to ensure the complete objectivity of mathematics, in a stronger sense than its intersubjective agreement interpretation: in particular for foundationalists mathematics should not rely on the use of direct introspection, in particular visual one, while Euclidian geometry relies on this sense, for instance in the recognition of a geometrical point (its undividedness). In fact this recognition consists in a transcendental operation that can not be translated in machine terms (and thus is not objective in the sense of the human-independent universal observer). However the difficulty of this interpretation of objectivity is that it relies on the viability of the machine, itself evaluated upon the regularity of the agreement of humans about the machine.

As a matter of fact, integrated information theory's system not only fails to be self-evident but also to find a shelter in modern axiomatic systems, notably by standing in contradiction with the reason of their existence: the possibility to prove completeness, decidability, or even the mutual interdependence of its elements.

3.1.1.3 Dynamical designations can not be axioms: As observed in the last section, some of the terms used in integrated information theory's system, such as *phenomenal distinctions*, are not pointing at precise objects in the Experience (precisely for no definition of *phenomenal distinctions* has been proposed on which the author and the reader can agree on, at least tacitly). As a consequence they are dynamical designations.

Their dynamical character can be seen despite the stability of the system (identity and number of the statements) in the evolution of its formulation through the publications. For instance the **Existence** statement, according to which consciousness existence follows in deduction (if these two statements are not identified by the authors) from the existence of *I*, is reformulated in the second version into **Intrinsèque existence**, for which *my experience* exists and it exists *for me*. The statement of **Composition** also undergoes some transformation: from that the experience *consists in aspects* to that it is *composed of phenomenal distinctions*, that are *bounded*. This is an important shift since the term 'aspects' is associated with passivity (the aspect presents itself) while a 'distinction' is associated with an action. The meaning of the statement is by this shift changed in a fundamental way. The first version of the statement of **Information** is centered on the relation of one single experience to other possible ones, where the second version is centered on the experience itself, its "form", and then differs from other experiences by having a different "form" (which is reflected in the term used to specify the 'axiom' in the first sentence: "informative" vs "specific"). A similar semantic shift occurs in the statement of **Integration**: from "integrated" to "unitary". Moreover the second version makes the precision that this "integration" concerns also phenomenal distinctions. Similarly, the axiom of **Exclusion** shifts from "exclusive" to "definite"; this shift concerns the suppression of the mention of that there is only one experience and not a superposition of experiences.

Unless the nature of this fundamental system is made clearer it falls under Wittgenstein's analysis of metaphysical statements, in agreement with Kant's more specific denial of philosophical propositions from the possibility to have axiomatic character:

"Now since philosophy is merely rational cognition in accordance with concepts, no principle is to be encountered in it that deserves the name of an axiom. Mathematics, on the contrary, is capable of axioms, e.g., that three points always lie in a plane, because by means of the construction of concepts in the intuition of the object it can connect the predicates of the latter a priori and immediately". - Immanuel Kant, Critique of pure reason, p. 640.

Interpretation: According to Kant, what differentiates axioms from any philosophical statement is that they are constructed *in the intuition of the object*. In fact this relation of the designations to the intuition is possible for the reason (mentioned before) that since there is a one-to-one categorial mapping between concepts of Euclidian geometry and areas of experience they are pointing at, one tends to assimilate the concept with what it does point at as a designation. Because of this identification, one can consider that the concept is constructed in the Experience (or equivalently the intuition). Moreover, one can connect the predicates (or propositions, articulations of concepts) in the intuition through this identification.

Following Kant, I consider that integrated information theory's system is no actually **axiomatic**. This is fundamental, since the role of axioms is for a theory to be founded on them, while if one bases such a theory on dynamical designations, the whole theory can crumble down at any time. As a matter of fact, the characterisation of this system as axiomatic is a rough attempt to dogmatically staticise a set of dynamical designations in the project of expand it at the collective level. Despite this one can make sense of the self-evident truth of these statements, by differentiating in the designation a reference of the *truth* character to the formula and a reference of the term *self-evident* to the area of experience it is pointed at. A dynamical designation can be good enough for one person's mind for the reason that by neglection it objectifies the area of Experience that the designation is pointing at, or because the cognitive context in which this designation is self-made is enough to perceive the truth of the statement. In this sense a designation can be self-evident at the individual level and not at the intersubjective one. For this one has to refine it sufficiently so that it allows the possibility of an intersubjective agreement and then an intersubjective self-evidence judgement, without which the individual self evidence and the self-occlusion can not be differentiated.

3.1.2 Transcription of the 'axioms' in transcendental terms

In the spirit of impartiality I attempted to make sense of integrated information theory's fundamental system in a way that would avoid tautology and hypothesis. Before exposing my interpretation I need to make explicit some transcendental 'framework', which is implicit (and in particular unquestioned) in integrated information theory, and on which it is built.

3.1.2.1 A transcendental 'framework' for the theorisation of Experience: Temporal series of experiences:

I order to talk about the Experience - that I think of as the totality of appearances, meaning what is present to me without effort or process - one usually talks about an *experience*, which consists in a temporal section of the Experience (meaning everything that one experiences at a fixed time), or a part of it. Of course an experience is an abstraction which refers to a non-identical intuition. Sometimes (including time in consideration) one also thinks about the Experience in terms of oriented series of experiences, which are identified with the Experience in the intuition. This model is convincing for the reason that when the series of experiences is sufficiently dense, one cannot differentiate it from the area of Experience that it is meant to model (this is the principle of movies for instance, but also cinematics) [however there is no evidence that the identity of this sequence of experiences with the corresponding area of Experience is stable under direct introspection]. This does not mean that the whole of Experience follows this schema (an abuse of a modelisation). In particular, one can see in Husserl's ideas of protention and retention a contradiction to this phenomenal time linearity: at the phenomenal microscopic level, past and future are intertwined in a non-trivial way.

On the other hand, this model is good enough (in the same way as Newtonian physics are still considered as good enough in particular conditions despite the advances of Einsteinian physics) for some areas of Experience, for instance when one cannot notice any differentiation in it: when one extracts from it temporal sections one does not differentiate these temporal sections and as a consequence identifies (up to perception) these experiences. One can thus identify the considered

area of Experience to any extracted series of experiences in which time differentiation is not perceived.

The set of possible experiences:

The approach of integrated information theory is also founded tacitely on the notion of *set of possible experiences*. Besides the absence of guarantee that the totality of Experience can be thought in terms of experiences, there are multiple indeterminacies in this expression, which correspond to many ways to interpret them. A first indeterminacy is on the subject of these experiences: this set (if properly defined) can depend on the subject, who could be the reader or the author, or a minimal subject of experiences common to all human beings (if properly defined). Another indeterminacy is on what "can have" means. This could refer to the experiences that one would experience in particular physical conditions (third person perspective), or to the experiences that one can imagine having (first person perspective). In my view it is reasonable to define the set of possible experiences as the one that \mathcal{J} - designating here indeterminately any subject in a group in intersubjective interaction - can imagine having (in a first person perspective), so that any statement made can be falsified by one of the subjects in the group, by direct introspection in the frame of agreed terms. The collective generality of the statements such a group can make can only be evaluated in comparison with other groups (in other words, one cannot generalise straightly as in objective frameworks).

The definition of an enriched set of experiences

With combinations in series of transcendental operations (finite transcendental algorithms), one can construct experiences that are not extracted from the Experience: for instance, by twisting multiple squares cutted in the plane and gluing them together one can build what are called *manifolds* in mathematics : a remarkable non observed example (in the actual pre-theorisation Experience) being the Klein bottle ; another simpler and observable one is the torus. As a tool for the theorisation of Experience, one can (and usually does implicitly) rely on the definition of an **enriched set of experiences**, which consists in the set of possible constructions out of transcendental operations on possible experiences. Its purpose is to set a framework in which one can attempt to find an explanation of the form of the actual set of possible experiences, which is conceivable since it is possible to relate in the intuition the elements of the enriched set of experiences with elements of the set of possible experiences, as argued in the following quote of David Wallace [] in the article *Worlds in the Everett interpretation*, about the expressive power of the many-worlds interpretation of quantum mechanics:

"So, as a description of spacetime the many-instants interpretation leaves much to be desired: it contains arbitrariness, obscures structural features, and we ourselves cannot be said to exist in any single given instant. Nonetheless it is a description with merits. Specifically, it is a complete description: once we have specified all the contents of all the instants of time and the temporal relations between instants, we have the entire spacetime. Further, we have an existing intuitive grasp of what an instant of time is like, and from this we can gain some understanding of what sort of entity spacetime is. This understanding is enough to make contact between our experiences and the formalism of relativity, granting the theory predictive and explanatory power." - **David Wallace**

In practice, one defines a subset of this theoretical enriched set of experiences, generated over the set of possible experiences by combination of a finite set of transcendental operations. For the integrated information theory's system, it seems that this enriched set is co-defined with the model of experiences.

3.1.2.2 A system of elementary transcendental operations Along with this framework one can see integrated information theory's system as the identification of a set of elementary transcendental operations, constituted into a system for the purpose of constructing a transcendence model (in other words the extraction and processing by the \mathcal{J} of information contained in rough experience). For each of these elementary transcendental operations are of constitutive nature, I

shall call them **articles** of the model. I also made the choice to rename these articles in line with their transcendental nature.

The first article states the **reality** (corresponding to the 'axiom' of existence) of consciousness, understood as the transcendental relation of \mathcal{J} to the Experience, which modeled as a series of experiences such as described in the last section. The purpose of this article is to designate the direction of interest for the investigation. Let us note that it is possible to go further with a hypothesis on the nature of reality, for instance that the experience that \mathcal{J} have is causally related to the objective world (this relation is the ultimate aim of this study), more specifically to the part of me which belongs to this world (my body). The article of **distinction** (corr. to composition) states that \mathcal{J} can distinguish 'things' in any experience, thought as areas in this experience, whether they are identical or different from it. In the words of Husserl: "*All consciousness is consciousness of something*". As a matter of fact, consciousness in quotidian experience is the consciousness of multiple things that one identifies in the experience (whatever is the nature of this thing: a feeling, an object, a thought, etc). The distinction of multiple things in an experience (in other words at the same time) and the further analysis of their relations implies the possibility to display these things in a spatial frame (this corresponds to the Kantian intuition form in which things appear, which models the spatiality of an experience). The article of **exhaustion** (corr. to information) states then that the sequence of spatial displays of 'things' exhausts what experience is, in the sense that an experience is exactly how it differs from other experiences, and it differs by what it contains. Although this article is similar in its formulation to the axiom of intentionality in sets theory, these statements are of different nature - the axiomatic nature comes from the fact that intentionality derives from the definition of the framework (the definition of sets). Moreover it is involved in the discourse about experience in the sense that it allows the comparison of an experience to another one in the model. Also it implies the deterministic character of experience (in particular it is not strictly probabilistic). The article of **non-separability** (corr. to integration) relies on the operation of separation of an experience into two parts (according to the spatiality of things display). This assumes the use of a more abstract spatiality, where the place of a things *is* its identity. The article states that an experience is different (in the sense that one does not recover the same things in these experiences through exhaustion) from any experience obtained by a non-trivial separation (for instance a blue sphere is different from concatenation of the color blue and the form of the sphere). It is different from the hypothesis of integration, made in integrated information theory, that there exists a physical phenomenon which explains the non-separability (as a 'force' opposed to the action of separation), in particular for the reason that integration can take multiple forms, which is not the case for the non-separability. Finally the article of **non-extractibility** (corr. to exclusion) uses the operation of *suppression*, acting on experiences that are under separation by forgetting one of the after separation parts. I shall call the composition of separation and suppression an **extraction**. The article states then that an experience is different from any experience obtained by extraction from it (in particular one can derive that a possible experience is different from any experience that *extends* it).

Comments:

Although this constitutive system satisfies some of the properties which are required for the foundation of a theory - for instance that they are coherent - this does not derive in the same way as for an axiomatic system: in this case, it derives from the constitutive nature of the system, while in logics this comes from the impossibility to derive two contradictory propositions by the combination of the axioms. Moreover the statement of self-evidence is misleading for the reason that it leads to the (usual in mathematics) identification area by area of the in-construction model of Experience with the Experience itself, and the illusory tautological character of statements which have constitutive nature and of which it does not make sense to talk about truth (as evoked earlier).

3.1.3 Further staticisation by projection in a fixed formal context

After theorising the Experience into a fundamental system, the direction that the theory takes is to search for a connection between the structure of Experience as such and the objective world (in particular with brain mechanisms) in order to find in physics the origin of phenomenal experience and explain its structure out of the brain mechanisms properties. Its strategy is to produce a mathematical version for each of the elements of its fundamental system in order to realise this connection through the medium of mathematical language. In the following I analyse the mode of production ('formalisation' according to integrated information theory's proponents, amongst other terms such as 'translation' or 'derivation') of these mathematical versions and the epistemic difficulties it generates, in particular when one tries to go in the inverse direction (from the formalism to reality), for instance accounting for aspects of reality (such as phenomenal spatiality) in the constructed formalism. Besides the manifest dynamicity of the designation of the mode of production itself, it is possible once again to interpret these difficulties in terms of statical and dynamical discourse phases.

3.1.3.1 Projection of the system in a fixed formal context The current modelisation paradigm for brain architecture is currently focused on neural networks, which are modeled as finite probabilistic dynamical systems. In this formal context a physical system such as the brain is considered as the collection of a finite set of units, where each of these units' state is considered to be in a finite set of possible sets, each of which are represented with a symbol. The state of the whole system (meaning the collection of its units states) is updated regularly according to some transition probability measures. In other words every update consists in changing the state of every unit randomly in a way such that the probability to change the unit's state is determined in advance and depends on the current state of all the other units. In particular learning neural networks machine learning and artificial intelligence belong to this formalism. The success of these networks is correlated with the impressive achievements in automatic learning and elaboration of game strategies and is the reason why we tend to assimilate human intelligence and brain architecture with this neural networks formalism. Despite the progression in understanding of the glia's importance for human cognition, this identification has still an important impact on brain modelisation. As a matter of fact, integrated information theory follows this schema.

The 'formalisation' of the fundamental system begins with the 'formalisation' of its first element (the 'axiom' of existence) using another dynamical identification between reality (or existence for a subject) and causality known as the Eleatic principle:

Eleatic principle[[Colyvan]]: **An entity is to be counted as real if and only if it is capable of participating in causal processes.**

The interpretation of this principle in integrated information theory replaces the word 'real' with 'existing' and 'participating in causal processes' with 'having a cause and an effect'. The interest of this identification is that it is possible to find an interpretation of causality (in terms of causes and effects) in the context of finite probabilistic dynamical systems (in particular there exists already some formalisation for causality developed by the mathematician J.Pearl [Pearl]). In particular the existence 'axiom' is 'formalised' into the principle according to which what exists for a subject has a cause and an effect on this subject. This 'formalisation' is a projection: it is a possible interpretation of the Eleatic principle formulation given a priori the context of finite probabilistic dynamical systems, but there is no reason to identify this principle's formula with the reality that it points to, and a fortiori its formal counterpart with of this reality. After that the 'formalisation' of the system's other element follow a similar line, projecting their formulation into the contextual formalism of finite probabilistic dynamical systems using the same 'formalisation' of causality.

In this production the theory also still relies on a form of the constancy hypothesis (mentioned in the introduction) which identifies in a strong sense any 'atomic' experience (a bit of information about it) with 'atom of space and time in a certain state' (which in particular 'exists' according the

theory's interpretation of existence). The importance of this hypothesis lies in the possibility to apply the principles ruling the Experience to its projection on the formal context by isomorphism. This hypothesis is in fact a strong position: one can imagine a lot of other a priori plausible ways (even if mathematical) for the Experience to be causally related to the objective world.

Along its 'formalisation' process the theory produces a formalism which is 'somehow' related to reality but is not guaranteed to talk about it. Furthermore the use of the Eleatic principle as well as the reasoning structure used in the theory's construction should recall some metaphysical argumentations criticized by Kant in his *Critique of Pure reason* (p.570, l. 11), such as Leibnizian proof of God's existence *a contingentia mundi*, also called cosmological proof: "*If something exists, then an absolutely necessary being also has to exist. Now I myself, at least, exist; therefore, an absolutely necessary being exists.* " According to Kant, this proof, which "in order to ground itself securely, [...] gets a footing in experience, and thereby gives itself the reputation that it is distinct from the ontological [Cartesian] proof". But as soon as this is done, "reason says farewell to it [experience] entirely and turns its inquiry back to mere concepts", and in fact reduces itself to the ontological proof.

This similarity should be a hint of the theory's epistemic instability. Moreover as it is deductively detached from its origin, the ability of the formalism developed in this theory to talk about reality is subjected to ostensive proof, which is still lacking. On the other hand this does not imply that it is not possible at all to use mathematical language to talk about phenomenal experience as such, but a fertile way to do it may come from another formal context than the current paradigm's one [consider for instance the formalisation of computation by A.Turing which appeared in a completely different setting than former attempts by A.Church (lambda calculus) and K.Gödel (computable functions)]

3.1.3.2 Procustean cut on phenomenal spatiality One of the most recent outcomes of the theory was an attempt to deploy it in order to *explain* the observations of a 'phenomenology' (understood here etymologically as the discourse on Experience as such) of space, in other words the any experience's spatiality - chosen for its 'introspectable' character, in the sense that spatiality under direct introspection has more structure than the only sensation of space, contrarily to color or pain for instance.

Beyond the 'formalisation' of its fundamental system, the theory conjectures an identity between a phenomenal experience experienced by a physical system (which does have an experience if it verifies a particular mathematical condition which consists roughly in being an integration maximum amongst other abstract systems obtained by extraction from the whole physical world) and this system's *cause-effect structure* at the time of experience. This cause-effect structure can be thought of roughly as the set of subsystems that have a cause and an effect, both in an irreducible way - which means that the causal relation it has with the cause or effect can not be reduced to multiple independent causal relations.

In this setting, in order to account for the spatial character of any experience that \mathcal{J} can imagine having, the theory adopts the following argumentation:

1. It is 'known' that bidimensional grid-like neural networks "map" visual space and body space (terms used in [Haun Tononi]), in the sense that to every *position* - or neuron - in the grid-like network corresponds to a position in the considered phenomenal space - in other words there is a one-to-one correspondance between the network and the phenomenal space (in coherence with the constancy hypothesis). As a consequence it is probable that one could account for the 'existence' of space, or its 'presence' in an experience with a causal analysis of these grid-like neural networks, using the 'formalisation' of the fundamental system into causal terms.
2. If one computes the cause-effect structure (as described above) of a simplified finite grid neural network, one can observe that the set of groups of this grid's units that 'exist' (meaning that have a cause and an effect in the dynamics of the network) satisfies some structural properties,

amongst which the following ones: 'most' of these groups contains strictly another group which is in the set, and another one that strictly contains it (of course that can not be verified for the group of all the units of the grid).

3. One can observe that \mathcal{J} can distinguish in any experience some 'things' (which although not properly defined in the text [Haun Tononi] can be thought of as areas delimited by closed paths when considered in the visual phenomenal space for instance) in this experience which are invariant from experience to experience and whose set satisfies the same structural properties as the set of 'existing' groups of units in the grid-like network.
4. As a consequence the bidimensionnal grid-like neural networks account, brought their cause-effect structure, for some aspects of the spatiality of any experience. This supports (according to the theory) the conjecture of identity between the cause-effect structure of the physical system (my body) and the experience that \mathcal{J} have, although the identity is only partial at this point.

I identified multiple direct difficulties with this argumentation:

1. Partiality of the correspondance:

The properties chosen to make a correspondance between experience's spatiality and grid-like neural networks' cause-effect structure are manifestly ad hoc: in fact these properties are not specific of any of the correspondance's terms - for instance any finite graph's parts set statisfies the same properties. In more trivial words this segment of the theory's argumentation is similar to an identity statement between a tomato and the little red book, based on the fact that they are both red.

2. Partiality of the representation:

As a matter of fact, the 'characterisation' of phenomenal spatiality as a set of areas misses various aspects of this spatiality, such as for instance, its continuous presentation or its orientation (top-bottom and left-right), which manifestly belong to the Experience, in the sense that they are immediately accessible, for instance when one evaluates the relative positions of two areas of space. Moreover the theory's account of spatiality does not take into account the variations in the sensation of space - for instance when one closes the eyes. It is thus manifest that, in order to make the correspondance they draw appear to be effective, the theory operations a procustean cut on spatiality by supressing from its representation (in the text) some of its fundamental properties.

3. On the actual 'presence' of space areas:

Even more importantly if one thinks of the experience as everything that appears - without cognitive effort - in the Experience at a time, then the areas that are used in the above argumentation - in order to characterise phenomenal spatiality - are not strictly present in the experience without me 'creating' them by my consideration (by further cognitive process) or by extracting the presence an object 'supported' by this area of space (which can be an actual object, a closed path drawn on a sheet of paper or imagined in my mind by a similar process). A proper account of spatiality should also exhibit this ontological dependance of the areas of space on objects that are 'present' in an experience.

Furthermore phenomenal spatiality is better represented and understood in already existing mathematical notions of space - in fact the understanding of spatiality belongs properly to mathematics and its formal characterisation to mathematicians. Although this does not lead to an explanation of phenomenal space's form in physical terms, there exists a better account of this space with bidimensional manifolds with (visual space) or without (body space) edges, augmented with an orientation, some metrics, etc.

Moreover even if the correspondance between phenomenal spatiality and the cause effect-structure of a neural network was more complete, this would not support the identity statement

- which in fact is not necessary for an understanding of Experience and its origin in the physical world - but only a categorial mapping between concepts and areas of Experience - which is enough to enable one to act on the Experience. For an analogy with mathematics, in category theory, when there exists a functor between, for instance, an algebraic category and a topological one, the categories may have the same structure (identified by the functor) but they stay phenomenally different.

In terms of the statical and dynamical phases of the discourse, the theory's account of phenomenal spatiality relies on artificial staticisation of two dynamical designations: the correspondance between grid-like neural networks and phenomenal spatiality, and the characterisation of this spatiality with a set of areas. These artificial staticisations are transported to the discussion on this matter (at the intersubjective level) in which the argumentation in favor of the theory takes the form of a demand for an understanding effort towards the theory's statements in this particular context. This effort consists simply in the suspension of critical thinking on introspection and the integration of this pair of staticisations. This way the 'phenomenology' conducted by integrated information theory stands in complete opposition to the phenomenology professed by the authors, as one can see in F.Varela's following description:

"In PhR the skill to be mobilized is called bracketing for good reasons, since it seeks precisely the opposite effect of an uncritical introspection: it cuts short our quick and fast elaborations and beliefs, in particular locating and putting in abeyance what we think we 'should' find, or some 'expected' description". - **F.Varela** [V96].

In the present context, what one 'should' find is precisely what one should accept in order to 'understand' the theory's statements. As a matter of fact, the intersubjective expansion of this theory relies on the absence of intersubjective agreement on the meaning of what is said by the theory, which thus does not say anything objective, even in terms of the intersubjective interpretation of objectivity. Beyond coercive persuasion the only factor of its expansion is the persuasive force of a collection of improperly defined analogies (that they call explanations) which are improperly interconnected, this force coming from the number of these analogies and not from their actual epistemic force. Despite the fact that the use of causality is widely spread in philosophy of mind and its vocabulary already present in Kant's writings, one can however attribute to this theory the merits of perceiving the possibility of using causality in the formalisation of phenomenal experience as such.

On a broader scope, the limit between what belongs to an experience and my actions on this experience at this time is itself dynamical. In other words it is difficult to distinguish them, in particular when the intensity of the actions is extremely small, or when one objectifies his or her own thoughts for instance - seeing them as part of the Experience. If this designation was static, it would enable a quantification of anything's presence to my mind - as the quantity of cognitive effort needed to isolate this thing from its context - and the possibility to share this without any hypothesis on the nature of this quantification (which would derive from a theory which is reductionist, fundamentalist and unificationist a priori), and thus construct an objective representation of the structure of any experience. I believe that this is the kind of objectivity that is aimed at by the theory, but artficed through the staticisation of the limit's designation. Despite its manifest dynamical character in the discussions on this matter, in which the designations of the areas of space mentioned above oscilate between a their description as attention focus actions (contradicting their existence in the experience by definition but used for persuasion) or as simply 'there'. It is remarkable that the counter-critic from the theory on this matter relies on this very difficulty from the other to determine the limit between what is perceived and what is cognised and thus to propose another (viable) solution: the absence of response from the interlocutor is interpreted as an argument in favor of the theory (while it should be invalidated for its artificial staticisation).

Moreover in this interpretation, there are other aspects of phenomenal spatiality that are not accounted by the theory's approach, for instance the greater presence of square areas compared to

others, or the greater presence of the totality of the areas occupied by objects in the experience compared to subgroups of objects.

Let us also notice some more epistemic difficulties in the interpretation of existence as presence intensity: should one consider in the experience what appears independantly from only conscious cognitive actions or all of them ? In the latter case, how can J distinguish parts of an experience that depend on cognitive actions that J do not perceive ?

3.1.3.3 Conclusive remarks From what I understand of integrated information theory's formalism, its derivation from the concept of consciousness is not rigorous enough (in the sense that there is a loss of meaning in this process) to expect with certainty any significant explanation phenomenal experience physical origin to come out of it. Despite the fact that the 'explanation' proposed is not accurate enough (which one can see on the attempt of account for spatiality), one could attempt to retroact on the formalisation in order for it to be more adapted to the reality of Experience, by progressive rectifications. However this strategy, which relies on the tension that exists between the current formalism and the observation of the phenomenal world, is not guaranteed to provide a meaningful direction.

This might be different in an application of the formalism to some microbiological systems, whose structure are better understood (although not in a systematic manner). This would yield an orientation of the theory towards a systematic understanding of *organisedness* (including for instance functional division, or the character of autonomy) of these systems, in which the 'axioms' of integrated information theory would merely play the role of principles, in the dynamic definition of causal structure, and in a dialogue with micro-biological data. The non-foundational character of the 'axioms' in this context yields the possibility of purely mathematical and physical principles in this definition, as well as the possibility of other formalisms, which would at the same time be more accurate and more tractable (where the formalism of integrated information is clearly untractable, bounding its development).

As a matter of fact, despite the interest of this theory in producing concepts and meaning at the border between cognitive sciences and philosophy, which is important for the problem of consciousness' physical origin, this conceptualisation comes with the cost of tractability and intuitiveness at the same time. Despite these issues, the theory's visibility can be explained by its ability to produce concepts - granting it Deleuzian value. This in turn comes from the formalisation process itself, in which the separation between sciences and philosophy is just broken, without an understanding of the reason why they were separated - in fact, multidisciplinary is replaced here with non-disciplinary. In the present section I used the notion of dynamical and statical phases of the discourse in order to analyse the way these phases interact in this theory, yielding the description of a 'knot' of dynamical and statical designations. In the following my aim is to use the phases separation criterion (presented in the first section) in order to provide an alternative 'formalisation' for the transcendental intretation of integrated information theory's fundamental system. This alternative makes use in particular of the localisation principle (formulated in the present section), in order to create a coherent formal 'contact surface' (a window) between the disciplines of mathematics and phenomenology, allowing to initiation of a dialogue while preserving the intuitiveness and the tractability in the formalisation. In this expression I understand a surface as a structure that one can explore from both epistemic formations but at the same time applies constraint on the exploration, avoiding the action of the will in the constitution of a theory about it.

3.2 Towards of a systematised approach of objectness: intuitions

The strategy that I will follow is to try to understand better the observations of integrated information theory's fundamental system in the transcendental interpretation. For its elements have a transcendental nature, there are two ways to progress in their understanding using formalism: projecting them onto a fixed formal context and producing formal objects representing them by analogy (with loss of meaning); or restricting their domain (as operations) to 'formal' objects. In doing so, one 'encapsulates' these operations in a formal context which participates to what

they actually are. Following this theoretical orientation, I choose to focus on the distinction operation, which is the most 'graspable' (or in other words introspectable, following the observation in [Haun Tononi] that spatiality is more introspectable than emotions or colors) - the reason being that designations produced in this direction are more stable intersubjectively. Along with the attempt of explaining this phenomenon, one may find along the way a possible explanation for other ones in the fundamental system (in particular integration, or various forms of integration, underlying non-separability). Despite the fact that one would find this by analogy, it would be expressed in formal terms depending mathematically on the formal representation of the objects towards which the transcendental operation is restricted, themselves consisting in statical designations. The closest mathematical type (in the above typology) to this quest is the reflexive one: in the same way Turing machine 'explains' computation and category theory 'explains' the cognitive instantiation of mathematical objects, I am searching for formal objects that could explain phenomenal objectness (the distinction of objects in the phenomenal field), in experiences that can themselves be described formally. In this section I expose some intuitions which I extracted from past decades theoretical ground, explaining the form of the formal framework that I will introduce in the last section of this text.

3.2.1 From the third person perspective to the first person one

The first intuition that I want to communicate is that of the difficulty to keep the first person perspective (which is what to be understood) at the center of attention.

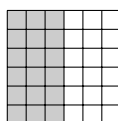
In particular I have noticed that most of the contemporary theories of consciousness rely on a form of concept tokening analogy (amongst which the *attentional schema theory* of M.Graziano or the *strong AI* (term of J.Searle) supported by D.Dennett): since it is possible to token the concept of consciousness in one's mind using a specific device, whether it is a ventriloque's puppet or an elaborated algorithm, meaning that this device appears as conscious in the intuition, then this device is conscious in its full sense, meaning that it should have phenomenal experience. However since one cannot reasonably conceive that this device is experiencing anything, it is concluded that phenomenal experience itself should be an illusion. As a matter of fact here the unquestioned analogy-based reasoning prevails on immediate intuition of phenomenal experience, which is to forget the nature of concept-tokening analogy phenomenological progression: its result should be considered as a good way to give an orientation to the intuition for the purpose of understanding of consciousness and not as a definitive answer.

In fact the flaws of this reasoning can be seen with J.Searle's *Chinese room argument*: this argument consists in imagining oneself (an english speaker) in a closed room with an opening wich is small enough so that one is not noticeable from the outside and being in charge to translate in english a text, input from outside, on a tape of paper written with chinese symbols, according to a symbol-manipulating program written in english, and then output the translated text. Despite the fact that from the outside one perceives the room as understanding Chinese (as the room is the only thing that one perceives and is interpreted as the translation's actor), this is in reality not true since *I* can imagine myself being the one in the room even if I don't understand Chinese. As a consequence consciousness from the third person perspective is the illusion while consciousness from the first person perspective does exist (and thus those two forms of 'consciousness' should not be identified!)

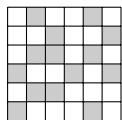
My point here is that the use of concept-tokening analogy has limits, in particular when it is applied to consciousness. This is the level at which the reasoning should be changed, in particular to preserve the object of study (phenomenal experience, consciousness from the first person perspective). This is what one should focus on before going further (respectively to the distortion of scientific time). As soon as one has a proper way to approach phenomenal experience systematically (which should obey the phases separation criterion) one can try to explain phenomenal experience in formal terms before going even further with a physical explanation (on the model of the decomposition of the universal gravitation theory's genesis, into Kepler's mathematical explanation of the partial observation of planets movement and Newton's physical explanation based on Kepler's work).

3.2.2 Real patterns

Surprisingly enough I found the possibility of combining a formal approach and the first person perspective in the following paper by D.Dennett: *Real patterns* (1991). In this article, D.Dennett illustrates his *mild realism* on beliefs and mental states (questioning their ontological status: do they exist "out-there" or are they just the product of our imagination?) with the idea that they are "as real as electrons or centers of gravity" which, although they can be considered as useful fictions for our understanding of the world, are real in the sense that they 'belong' to what is observed "out-there" (such as for instance the planets movement formal description) as behavioral patterns. In order to do that he uses visual *patterns* whose reality is identified with their distinguishability over the background. For some simple patterns this reality can thus be explained in principle by the possibility to describe them with some proposition. For instance the pattern



can be described as "*a square divided in two equal vertical rectangles, one gray and the other white*", while in order to describe a 'random' pattern, for instance the following one:



one has to collect every bit of information, which makes the description a lot longer.

The proposed explanation relies on a pre-established interpretation 'program' in the human mind, and a priori there is no reason that every person has the same 'program' and will perceive the same patterns as others. Moreover patterns' saillance may depend on expertise, as chessboard patterns can be detected more efficiently by expert chess players than random players. The fact that there are some intersubjectively shared patterns could be explained by the common use that we have for them (in predictions for instance). However as one can see in the Game of life (a cellular automaton defined by the mathematician J.H.Conway and taken as example by D.Dennett), there are some patterns that are irresistibly saillant, independantly of any intention towards them, which makes them real. In the last part of his text Dennett extends these intuitions to folk psychology, where mental contents and beliefs can be seen as patterns.

A remarkable aspect of Dennett's position in this article is the use of a formal object (J.H.Conway's Game of life) as a visual experience (or a temporal series of visual experiences) in order to capture the distinction of patterns as a phenomenon (that I would abstract saying that they are particular "objects"). I interpret this distinction as visual (and as a consequence perceptual) since it is not intentional but appears to me without effort. Dennett was interested in questioning the reality of these patterns and did not go further in the use of this model. For my purpose here it is actually important to notice that by this stance one captures not life itself in J.Conway's model, but also the distinction of living 'things' - from the first person perspective! - in a particular experience which can be described formally (by describing each bit of information according to a pre-fixed visual code).

Furthermore the fact that this model can be considered at the same time as a visual experience and a mathematical object, that can be studied as such, forming what I may call a **porous contact** between phenomenal experience and mathematics, yields some hope to capture formally the phenomenon of distinction. In my view patterns that one distinguishes are *real* in the strict sense, as my conscious experience is real before anything else.

Remark 2. *I would rather use the term of reality instead of existence used in integrated information theory, as it bears metaphysical pre-conceptions that are difficult to extract.*

My strategy is to build a framework for the study of distinction (or phenomenal objectness) over a class of formal objects which contains this singular model (and more generally cellular automata) in a more systematic way than Dennett. The reason of this extension beyond the class of cellular automata is that it is difficult to theoretise objectness over this class only for tractability reasons. On the contrary the larger class that I will present in the following contains models that are simple enough (in particular in terms of causality) to support tractable explaining formalism and complex enough to provide significant benchmarks for any attempt of uniform explanation of the distinction phenomenon. This uniformity is fundamental for various reasons: first, that a uniform explanation over a large class of models would explain objectness as such and not objects specific to a model or a small group of models; second, that the proposed explanation would be more likely to be independent of the perceiving subject (or the evolution of a subject with time) and thus to have an objective counterpart.

This approach will necessarily involve a dialogue between phenomenology (in the discourse about phenomenal experience) and its principles as well as mathematics and their principles, and this is where one should care about the conditions of possibility of such a dialogue. Moreover I choose to focus on visual experience for the richness of its introspection and assume that perception of the kind of experiences considered are not impacted by other senses (which seems reasonable hypothesis given their abstract nature). Naturally it can be objected that patterns in Conway's cellular automaton are distinguished because they token the concept of living being: thus are not strictly visual patterns in the sense that their distinction may involve other factors than pure perception. However we will also consider more abstract models for which this effect disappears.

3.2.3 Worlds 'in' the World: an *isolation* operation

An important aspect of J.Conway's cellular automaton is that, when considered as the description of visual experience series, these series of experiences are considered in a group of other ones which is cognitively isolated from the total enriched set of experiences derived from the set of possibles experiences.

For this model the isolation factor for grouping is the abstraction act inherent to mathematical definition. However this is not the only one possible: for instance one can encounter natural isolation when looking into a microscope's eyepiece: all the visual experiences that one has there are considered together as isolated from ordinary experience, and grouped into the 'microscopic world' (a world 'in' the World). One can find another example in the 'animal world' or the 'natural world' for which the isolation intensity is lower.

In such a 'world in the World', the perception of an experience (or a temporel series of experiences), in particular the distinction operation, is dependent upon the whole world it belongs to, with gradual intensity according to - possibly multiple - isolations. In the context of Euclidian geometry for instance, a straight line has a meaning in this specific context while it does not have any particular meaning in the whole set of possible experiences. There is a certain objectivity for this concept for the reason that the actual isolation operation is not consciously articulated.

As much as objectness the worldness of a world does participate to the concept of consciousness. In fact the actual set of past experiences is specific to the person's singularity and her consciousness, in particular for the reason that this *world* has an impact on what \mathcal{I} am, determining my thoughts and what I perceive in the World, what \mathcal{I} evaluate as possible, and thus my decisions and actions. This is, I believe, one reason of the inclusion of the concept of world in Heidegger's concept of *dasein* (being-to-the-world).

For my purpose the interest of the isolation transcendental operation is that it is possible to avoid projection of its concept and use instead domain restriction in order to formalise it: indeed, for any model such as Conway's cellular automaton, one can conceive an actual phenomenal grouping when considering a microscope in whose eyepiece one can see the series of visual experiences described by the model and only those. These models can be then used to describe a particular (ideal) 'world in the World'. In the following, I shall call them **micro-worlds** (the term 'micro' referring to the fact that these worlds contain only a very small quantity of in-formation (as the action of the world on general distinction)).

Furthermore the objectivity of the concept of world in the World allows the application of the localisation principle (in particular through the mathematical descriptibility of micro-worlds), which is a fundamental factor in the scientification of the investigation. The mathematical descriptibility also allows the possibility of an exhaustive collection of distinctions in micro-worlds as well as the possibility of an exhaustive search for an explanation (that I shall call an *ontology*) for the whole set of distinctions, and the robustness of its results (in that exhaustiveness also prevents from neglecting hidden factors in actual distinction and thus the in-significance of the explanation).

In order to defined a formal framework for the study of micro-worlds, I will also assume the following hypotheses:

1. that the visual experiences considered are infinite ones: what one actually sees is considered as a part of an infinite 'experience' that is experienced by parts, by shifting it ad indefinitum. (in mathematical terms, the infinite experience is the inverse limit of a series of experiences obtained by extension). This infinite experience is conceivable in the enriched set of experiences of a micro-world (which is what I will actually consider). This hypothesis serves the formal analysis by neglecting border effects, and is reasonable for the reason that what we perceive in the center is perceptively not affected by what lies at the limit of the visual scope. In fact it is present in Euclidian geometry, where it does not affect the truth of any proposition, but grounds the possibility of the representation of arbitrarily large figures.
2. that any of the micro-worlds are invariants by shifting, meaning that if an experience is possible in this world, then any of the experiences obtained by shifting (in any direction) this one is possible in this world.

Let us notice that the transition to a mathematical framework relies non-reductibly on the transcendental one; this transition mainly consists in the objective progression on the basis of what is the object of a (potential) intersubjective agreement.

3.2.4 Objectal complexity

Along with the search of a formal explanation for the phenomenon of distinction, one can search for a quantification of the 'richness' or 'complexity' of any experience with respect to distinction (which is different from a 'quantification of consciousness' but closer than any other existing quantities): a form of *objectal complexity*. The possibility of a mathematical definition of such a quantification can at the same time serve as a test for a formal 'ontology' and be derived from it. Such a notion of complexity would be specific to perception and different from algorithmic complexity developped in the last XXth century: as a matter of fact, although a random pattern is 'complex' in this sense, meaning that it is difficult to describe with an algorithm (bit by bit description), it does not mean it is perceptually rich: no 'thing' is perceived in it. On the contrary, the experience that *J* have contains many things and types of things, articulated in many ways, and this constitutes its richness.

3.2.5 Conclusive remarks

I would like to notice that the framework in construction in this text is specific to a particular approach, derived by un-knoting integrated information theory's knot of static and dynamic designations, and can not a priori be generalised outside of this context.

This framework allows in particular to search for ontologies derived from the ones proposed by integrated information theory [AOT] by re-ordering the mathematical operations involved in them (for instance cutting, summing, etc) in order to build another formalism which would be tractable - which would possibly allow an ostensive proof of its expressiveness (in the same way as it is possible to re-order the terms of a converging series in order to compute its sum).

The theoretical orientation that I propose in the following section is not strictly speaking mathematical for the reason that, by the nature of its object (the understanding of the distinction transcendental operation as such) it has to include intradiscursively (and not paradiscursively as

in mathematics) a dynamical phase. It is close enough to reflexive mathematics in the sense that the elements of the discourse (the ontologies), which derive from the practice of phenomenology, are specific to a transcendental position and not to the object only towards which it is directed. However in reflexive mathematics, this object is mathematical and extracted from its experiential context, while my approach is meant to place this context back at the center, while ensuring that the description is made in a reliable way.

The phases separation strategy is implemented in a way that preserves the dynamical (for the collection of distinctions description, specifically phenomenological) and statical (for the formal ontological statements relating these description to the description of the model) nature of the discourse types involved in the dialogue; and what permits this dialogue is the existence of a 'porous contact' between the formal world and the world of perception in the micro-worlds, which can be considered at the same time as an experience (in its plain sense) and a formal object.

4 A formalism of micro-worlds

In this section I introduce a formal framework, for the study of objectness, inspired by the intuitions described above. It is composed of a series of definitions (micro-world, ontologies, etc) and a series of lemmas, whose purpose is mainly to prove the formalism's tractability, in other words the possibility to prove non-trivial and intuition-based mathematical results. Let us start with the formal definition of micro-worlds.

4.1 Micro-worlds

Let \mathcal{A} be a finite set of square tiles, that I call **alphabet**, such as the following:

$$\mathcal{A}_{01} = \left\{ \begin{array}{|c|} \hline \bullet \\ \hline \end{array}, \begin{array}{|c|} \hline \bullet \\ \hline \end{array}, \begin{array}{|c|} \hline \bullet \\ \hline \end{array}, \begin{array}{|c|} \hline \bullet \\ \hline \end{array} \right\}$$

Definition 1. A *micro-world* on alphabet \mathcal{A} is a subset W of $\mathcal{A}^{\mathbb{Z}^2}$ (the set of colorations of an infinite square grid with elements of the alphabet \mathcal{A}) that is shift-invariant, meaning that for all $w \in W$ and $v \in \mathbb{Z}^2$, $(x_u)_{u+v} \in W$. The elements of a micro-world shall be called **data configurations**.

Each of the data configurations of a micro-world is considered as the (mathematical) description of an experience in this micro-world, from which it should be differentiated.

Let us call **pattern** on alphabet \mathcal{A} an element p of some $\mathcal{A}^{\mathbb{U}}$, where \mathbb{U} is some part of the grid \mathbb{Z}^2 : this set is called the **support** of the pattern p . It is straightforward that a subset of $\mathcal{A}^{\mathbb{Z}^2}$ is a micro-world if and only if all of its data configurations w obey a set of rules, under the form \mathcal{R}_p : " p does not appear in w ", where p is a pattern on alphabet \mathcal{A} . The set of patterns associated to these rules is denoted \mathcal{F} and its elements are called **forbidden patterns** (in the micro-world W). For example, the rules defined by the set of patterns \mathcal{F}_{01} whose elements are all the possible rotations of the following patterns:



on alphabet \mathcal{A}_{01} yields a micro-world W_{01} which consists in the set of pictures as the one on Figure 1.

Remark 3. The definition of micro-world considered here is limited to static bidimensional visual experiences, but the definition can be extended to dynamic experiences, such as series updates of a bidimensional cellular automaton, by considering shift-invariants subsets of $\mathcal{A}^{\mathbb{Z}^3}$ (the update rules of the cellular automaton being translated straightforwardly into rules of the micro-world). Moreover one can also extend this formally to subsets of \mathcal{A}^V for a finite graph whose vertex set is denoted V . However the rules would not necessarily be applied uniformly, which makes the study of these systems a priori more difficult. Moreover it is possible that the graph's structure might influence what is distinguished in the experience. By choosing to restrict to the graph \mathbb{Z}^2 I cancel this potential phenomenon.

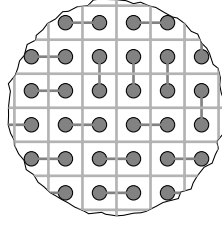


Figure 1: A microscope eyepiece representation of a data configuration in the micro-world W_{01} .

Remark 4. Let us notice that although we consider here only micro-worlds generated by a finite set of rules, which yields only simple objects that seem far from common experience, the following analysis can be extended to micro-worlds generated by an arbitrary set of rules, yielding much more complex behaviors. Another way to extend the investigation would be to define micro-worlds in a continuous setting, taking for example the set of possible figures of Euclidian geometry.

4.2 Ontologies

Notation 1. Let us denote \mathcal{L}_A the set of all the possible patterns on the alphabet A and \mathcal{L} the set of all the possible patterns on any alphabet. The **language** of a micro-world W , denoted $\mathcal{L}(W)$, is the set of patterns that appear in at least one data configuration of this micro-world.

For a set S , we will denote $\mathcal{P}(S)$ the set of its parts.

Definition 2. Given a micro-world W , let us call an **agreement** over X any function $W \rightarrow \mathcal{P}(\mathcal{L}(W))$.

The Definition 2 captures the intuition of an intersubjective agreement on the set of patterns that one distinguishes in data configurations of a micro-world. This definition enables the description of temporary intersubjective agreements, potentially conflicting. As a consequence it allows the integration in the discourse of the intersubjective agreements progression along the construction of a formal explanation of the distinction phenomenon, and materialises the intradiscursive shift of the phases limit. Given an agreement we will search for an 'explanation' for this agreement which depends mathematically on the micro-world's formal description.

Notation 2. The set of all the micro-worlds will be denoted by \mathcal{U} .

Definition 3. An **ontology** is a function π which to a micro-world $W \in \mathcal{U}$ associates a function $\pi(W) : W \rightarrow \mathcal{P}(\mathcal{L}(W))$ such that there exists an algorithm which taking as input a micro-world W outputs a second-order logical formula which has a pattern free variable p and a data configuration free variable w such that for all $w \in W$, $\pi(W)(w)$ is the set of patterns which satisfy this formula.

Definition 4. Given an ontology π , a set \mathcal{D} of micro-worlds and a family of agreements $A_{\mathcal{D}}(A_W : W \rightarrow \mathcal{P}(\mathcal{L}(W)))_{W \in \mathcal{D}}$, the ontology π is said to be **compatible** with $A_{\mathcal{D}}$ when for all $W \in \mathcal{D}$ and for all $w \in W$, $\pi(W)(w) = A_W(w)$.

In line with the intuition, I call **epistemic force** of an ontology the cardinality - the number of its elements - of an ontology's domain of compatibility with 'natural' agreements: the more an ontology has significance for our purpose the more it has epistemic force in this sense.

Remark 5. In Definition 3 it is important to note that an ontology explains an agreement in the sense that it derives the agreement from the description of the micro-world, thus extending the strict agreement with a mathematical structure, in the same spirit as Kepler completed the planets movement description with a mathematical object (an elliptic path). A fundamental aspect of this explanation is its exhaustiveness: the ontology explains the presence of some patterns in the agreement as well as the absence of the others.

Given a 'natural' family of agreements $(A_W)_{W \in \mathcal{U}}$ we would ideally like to find an ontology on this domain for this family of agreements (it would have maximal expectable epistemic force). However there is a priori no reason to think that such an ontology actually exists. As a consequence we will search instead for ontologies with maximal compatibility domain, given a natural family of agreements (which in practice will be constituted along the ontology's construction), on the basis of small compatibility domain ontologies (we will recycle the idea of causality in order to find them). Let us note that the difference between conflicting 'natural' agreements may often come from the inclusion of certain patterns in the agreement more than their exclusion: it is easier to agree on that certain patterns do not 'make sense' in the micro-world and are not distinguished than decide that they do indeed 'make sense'. Moreover one important criterion for the selection of an agreement amongst conflicting ones might be the extension of the ontology's domain itself, in other words the growth of its epistemic force. Another possible criterion is the restriction to agreements of patterns of which any data configuration is made (intuitively distinction should satisfy an optimisation property in terms of quantity of information). These criteria should be considered so far as only principles of investigation and should be refined along with the framework's development.

Remark 6. *Although the formalism presented here does not include probability distributions, 'probabilistic ontologies' are not excluded, and it is possible in principle to define the integrated information theory's 'ontology' in this framework, which would be based on natural probability measures such as the micro-world's (considered as a symbolic dynamical system) invariant measures. The construction of the present framework allows the comparison with other (in particular more tractable, and not based on a particular probabilistic formal context) ontologies.*

4.3 Phenomenology of some micro-worlds

With these definitions in mind, let us illustrate them on some simple examples of micro-worlds and illustrate how phenomenology and mathematics get intertwined along with the definition of ontologies compatible with natural agreements.

4.3.1 Definition of some naive ontology

Let us consider some simple example of micro-world W_{00} defined by the following alphabet and forbidden patterns:

$$\mathcal{A}_{00} = \left\{ \begin{array}{|c|} \hline \text{blue} \\ \hline \end{array}, \begin{array}{|c|} \hline \text{white} \\ \hline \end{array} \right\}, \quad \mathcal{F}_{00} = \left\{ \begin{array}{|c|c|} \hline \text{white} & \text{blue} \\ \hline \end{array}, \begin{array}{|c|c|} \hline \text{blue} & \text{white} \\ \hline \end{array} \right\}$$

This yields a set of pictures as the ones on Figure 2.

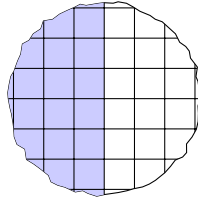
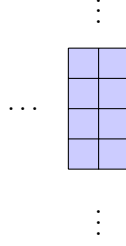


Figure 2: A microscope eyepiece representation of a data configuration in the micro-world W_{00} .

In any data configuration in this micro-world, I assume anyone would distinguish (and phenomenologically observe this distinction) simple possible 'significant' objects: the 'blue left half planes', 'white right half planes', 'blue plane' and 'white plane', such as the following one:





Notation 3. I will thus denote and consider in the following the agreement A_{00} which to a plane configuration will associate the set of patterns reduced to this configuration and to a half plane one will associate the set of its two half planes.

In the data configuration on Figure 2 for instance one can distinguish a blue left half plane and a white right half plane. This type of patterns do 'make sense' in this micro-world, while the following 'arbitrary' one does not:

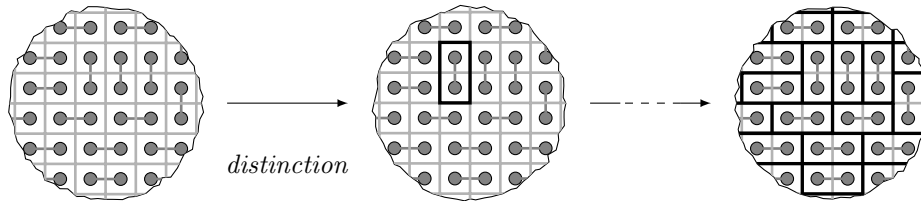


Let us say that a subset of \mathbb{Z}^2 is connected when one can draw a path in \mathbb{Z}^2 starting from any position in this subset, ending to any other and staying in the subset along the way. One can find straightforwardly a (naive) ontology π_0 on the domain \mathcal{D} that is reduced to the world W_{00} ($\mathcal{D} = \{W_{00}\}$), whose image by the ontology is a formula - which intuitively exists but I won't make it explicit: in the following I will assimilate the formula with the set of patterns it describes - describing for any $w \in W$ a set of patterns whose elements are the maximal patterns p in \mathcal{A}_{00}^U such that for all position in U has colored with the same symbol, U is a connected subset of \mathbb{Z}^2 and the restriction of w to U is p .

4.3.2 An obstacle to the naive ontology: dimers

Let us consider the example of the micro-world W_{01} . After considering some data configurations in this micro-world, I distinguish the following patterns (and assume that this leads to a natural intersubjective agreement):  and  that are called *dimers* (giving their name to the corresponding statistical physics model), and only them.

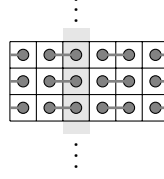
The principle of exhaustion exposed in the first part (interpretation of integrated information theory's fundamental system in transcendental terms) can be illustrated as follows:



This corresponds to the intuition that the experience a data configuration describes in W_{01} is *made of* dimers. The dimers are in this sense the 'constitutive elements' of this micro-world (as atoms in the physical world), and are distinguished for this reason. In other words they are natural elements for a uniform description of any data configuration in W_{01} . Following an idea

of D.Dennett expressed in his article *Real patterns*, this last interpretation could be used for the definition of an ontology, whose compatibility domain would include W_{01} .

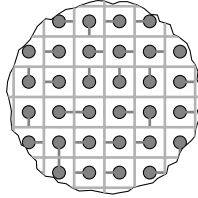
The dimers patterns constitute an agreement that I will denote A_{01} such that for all $w \in W_{01}$, $A_{01}(w)$ is the set of patterns that appear in w . One can see that the ontology π_0 fails to extend onto the domain $\{W_{00}, W_{01}\}$ for the agreements A_{00} and A_{01} , since in some data configurations w of W_{01} , this ontology distinguishes patterns that are not in the agreement $A_{01}(w)$, as the highlighted infinite column in the following picture:



This pattern is intuitively excluded from the agreement (intuitively it 'only happened' in the data configuration, in other words there is no particular reason for its presence that derive from the micro-world itself). Moreover, the ontology π_0 does not distinguish the dimers themselves. At this point we should like to search for another ontology which is compatible on domain $\{W_{00}, W_{01}\}$ with agreements A_{00} and A_{01} . Such an ontology will use here causality and will be presented later. In the rest of this section we will consider other simple examples of micro-worlds and briefly give some other interesting phenomenological observations.

4.3.3 Effet of changing the rules

Let us consider the world W'_{01} whose alphabet is \mathcal{A}_{01} and forbidden patterns are $\mathcal{F}'_{01} = \emptyset$. This micro-world is thus obtained from W_{01} by removing all its rules. A typical data configuration is then the following:



The difference with the micro-world W_{01} is that dimers are not distinguished anymore (some dimers are present in data configurations but they 'just happen' to be there). One distinguishes instead dimers' elements as constituents of this new micro-world. In other words the 'principle' bounding these elements in W_{01} disappeared. This illustrates the influence of the micro-world (and in particular the rules) on the distinction phenomenon.

4.3.4 Movement in the micro-world

Another way to perceive that dimers 'make sense' in the world W_{01} is the intuition of *possible elementary movement* in this micro-world: one can move from one experience to another by an elementary perturbation which consists in changing one symbol of the data configuration into another, potentially breaking the rules of the micro-world, and then adapting the data configuration to the perturbation - meaning changing a set of symbols which is minimal knowing that the resulting data configuration belongs to the micro-world. One can see an example on Figure 3.

It is natural to see an elementary movement in W_{01} as a displacement of the first data configuration's dimers which are preserved along the movement. One could use this idea for the definition of another ontology which would be based on the structure of possible elementary movements - in

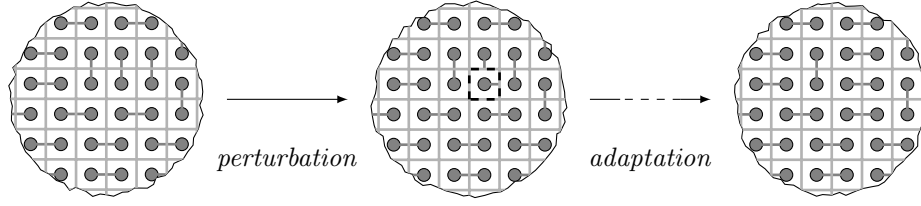


Figure 3: Example of possible movement in the microworld W_{01} from the leftmost picture to the rightmost one.

other words the possible reactions to an elementary perturbation, and what is preserved through these perturbations.

4.3.5 Spaced dimers

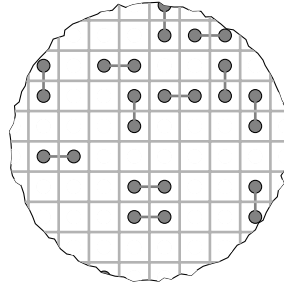
Let us consider the micro-world on alphabet

$$\mathcal{A}_{10} = \left\{ \begin{array}{|c|} \hline \bullet \\ \hline \end{array}, \begin{array}{|c|} \hline \bullet \\ \hline \end{array}, \begin{array}{|c|} \hline \bullet \\ \hline \end{array}, \begin{array}{|c|} \hline \bullet \\ \hline \end{array}, \begin{array}{|c|} \hline \\ \hline \end{array} \right\}$$

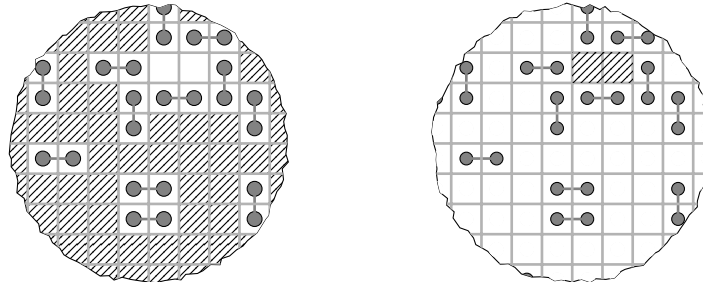
which is defined by the following set of forbidden patterns \mathcal{F}_{10} whose elements are the patterns obtained by any rotation of the following ones:



An example of data configurations in this micro-world is the following one:



While the possible elementary movements in the micro-world W_{01} corresponds to the displacement of dimers whose movement force the movement of the other dimers, movement in W_{10} is made a priori more *flexible* by the presence of 'space' between the dimers. Data configurations can be described in W_{10} as 'made' out of dimers with arbitrary space between them. In this micro-world I distinguish the dimers and the maximal 'spaces' enclosed by them such as the dashed patterns on the following figure:



Let us note that the denomination of 'space' is intuitive, and I do not mean to explain here how the types of distinctions of a 'space' and 'dimers' differ. However, at this *philosophical time*, one can notice that a space pattern tokens the more general concept of space: as a consequence this makes this type of micro-worlds interesting for a formal approach of phenomenal space. Moreover the 'space' patterns considered here echo the intuition of ontological dependence of spaces on 'objects'. In this phenomenology of space one should be tempted to identify a 'space' pattern with the *support* of this pattern (where the support of a pattern in $\mathcal{A}^{\mathbb{U}}$ is \mathbb{U}). One can then talk about the 'possible spaces' of this model (here the set of connected subsets of \mathbb{Z}^2) and as a consequence about 'spaces occupied by objects' (here the dimers). Although often identified, they should be differentiated when collecting more systematically 'phenomenological data' about phenomenal spatiality. In particular the 'space' patterns are more closely related to phenomenal spatiality than their abstraction, in the sense that they in themselves token more intuition related to this spatiality (for example the transcendental algorithm by which one recognises 'a space': focusing one undividable area in it and expanding in every direction at the same time until reaching the limits).

4.3.6 Drifting curves

In order to illustrate the diversity of the micro-worlds class in terms of the distinction phenomenon, as well as an example of conflicting possible agreements, let us introduce another micro-world, denoted W_{11} . This micro-world is on the alphabet

$$\mathcal{A}_{11} = \left\{ \begin{array}{c} \square \\ \text{II} \\ \text{II} \\ \text{II} \\ \text{II} \end{array}, \begin{array}{c} \square \\ \text{II} \\ \text{II} \\ \text{II} \\ \text{II} \end{array}, \begin{array}{c} \square \\ \text{II} \\ \text{II} \\ \text{II} \\ \text{II} \end{array}, \begin{array}{c} \square \\ \text{II} \\ \text{II} \\ \text{II} \\ \text{II} \end{array}, \begin{array}{c} \square \\ \text{II} \\ \text{II} \\ \text{II} \\ \text{II} \end{array} \right\}$$

and is defined by the set of forbidden pattern \mathcal{F}_{11} defined as the union of the two following sets:

$$\left\{ \begin{array}{c} \square \\ \text{II} \\ \text{II} \\ \text{II} \\ \text{II} \end{array}, \begin{array}{c} \square \\ \text{II} \\ \text{II} \\ \text{II} \\ \text{II} \end{array}, \begin{array}{c} \square \\ \text{II} \\ \text{II} \\ \text{II} \\ \text{II} \end{array}, \begin{array}{c} \square \\ \text{II} \\ \text{II} \\ \text{II} \\ \text{II} \end{array}, \begin{array}{c} \square \\ \text{II} \\ \text{II} \\ \text{II} \\ \text{II} \end{array}, \begin{array}{c} \square \\ \text{II} \\ \text{II} \\ \text{II} \\ \text{II} \end{array}, \begin{array}{c} \square \\ \text{II} \\ \text{II} \\ \text{II} \\ \text{II} \end{array}, \begin{array}{c} \square \\ \text{II} \\ \text{II} \\ \text{II} \\ \text{II} \end{array}, \begin{array}{c} \square \\ \text{II} \\ \text{II} \\ \text{II} \\ \text{II} \end{array}, \begin{array}{c} \square \\ \text{II} \\ \text{II} \\ \text{II} \\ \text{II} \end{array} \right\}$$

$$\left\{ \begin{array}{c} \square \\ \text{II} \\ \text{II} \\ \text{II} \\ \text{II} \end{array}, \begin{array}{c} \square \\ \text{II} \\ \text{II} \\ \text{II} \\ \text{II} \end{array}, \begin{array}{c} \square \\ \text{II} \\ \text{II} \\ \text{II} \\ \text{II} \end{array}, \begin{array}{c} \square \\ \text{II} \\ \text{II} \\ \text{II} \\ \text{II} \end{array}, \begin{array}{c} \square \\ \text{II} \\ \text{II} \\ \text{II} \\ \text{II} \end{array}, \begin{array}{c} \square \\ \text{II} \\ \text{II} \\ \text{II} \\ \text{II} \end{array}, \begin{array}{c} \square \\ \text{II} \\ \text{II} \\ \text{II} \\ \text{II} \end{array}, \begin{array}{c} \square \\ \text{II} \\ \text{II} \\ \text{II} \\ \text{II} \end{array}, \begin{array}{c} \square \\ \text{II} \\ \text{II} \\ \text{II} \\ \text{II} \end{array}, \begin{array}{c} \square \\ \text{II} \\ \text{II} \\ \text{II} \\ \text{II} \end{array} \right\}$$

This yields the following type of pictures:

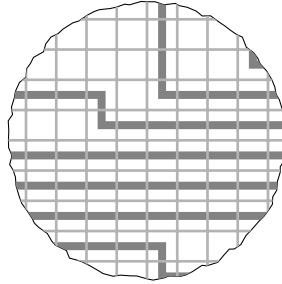
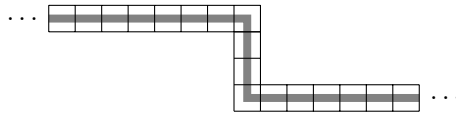


Figure 4: An experience in the world W_{11} .

The phenomenology in this micro-world is similar to the one of the spaced dimers, and one can distinguish the following type of pattern that I shall call a **curve**:



as well as the spaces between them. These patterns form one possible agreement. The subtlety here is that one could also distinguish the various parts of the curves: the segments and corners

which constitute them. This second agreement would be in conflict with the first one, and the ambiguity comes from the ontological status of the parts, since they completely derive from the knowledge of the curves themselves. The selection of one of these agreements should actually come out of the whole ontological process.

In the following, I will denote A_{11} the agreement which consists in the distinction in any data configuration of all the curves if it contains at least one and else the whole data configuration (thought as an infinite pattern).

4.4 Some causal ontologies

In this section I define some ontologies which go a little bit further in the ontological process (in the sense that their compatibility domains are larger or may be made larger in principle) than the naive ontology presented in the last section. As causality notions are widely present in contemporary philosophy of mind, in particular in theories relating mental content (distinguished areas of Experience) to causality (such as the integrated information theory reviewed above, but also W.Salmon's 'counterfactually robust causal relata'), I take causality in general as a theoretical ground for the definition of ontologies (in a similar way at the time of J.Kepler, one should have suspected to be able to describe astrological phenomena with geometrical figures); the difference with the aforementioned theories lies in the systematisation of the search for causal ontologies. After providing some definitions related to causality, I introduce two causal ontologies and prove formally that they are compatible with some natural agreements for the micro-worlds.

4.4.1 Definitions

4.4.1.1 Causal graphs Let W be a micro-world on alphabet \mathcal{A} which is generated by a set of forbidden patterns \mathcal{F} .

Notation 4. Let us denote \mathcal{R} the relation between elements of $\mathcal{L}_{\mathcal{A}}$ such that for any two patterns $p \in \mathcal{A}^{\mathbb{U}}$ and $q \in \mathcal{A}^{\mathbb{V}}$, $p\mathcal{R}q$ if and only if p and q do not overlap (the intersection of \mathbb{U} and \mathbb{V} is the empty set) and $\forall w \in W, (w|_{\mathbb{U}} = p \Rightarrow w|_{\mathbb{V}} = q)$. In this case we say that p **causes** q , or equivalently that p is a **cause** of q .

For any pattern $p \in \mathcal{A}^{\mathbb{U}}$ a sub-pattern p' of p is another pattern $\mathcal{A}^{\mathbb{U}'}$ with $\mathbb{U}' \subset \mathbb{U}$ such that the restriction of p to \mathbb{U}' is p' .

Notation 5. In the following, for any two patterns $p \in \mathcal{A}^{\mathbb{U}}$ and $q \in \mathcal{A}^{\mathbb{V}}$ whose restrictions on $\mathbb{U} \cap \mathbb{V}$ coincide, I will denote $p \cdot q$ the pattern in $\mathcal{A}^{\mathbb{U} \cup \mathbb{V}}$ whose restrictions on \mathbb{U}, \mathbb{V} are respectively p and q . We generalise this straightforwardly to any number of patterns.

Definition 5. For any pattern p , we will call **decomposition** of p any tuple of non-overlapping sub-patterns (p_1, \dots, p_n) such that $p = p_1 \cdot \dots \cdot p_n$.

Definition 6. We say that the relation $p\mathcal{R}q$ is **irreducible** when there is no non trivial decomposition (p_1, p_2) of p and (q_1, q_2) sub-patterns of q such that $p_1\mathcal{R}q_1$ and $p_2\mathcal{R}q_2$.

Definition 7. The **causal graph** of W , denoted $\mathcal{G}(W)$, is the directed graph:

1. whose vertices are the patterns p in $\mathcal{L}_{\mathcal{A}}$ which are related (through \mathcal{R}) irreducibly with some pattern q
2. and there is an arrow pointing from a vertex p to another one q if and only if $p\mathcal{R}q$ and this relation is irreducible.

As well the causal graph of a data configuration w in W , denoted $\mathcal{G}(w)$, is the subgraph of $\mathcal{G}(W)$ whose vertices are the patterns that appear in w .

Definition 8. An irreducible relation $p\mathcal{R}q$ is said to be **simple** when for any pattern p' ,

$$(p\mathcal{R}p' \text{ and } p'\mathcal{R}q) \Rightarrow p'\mathcal{R}p.$$

4.4.1.2 Generators of the causal graph

Notation 6. For a pattern $p \in \mathcal{A}^{\mathbb{U}}$, let us denote $|p|$ the cardinality of \mathbb{U} .

Definition 9. Given W a micro-world, a **generator** of its causal graph is an irreducible relation $p\mathcal{R}a$, such that $|a| = 1$.

Lemma 1. A relation $p\mathcal{R}a$ such that $|a| = 1$ is a generator of the causal graph $\mathcal{G}(W)$ if and only if there is no sub-pattern q of p and different from p such that q is a cause of a .

Proof. Let us assume the relation $p\mathcal{R}a$.

1. \Rightarrow **by contraposition.** If there exists q sub-pattern of p different from p such that $q\mathcal{R}a$, then there exists another non trivial pattern r which does not overlap q such that $p = q \cdot r$. As a consequence, we have $q\mathcal{R}a$ and $r\mathcal{R}\emptyset$, which means that $p\mathcal{R}a$ is not irreducible.
2. \Leftarrow **by contraposition.** If the relation $p\mathcal{R}a$ is not a generator, it means that one can write $p = q \cdot r$ such that q causes a and r is not empty. In particular q is a sub-pattern of p which is different from it such that q causes a .

□

Lemma 2. Any irreducible relation $p\mathcal{R}q$ can be **decomposed into generators** in the following sense: there exists a finite sequence of patterns $\bar{p} = (p_1, \dots, p_n)$ such that $p = p_1 \cdot \dots \cdot p_n$, and another sequence of distinct patterns $\bar{q} = (q_1, \dots, q_n)$ such that for all i , $|q_i| = 1$, and $q = q_1 \cdot \dots \cdot q_n$, and the relations $p_i\mathcal{R}q_i$ hold and are irreducible. The pair (\bar{p}, \bar{q}) is called a **decomposition** of the relation $p\mathcal{R}q$.

Remark 7. As a direct consequence a simple relation can be decomposed into simple generators.

Proof. It is trivial to find a sequence of distinct patterns (q_1, \dots, q_n) such that for all i , $|q_i| = 1$ and $q = q_1 \cdot \dots \cdot q_n$. Since $p\mathcal{R}q$, for all i the relation $p\mathcal{R}q_i$ holds. As a consequence there exists a minimal sub-pattern p_i such that $p_i\mathcal{R}q_i$: since it is minimal, the relation $p_i\mathcal{R}q_i$ is a generator. Since all the patterns p_i are sub-patterns of p , $p_1 \cdot \dots \cdot p_n$ is also a sub-pattern of p . Since this pattern causes q and that $p\mathcal{R}q$ is irreducible, then $p = p_1 \cdot \dots \cdot p_n$. □

Lemma 2 can be generalized straightforwardly in the following one:

Lemma 3. For any relation $p\mathcal{R}q$ there exists a finite sequence of patterns $\bar{p} = (p_1, \dots, p_n)$ such that $p = p_1 \cdot \dots \cdot p_n$, and another sequence of distinct patterns $\bar{q} = (q_1, \dots, q_n)$ such that for all i , $|q_i| \in \{0, 1\}$, and $q = q_1 \cdot \dots \cdot q_n$, and the relations $p_i\mathcal{R}q_i$ hold and are irreducible. The pair (\bar{p}, \bar{q}) is also called a **decomposition** of the relation $p\mathcal{R}q$.

Definition 10. Let us consider an irreducible relation $p\mathcal{R}q$, and (\bar{p}, \bar{q}) a decomposition of this relation. The **connection graph** of this decomposition is the undirected graph whose vertices are the elements of \bar{p} and two of these patterns are connected by an edge when they overlap non-trivially.

Proposition 1. A relation $p\mathcal{R}q$ is irreducible if and only if the connection graph of any of its decompositions is connected.

Proof. Let us assume that the relation $p\mathcal{R}q$ holds. Let us prove the direction (\Leftarrow) by contraposition (the other direction has a similar proof). Let us assume that $p\mathcal{R}q$ is not irreducible. As a consequence one can write $p = p_1 \cdot p_2$ such that p_1 and p_2 do not overlap and are not trivial and $q = q_1 \cdot q_2$ such that p_1 causes q_1 and p_2 causes q_2 . By decomposing both relations $p_1\mathcal{R}q_1$ and $p_2\mathcal{R}q_2$ into generators one finds a decomposition of $p\mathcal{R}q$ whose connection graph is not connected. □

4.4.1.3 Interpretation of causal relations as in-information The notion of causal graph encapsulates a notion of information which is fundamentally different from Shannon's one (in a similar way as the one posited by integrated information theory) and fits the intuitive notion of communication (or information transport) in micro-world, notably used in symbolic dynamics. In fact one can see a causal relation $p\mathcal{R}q$ as an in-information in the sense that the presence of p gives form to the data configuration over q 's support. As a consequence the causal graph contains all the in-formations inside the data configuration.

On the other hand in terms of the more classical Shannon's notion of information, each symbol of a data configuration is usually thought as a *bit of information*. This information is external (corr. to extrinseque in integrated information theory) as it concerns an external observer of the data configuration, where the in-formation is internal (corr. to intrinseque in integrated information theory).

In order to perceive the pertinence of this notion of in-information, it is useful to see in-information and (bits of) information as particular cases of the same object type. In particular one can see any causal relation $p\mathcal{R}q$ as an operator which in-forms any data configuration in which p appears that q is present. On the other hand and with a bit of translation effort one can interpret bits of informations in similar terms.

For any symbol a in the alphabet \mathcal{A} of a micro-world W , and $\mathbf{u} \in \mathbb{Z}^2$, let us denote $[a]_{\mathbf{u}}$ the set of data configurations $w \in W$ such that $w_{\mathbf{u}} = a$.

Given an unknown data configuration w , the operator acting on the set of $\mathcal{A}^{\mathbb{Z}^2}$'s subsets such that the image of a subset S is $S \cap [a]_{\mathbf{u}}$ is an in-information to the observer that the data configuration considered (if there is one) which is known to be in S has symbol a on position \mathbf{u} . One can interpret the perception of this symbol, itself an information, in a data configuration as an in-information of the observer, which acts on the observer's representation. In a sense the data configuration *contains* this in-information as a potential, which is here external and not internal.

Starting from no knowledge on the configuration, meaning that we only know that the configuration is in $\mathcal{A}^{\mathbb{Z}^2}$, we know as an end result of the operator that the configuration is in $[a]_{\mathbf{u}}$. Furthermore one obtains as the result of an infinite sequence of in-formations which specify every position \mathbf{v} to be superimposed with the symbol $w_{\mathbf{v}}$, the set $\{w\}$ which is identified with w (one arrives this way to a complete information about the data configuration). I believe that the formulation (widely used) that the data configuration w 'contains' the bit of information a makes forget that the nature of the information is of an in-information of the observer.

One can also see an internal in-formation in terms of information: if a pattern p causes a pattern q , then the pattern p contains in itself (as a potential) the presence of the pattern q .

In the following I will define some ontologies based on causal relations defined above.

4.4.2 Causal stability

The first causal ontology that I will present here is based on an intuition that is similar to the one rooting the definition of the *functional clustering index* [TMRE98], which is that in a complex system clusters are formed as an overcoming of external interactions, thought as exchanges of information, by internal interactions. The main difference with the causal ontology presented here is the way the information exchanges (in other words communication) are formalised: instead of using relative entropy (in line with classical information theory), I use the combinatorial formulation of causal relations. This slight difference makes the ontology's definition more tractable and let us envision more progress in the whole ontological process.

Notation 7. Let us consider p a pattern which appears in a configuration w . I will denote $\mathcal{C}(p, w)$ the set of patterns q which appear in w with $|q|$ minimal such that $p\mathcal{R}q$ or $q\mathcal{R}p$.

Definition 11. Let W be a micro-world on alphabet \mathcal{A} . A pattern $p \in \mathcal{A}^{\mathbb{U}}$ which appears in a data configuration $w \in W$ is said to satisfy the **first-order causal stability** condition for w when for all a sub-pattern of p such that $|a| = 1$ and $q \in \mathcal{C}(a, w)$ which appears in w , $q \subset p$.

Remark 8. Let us note that although the intuition of this definition is in line with the functional clustering index, its precise form is a bit more complex: this results from the confrontation of the intuition with phenomenological observation in a precise mathematically describable framework. In fact this confrontation does not consists in the simple projection of the intuition in an arbitrary fixed formal context but is subjected directly to the tractability condition: the formalisation choices follow then from this condition. For instance the restriction on the patterns size in Definition 11 is imposed in order to extend the ontology's compatibility domain from $\{W_{00}\}$ to $\{W_{00}, W_{01}\}$.

Notation 8. Let us denote π_c^0 the ontology such that for all micro-world W and all $w \in W$, $\pi_c^0(W)(w)$ is the set of minimal (for inclusion) patterns in $\mathcal{L}(W)$ that satisfy the first-order causal stability condition for the configuration $w \in W$.

The following proposition states that the ontology π_c^0 is compatible on domain $\{W_{00}, W_{01}\}$ with agreements A_{00} and A_{01} :

Proposition 2. We have for all $w \in W_{00}$, $\pi_c^0(W_{00})(w) = A_{00}(w)$ and for all $w \in W_{01}$, $\pi_c^0(W_{01})(w) = A_{01}(w)$. However the equivalent statement for W_{11} is not true.

Proof. 1. **For all $w \in W_{00}$, $\pi_c^0(W_{00})(w) = A_{00}(w)$.** Let us make a list of the patterns which satisfy the first-order causal stability for a data configuration of w of W_{00} .

- **If w is a blue or white plane:** it is straightforward that for any two distinct patterns a and b which appear in W such that $|a| = |b| = 1$, $a\mathcal{R}b$ or $b\mathcal{R}a$. As a consequence any pattern whose support is strictly included in \mathbb{Z}^2 does not satisfy first-order causal stability. On the other hand the whole plane pattern is causally stable, and thus minimal to be so.
- **If w consists in two half planes:** the reasoning is similar to the first point, except that it is done for each of the two half planes; the conclusion is that the first-order causally stable patterns in w are exactly the two half planes.

Since these two cases exhaust the possible data configurations in W_{00} , the set of patterns distinguished by the ontology for W_{00} in any configuration w is equal to $A_{00}(w)$.

2. **For all $w \in W_{01}$, $\pi_c^0(W_{01})(w) = A_{01}(w)$:**

- It is straightforward that any pattern a such that $|a| = 1$ in a data configuration of the dimers micro-world is not first-order causally stable. Indeed it belongs to a dimer pattern: let us denote a' the other part of this dimer. We have that a' causes and is caused by a while it is outside the pattern a .
- On the other hand, a dimer is first-order causally stable in any data configuration: indeed, for any sub-pattern a of this pattern such that $|a| = 1$ and any configuration w , the set $\mathcal{C}(a, w)$ is reduced to the other part of the dimer. This together with the first point implies that a dimer is a minimal pattern to be first-order causally stable.
- Any other pattern in the language of the micro-world can else be decomposed into dimers and thus contains strictly a first-order causally stable pattern and as a consequence can not be minimal to be so, or contains at least a half dimer and thus is not first-order causally stable.

3. However this ontology is not compatible with agreement A_{11} since the following pattern p in the agreement A_{11} does not satisfy the first order causal stability condition for any data configuration:



Indeed the symbol a colored with light gray can not be caused by a pattern whose size is 1: a symbol can not cause any other symbol which is at distance greater than one from the first

one, and one can check by listing all the possibilities that no symbol in the neighborhood of a position can cause a symbol on this one. Since for all w in which p appears the pattern highlighted with north east lines, which is not a sub-pattern of p , causes this symbol, it is in $\mathcal{C}(a, w)$. Thus it contradicts the first-order causal stability condition for p . \square

Let us consider some possible refinement in the definition of π_c^0 which is compatible with agreements A_{00}, A_{01}, A_{11} .

Notation 9. For any configuration w and patterns p and $q \in \mathcal{C}(p, w)$ which appear in w , let us denote $\delta(p, q, w)$ the number of symbols that are neighbors of q in the finite graph $G(p, q, w)$ defined as follows:

1. its vertices are the elements of $\mathcal{C}(p, w)$,
2. and there is an edge between two of these vertices if the patterns overlap non-trivially.

We call $\delta(p, q, w)$ the degree of q in $\mathcal{C}(p, w)$. Let us also denote $\mathcal{N}(p, w)$ the set of patterns $q \in \mathcal{C}(p, w)$ such that $\delta(p, q, w)$ is maximal.

Definition 12. Let W be a micro-world on alphabet \mathcal{A} . A pattern $p \in \mathcal{A}^{\mathbb{U}}$ is said to satisfy the **second-order causal stability** condition when for all sub-pattern a of p such that $|a| = 1$ and $q \in \mathcal{N}(a, w)$, q is a subpattern of p .

Notation 10. Let us denote π_c^1 the ontology such that for all micro-world W and all $w \in W$, $\pi_c^1(W)(w)$ is the set of minimal patterns which satisfy the second-order causal stability and appear in w .

Lemma 4. A pattern in the language of the world W_{00} or W_{01} is first-order causally stable for a configuration w if and only if it is second-order causally stable for this configuration.

Proof. Indeed for W any of the micro-worlds W_{00}, W_{01} and a a pattern in its language such that $|a| = 1$ that appear in a configuration w , any element q of $\mathcal{C}(a, w)$ satisfies $|q| = 1$. As a consequence these elements do not overlap and for all of these q , $\delta(p, q, w) = 0$. As a consequence, $\mathcal{N}(a, w) = \mathcal{C}(a, w)$. This implies that the first-order and second order causal stability conditions are equivalent for every w . \square

Proposition 3. The ontology π_c^1 is compatible with the agreements A_{00}, A_{01}, A_{11} .

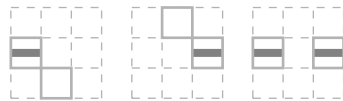
Proof.

We have that for all $w \in W_{00}$, $\pi_c^1(W_{00})(w) = A_{00}(w)$ and for all $w \in A_{01}$, $\pi_c^1(W_{01})(w) = A_{01}(w)$: this follows from Lemma 4. Let us prove that $\pi_c^1(W_{11}) = A_{11}$:

1. **Any curve pattern is minimally second-order causally stable in the configurations in which it appears.** Indeed, let us consider the pattern which consists in a symbol in the curve pattern and its Moore neighborhood in some configuration w . Let us assume that this pattern is for instance as follows:



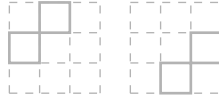
Here the central symbol a is caused by exactly the following patterns



whose support has minimal cardinality (two) and causes no pattern. In the list above the patterns the first and second ones have degree 1 in $\mathcal{C}(a, w)$. As a consequence $\mathcal{N}(a, w)$ is

reduced to the third one, which has degree 2 and is contained in the curve pattern which contains a . The same reasoning applies for any other position in the curve and any configuration w (by exhaustive listing of the 3×3 square patterns centered on a curve symbol). Moreover if any number of symbols are removed from a curve pattern without leaving it empty, consider b any of the left positions that are in the (Von Neumann) neighborhood of a removed position: the pattern in $\mathcal{N}(b, w)$ is not included in the curve pattern.

2. Let us consider any other pattern in some configuration w . If it intersects a curve without containing it, then it is not second-order causally stable. If it contains one then it cannot be minimally second-order causally stable. Any minimally second-order causally stable distinct from any curve is thus included in one of the maximally connected areas that do not intersect the curve patterns in this data configuration. Consider such a pattern and consider a symbol a in this pattern. Let us assume that the Moore neighborhood of this symbol is all-white. The patterns in $\mathcal{C}(a, w)$ are the following ones:



These two patterns have degree 0 in $\mathcal{C}(a, w)$ and thus are both in $\mathcal{N}(a, w)$. If the Von Neumann neighborhood intersects a curve, then there is an element of $\mathcal{N}(a, w)$ which intersects the curve. If the data configuration has curves, this implies, by the repetition of this conclusion, that the pattern intersects one of the curves that are near its border. As a consequence it can not be minimally second-order causally stable. On the other hand if the data configuration has no curves, the pattern considered is equal to the whole data configuration. We deduce that the set of minimal second-order causally stable patterns are is A_{11} . □

4.4.3 Causal structure

In this section I define some ontologies based on the idea that clusters in complex systems are mathematically related to the structure of the system's causal graph. This more abstract intuition is followed for instance by the more recent approach of integrated information theory ???. I will present two of these ontologies, the first one compatible on agreement over $\{W_{00}\}$ and the second one over $\{W_{01}\}$. Both of them follow a common more precise outline: thinking about distinguished patterns in terms of special orbits in the causal graph.

Let us first describe the structure of W_{01} 's causal graph by listing its simple generators.

Lemma 5. *The following pattern and its translates are simple generators of the causal graphs of W_{01} :*

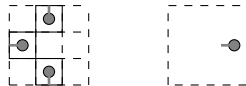


their images by rotation, as well as all the similar patterns in which the segment part of the symbols are not pointing to the symbol highlighted with light gray in the previous figure. As well the following pattern:



and all its translates and their images by rotation are simple generators of W_{01} 's causal graph. There is no other simple generator.

Proof. Let us consider a symbol a of the alphabet placed on a position of \mathbb{Z}^2 , and assume, without loss of generality, that it is the symbol \bullet . The only patterns on the Von Neumann neighborhood of this position that cause this symbol are the following ones:



For p equal to any of these two patterns, $p\mathcal{R}a$ is a generator which is simple since p does not cause any other pattern. Moreover if any pattern q causes this symbol then it causes one of these patterns: otherwise the pattern on the Von Neumann neighborhood of the position would not cause the symbol, which would mean that the pattern could not cause this symbol. As a consequence there exists p such that $p\mathcal{R}q\mathcal{R}a$ and the relation $q\mathcal{R}p$ does not hold. This implies that there is no other simple generators than the ones listed in the Lemma. \square

Notation 11. Let W be a micro-world and p, q two patterns on its alphabet. We denote $p \sim q$ when $p\mathcal{R}q$ and $q\mathcal{R}p$. This relation is said to be irreducible when both $p\mathcal{R}q$ and $q\mathcal{R}p$ are irreducible. It is said to be minimally irreducible when there is trivial pair of sub-patterns p' and q' respectively of p and q such that $p' \sim q'$.

Notation 12. Let us denote π_c^2 the ontology such that for all micro-world W and $w \in W$, $\pi_c^2(W)(w)$ is the set of patterns p such that there exists a minimal non-empty subset S of the causal graph's vertices set such that for all $s \in S$ if $q\mathcal{R}s$ or $s\mathcal{R}q$ is irreducible then q is in S , and p is the product of all the elements in S .

Notation 13. Let us also denote π_c^3 the ontology such that for all micro-world W and $w \in W$, $\pi_c^3(W)(w)$ is the set of patterns p such that there exists a minimal non-empty subset S of the causal graph's vertices set such that for all $s \in S$ and q which satisfy the following conditions (*):

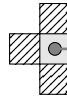
1. $s \sim q$ is minimally irreducible
2. the relations $s\mathcal{R}q$ and $q\mathcal{R}s$ are simple
3. and there is no other pattern q' such that $s\mathcal{R}q'$,

then q is in S , and p is the product of all the elements in S .

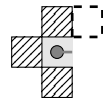
Proposition 4. The ontology π_c^2 is compatible with agreement A_{00} and π_c^3 is compatible with agreement A_{01} .



Proof. 1. **For all $w \in W_{00}$, $\pi_c^2(W_{00})(w) = A_{00}(w)$:** It is straightforward that the only irreducible relations are between symbols of the same color such that the first one is on the right of the second one if they are blue and on the left if they are white. As a direct consequence $\pi_c^2(W_{00})(w) = A_{00}(w)$.

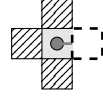
2. **For all $w \in W_{01}$, $\pi_c^3(W_{01})(w) = A_{01}(w)$:** it is sufficient to prove that the only patterns s, q that satisfy the conditions (*) are the two symbols of a dimer. Let us assume ad absurdum that there exists such patterns such that $s \sim q$ is not a reduced to a dimer. By minimality it can not contain any dimer. As a consequence, without loss of generality, one can assume that one of the simple generators in s is of the following form:



- (a) Let us assume that the lower or upper symbol is equal to $\begin{smallmatrix} \square \\ \bullet \end{smallmatrix}$ or $\begin{smallmatrix} \bullet \\ \square \end{smallmatrix}$. Without loss of generality, one can assume that this is the upper symbol. Since $s \sim q$ is minimal and $q\mathcal{R}s$ is simple, it has no dimer simple generator, and since q has to cause the upper symbol, the highlighted position on the following figure has to be in the support of q :






Moreover, this symbol has to be equal to  or  in order to not break local rules. As a consequence for similar reasons the symbol highlighted on the following figure has to be in s :



This means that the relation contains a dimer, and thus is not minimal.

(b) Let us consider now that the generator is as follows:



If the left symbol is  or  then else the upper symbol or the lower symbol can not be caused by q unless the relation contains a dimer. As a consequence, let us assume that this symbol is . In order for the upper position and the lower positions to be caused by q , and applying the same reasoning as above, the relation has to contain the following pattern:



Iterating this, one obtains that the relation has to contain the following:



This relation is minimal, but both s and q are related irreducibly to another pattern (obtained for instance by suppressing exactly one symbol from the other pattern).

□

I leave the reader on this section with the following question:

Question 1. *Is there an ontology based on the causal structure which is compatible with both agreements A_{00}, A_{01} ?*

5 Perspectives

5.1 Further research

In this last section I initiated the ontological process around some simple micro-worlds in order to prove the framework I propose. In further work on this subject one would try to extend the presented ontologies to micro-worlds with progressive complexity (in particular with respect to causality).

In the search for sufficiently simple micro-worlds one could rely on criteria such as the 'finite impact radius', meaning that any symbol can not cause any other symbol over positions that are sufficiently far, uniformly over the alphabet, which is verified by the micro-worlds considered in the last section. Ultimately, complex square tilings constructions such as R.Robinson's one [Robinson] can be considered as a benchmark for this development (in the sense that a significant ontology

should include it in its domain): although causality is non trivial in this micro-world, it seems reasonable to expect finding a relation between it and the patterns in a natural agreement in its data configurations.

In this direction I expect that the extension of the causal structure ontology will lead to a useful notion of causal structure, in the perspective of understanding information transfers in multidimensional subshifts of finite type and in dynamical systems in general.

Furthermore an analysis of Robinson tilings in causal terms would enable us to access more complex (in objectal sense of Section 3.2.4) recent and various constructions embedding Turing computations in tilings [see for instance [GS] [DRS]]. The interest of these constructions, besides being further benchmarks for the significance of ontologies, is to provide not only intuitive distinctions but also an intuitive decomposition of its configurations into 'functions' patterns that cooperate in order to enforce a certain behavior (which is the purpose of their construction).

5.2 Comments

The framework presented in the last section should not be considered as definitive but should be remodeled according to the intuitions gained in this direction. I would like also to stress the fact that the ontologies presented in the last section are not presented for their absolute meaning but as a step in a dynamical process, a long term project of combining mathematics and phenomenology in order to understand phenomenal objectness. The interested reader should feel free to search for completely different ones, as long as they have more meaning for the present discourse: just as the importance of philosophical concepts depends on their generality (and thus the possibility of using these concepts in various different situations), the importance of an ontology depends on its domain's width.

In fact one can think of the phases separation in this section as follows: each of the ontologies definition consists in a statical designation while the direction to take in the ontological process, signified by the position of particular ontologies of a certain type, is dynamical, as well as the description of some agreements for a micro-world is statical while the collection of the possible agreements as well as the most significant agreement are dynamical. As a matter of fact the discourse exhibits here an inclusion of the dynamical both intradiscursively (the agreements) and paradiscursively (the ontological progression). On the other hand the end goal of the ontological process is the creation of concepts that could be used to analyse objectness outside of the framework defined here, having a similar theoretical role as Turing machines in relation to computation in general.

References

- [F09] K.Friston. Predictive coding under the free-energy principle. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 2009, Vol. 364, Iss. 1521, pp. 1211–1221.
- [DCN11] S.Dehaene, J.-P. Changeux, L.Naccache. The Global Neuronal Workspace Model of Conscious Access: From Neuronal Architectures to Clinical Applications. In *Characterizing Consciousness: From Cognition to the Clinic?* Springer, 2011.
- [D91] D.Dennett. *Consciousness explained*. Little, Brown and Co, 1991.
- [V96] F.J.Varela. Neurophenomenology, a methodological remedy for the hard problem. *Journal of consciousness studies*, 1996, Vol. 3, Iss. 4, pp. 330-349.
- [PPRV00] B.Pachoud, J.Petitot, J.-M.Roy, F.Varela. Naturalizing Phenomenology: Issues in Contemporary Phenomenology and Cognitive Science. *Stanford University Press*, 2000.
- [P93] J.Petitot. Phénoménologie naturalisée et morphodynamique : la fonction cognitive du synthétique a priori. *Intellectica*, 1993, Vol. 2, Iss. 17, pp. 79-126.

- [A18] L.Albertazzi. Naturalizing Phenomenology: A Must Have ? *Front. Psychol.*, 2018, Vol. 9, Iss. 17, pp. 79-126.
- [Z14] D.Zahavi. Phenomenology and the project of naturalization. *Phenomenology and the Cognitive Sciences.*, 2014, Vol. 3, pp. 331-347.
- [Kahneman] D.Kahneman Thinking, fast and slow. *Farrar, Straus and Giroux*, 2011, New York.
- [Colyvan] M.Colyvan The indispensability of Mathematics. *Oxford University Press*, 2003.
- [Dennett] D.Dennett Real Patterns. *Journal of Philosophy*, 1991, Vol. 88, Iss. 1, pp. 27-51.
- [Pearl] J.Pearl Causality: Models, reasoning, and inference. *Cambridge University Press*, 2000.
- [AOT] L. Albantakis, M. Oizumi and G. Tononi. From the phenomenology to the mechanisms of consciousness: Integrated Information Theory 3.0 *PLoS Comput Biol*, 2014, Vol. 10, Iss. 5.
- [Haun Tononi] A. Haun and G. Tononi. Why does space feel the way it does? Towards a principled account of spatial experience *Entropy*, 2019, Vol. 21, Iss. 12.
- [M.Pokropski] M. Pokropski Phenomenology and mechanisms of consciousness: considering the theoretical integration of phenomenology and mechanistic framework. *Theory and Psychology*, 2019, Vol. 29(5), p. 601-619.
- [Bayne] T. Bayne On the axiomatic foundations of the integrated information theory of consciousness. *Neuroscience of consciousness*, 2018, Iss. 1.
- [MMMG] P. Maguire, P. Moser, R. Maguire, V. Griffith Is consciousness computable? quantifying integrated information using algorithmic information theory. *Proceedings of the annual meeting of the cognitive science society*, 2014, Vol. 36.
- [Laudan] L.Laudan The Demise of the Demarcation Problem. In Cohen R.S., Laudan L., Physics, Philosophy and Psychoanalysis. *Boston Studies in the Philosophy of Science*, vol 76. Springer, 1983.
- [Thagard] P.R.Thagard Why Astrology is a Pseudoscience. *Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1978, pp. 223-234.
- [TMRE98] G.Tononi, A.R.McIntosh, D.P.Russell, G.M.Edelman Functional Clustering: Identifying Strongly Interactive Brain Regions in Neuroimaging Data. *NeuroImage*, 1998, Vol. 7, Iss. 2, pp. 133-149
- [Z19] D. Zahavi. The practice of phenomenology: The case of Max van Manen. *Nursing philosophy*, 2019, Vol. 21, Iss. 7.
- [Robinson] R. Robinson. Undecidability and nonperiodicity for tilings of the plane. *Inventiones Mathematicae*, 1971, vol. 12, pp. 177–209.
- [DRS] B.Durand, A.Romashchenko, A.Shen. Fixed-point tile sets and their applications. *Journal of Computer and System Sciences*, 2012, Vol. 78, Iss. 3, pp. 731-764
- [GS] S.Gangloff, M.Sablik Quantified block gluing, aperiodicity and entropy of multidimensional SFT. *Journal d'Analyse Mathématique*, to appear.