Continuing the search for structure of the experiencing subject

Abstract

In the last decades, the development of advanced imagery techniques made possible a better understanding of the functioning of the brain as well as the formulation of cognitive theories on how conscious experience may rise from its activity. However, it is sometimes challenging to distinguish which of these theories are actually about consciousness (addressing '*easy*' problems instead of the hard problem). In this text, I put into evidence that, for two prominent of these theories, what makes them theories about consciousness are tacit relations between parts of the brain and part of the mind, and that these relations rely on mapping structures of these parts. While conceptualising this approach, I address some shortcomings of its implementation by the mentioned cognitive theories. It appears that the most important difficulty in implementing it is the lack of a method for discovering more structure in the mind which may be related to physical activity through structure mapping. I propose that this could be done by an axiomatisation of the structure of judgements about the experiencing subject which irreducibly involve a point of view.

This text is divided in two section. Section **I** exposes the structure-mapping strategy, presents two prominent cognitive theories of consciousness and discusses the importance of structure-mapping in what makes these theories about consciousness. This discussion addresses the shortcoming of their use of structure-mapping. Then Section **II** details and justifies the main proposition of this article, which is axiomatising the structure of judgements about the experiencing subjects which irreducibly involve a point of view.

I. On the structure-mapping strategy

Understanding how conscious experience may rise from some particular physical activity of a body means having unveiled a relation of causal nature between this physical activity and the *existence* of a mind related to the body. This is difficult for the evident reason that we do not have access to a mind in the course of its development. However it is in principle possible to see how a particular physical activity causes a particular *mental content*, for we do have access to

situations in which such a mental content is present and situations in which it is not, and thus can conjecture and potentially falsify a causal relation involving this mental content.[1] Ultimately it may be possible to conjecture how physical activity causes the mind's existence from a better understanding of how it causes mental contents. Furthermore, establishing strong relations (such as ones of causal nature) between beings in two huge domains of beings is only practically possible once we have an idea of which ones are related in a looser sense. This is where structure comes into play.

To be clear, I call **structure** any set of *elements* and *relations* between these elements. For instance, a neural network can be thought as a structure whose elements are neurons, and for which synapses, as contact areas between neurons, are thought as relations between these neurons. Areas in the visual field are also the elements of a structure, where there is a relation between two areas when one is included in the other. A **mapping** between two structures $\mathcal{S}_1$ and $\mathcal{S}_2$ is a function which associates an element (resp. relation) of $\mathcal{S}_2$ to each element (resp. relation) of $\mathcal{S}_1$, such that this function *preserves* the structure in the sense that a set of elements have a relation between them in $\mathcal{S}_1$ if and only if the corresponding elements are related by the corresponding relation in $\mathcal{S}_2$.[2]

The existence of a mapping between two structures does not imply that they are one and the same.[3] However the existence of a mapping between two structures may be seen as a hint that there may be another stronger relation of different nature between them (this relation may then be discovered by acquaintance). Such a relation could be an ontological identity (they are one and the same) or a relation of causal nature such as for instance, for a structure $\mathcal{S}_1$ found in the mind and a structure $\mathcal{S}_2$ found in the brain: whenever an element of $\mathcal{S}_1$ causes a physical event, the corresponding element in $\mathcal{S}_2$ also causes it.

The recourse to structure mappings in order to relate domains of things which have different nature (for instance mental contents and neural structures in general) is appealing and it is difficult to see a different general way to make such connections.[4] However this recourse if often tacit, and the difficulties it poses are thus not apparent. In the remainder of this part, I will present the two main cognitive theories of consciousness currently under spotlight, namely *Integrated information theory* and the *Global workspace theory*, in order to put into evidence how the structure-mapping reasoning appears, and the difficulties that these particular uses pose.

---

[1] Simple collections of such causal relations correspond to what D.Chalmers called *psychophysical laws*.

[2] The notion of mapping described here resembles the one of *isomorphism* in mathematics, however is thought here not as limited to mathematical objects, but applicable to parts of the experience of a subject.

[3] For a simple example in mathematics, one can consider the structures whose elements are respectively the ones of the sets $\mathbb{N}$ and $2\mathbb{N}$, where $\mathbb{N} = \{0, 1, 2, \ldots\}$ and $2\mathbb{N} = \{0, 2, 4, \ldots\}$. The relations are the pairs $\{m, n\}$ of $\mathbb{N}$ (resp. $2\mathbb{N}$) such that $m = n+1$ (resp. $m = n+2$). There is a mapping between these two structures which associates $2n$ to all n in $\mathbb{N}$, however it is clear that the two structures are different, because $\mathbb{N}$ is different from $2\mathbb{N}$.

[4] To be clear, I do not mean to say here that "all that there is, is structure" but that

Integrated information theory. — *(i) Presentation.* – Integrated information theory (IIT) is a theory about consciousness which has been developed by G.Tononi and collaborators during the past two decades and is known for its formalisation of conscious experience and its use in order to find the relation between the experience and the functioning of the brain. The origin of the theory lies in the work of O.Sporns, G.Tononi and G.M.Edelman in 1994 [7] in which is proposed a mathematical quantity, that they called *neural complexity*, meant to quantify, for a neural network, the balance of between *functional segregation* and *integration*. These two aspects of neural network correspond respectively to how much the network is divided into functional modules and how much these modules are interdependent (in terms of information). The discussion suggested that neural complexity is high for connectivity patterns exhibited by areas of the brain associated with consciousness, such as the cerebral cortex, so that in principle it would be possible to detect conscious experience of a physical system using (an equivalent of) neural complexity. This is still one of the main statements of IIT, although the quantity has been updated multiple times (the last definition can be found in [8]), in particular formalising in a sound way communication between parts of the network using *causation*, where an exchange of information is a causal relation between elementary events. This quantity would therefore measure how much a physical system is conscious, and this system would be conscious if the quantity is > 0. Furthermore the set of causal relations of the physical system has a structure in which it would be possible to read the complexes of concepts and phenomenal events constituting conscious experience.

It matters to notice that there is no mathematical proof, at least to the author's knowledge, for the conjecture of O.Sporns, G.Tononi and G.Edelman that neural complexity is the highest for architectures exhibiting most balance between functional segregation and integration.[5] Furthermore, although IIT provides a extensive conceptual apparatus, all the identities that it formulates are as speculative as this. The search for foundation, and concomitantly the idea that it is difficult to see how the human brain may generate conscious experience by considering it only and not the experience itself, led G.Tononi and C.Koch [9] to put forward a restricted set of propositions on conscious experience from which the theory stems. They claimed that these propositions are independent of any particular experience, meaning in principle that they are only about *experiencing* itself, and that they are self-evident[8]. This is the reason why they baptised them *axioms*. The derivation of the theory from these propositions is done through an association between each of these axioms and another corresponding proposition, called *postulate*, which is meant to express the underlying idea of the axiom in causal terms as a property that a neural network has if and only if it participates to the generation of conscious experience, thus constraining further speculation.

---

structures are probably the only tool there is to find relations between things which are not a priori related.

[5] Let us note here that later versions of this quantity may be the highest for other types of

(ii) Structure-mapping must be mathematically binding. – More recently, G.Tononi and A.Haun have applied the framework of IIT to the construction of a correspondence between some neural structures in the visual cortex which are correlated with processing of spatial information, namely *grid cells*, and the spatiality of conscious experience [1]. More precisely they found a correspondence between connected areas in the visual field and connected sets of neurons in a grid-like neural structure, which can be found naturally within the causal structure of this network. These are the elements of two structures $\mathcal{S}_1$ and $\mathcal{S}_2$. In each of these structures, elements are related to each other when one is included or includes another or intersects non-trivially another. Although the two structures can be mapped one onto the other, any set of things bear a similar structure on the elements of the set of parts of this set. Thus any two sets of things have structures which can be mapped one onto the other in the same way. This contradicts the purpose of structure-mapping made explicit above. In general, for a structure mapping between a structure $\mathcal{S}_1$ and a structure $\mathcal{S}_2$ to be meaningful, there should at least be other things which bear no structure which can be mapped to $\mathcal{S}_1$ or $\mathcal{S}_2$. In this case we say that the mapping is *mathematically binding*.[6]

This reveals one general difficulty of the structure-mapping strategy, which is that provided a structure within the brain, it is always possible to find a similar structure in the mind to which it can be mapped, for when thinking about the former, the mind creates a mental copy by *simulating* it. It is intuitive that nothing is learnt about the relation between the mind and the brain from such a mapping, and additional criteria are required on the kind of structures one should consider for this purpose. This difficulty is accentuated in the case of space that G.Tononi and A.Haun consider, which is, to be clear, not the space we live in, which is intuitively three-dimensional, but the space of experience as it is displayed to the conscious subject, which is two-dimensional. The reason is precisely that this sense of spatiality is encountered in all experience.[7]

(iii) Critics of the foundations. – The fact that the correspondence put forward by G.Tononi and A.Haun is not mathematically binding is not surprising when considering the shortcomings of IIT foundations. The idea of G.Tononi and C.Koch to ground the search for an explanation of how the brain generates conscious experience in experience itself is convincing, because it is not possible to explain anything (in particular phenomenal experience) without observing it beforehand. However little investigation has been done on these foundations. And as a matter of fact they fall short from another epistemological point of view which Kant has expressed in the Critique of pure reason:

"*Now it may seem natural that as soon as one has abandoned the territory of experience, one would not immediately erect an edifice with cognition that*

---

systems.

[6] The critics presented here are similar to the ones of M.H.A.Newman [20] about B.Russell's structuralism (page 3).

[7] The same remark applies to time. In my interpretation, this presence in all experience, of space and time for instance, is what Kant meant in his concept of pure a priori forms.

one possesses without knowing whence, and on the credit of principles whose origin one does not know, without having rest 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 cognition a priori and what domain, validity and value they might have."

Indeed, if the discourse leaves the ground of the experience of its matter, one should expect it to have little probability to have descriptive or explaining power about this matter (as it actually is). Beginning with subjective experience is not enough, one also has to resist the impulsion (cultivated by the method of science) of leaving it.

There are two points where IIT depart from this ground. The first one is the formulation of the axioms, which were the object of some critics by T.Bayne [6]. The main argument of T.Bayne is that each of these axioms fails to be an axiom, else because it is not self-evident, for it is not universally accepted, or it is not constraining, meaning that it is epistemologically sterile. One may answer to this argument that the formulation of these axioms is not understood well, however this would *still* mean that the axioms fail to be self-evident.[8]The second point is the translation of axioms into postulates, which relies on the assumption that propositions on conscious experience correspond to physical properties expressed in causal terms. Even if this is the case, there may be multiple ways to express them in such terms which should be directly investigated.

Global workspace theory. — *(i) Presentation.* – The Global workspace theory is a theory of consciousness which originates in A. Newell's research of a unifying interpretative framework for the functioning of the brain, in other words a *cognitive architecture* (see for instance [17]). A.Newell proposed the Soar architecture together with his collaborators J.Laird and P.Rosenbloombased, which embodies several hypotheses about computation structures underlying general intelligence, in particular the *blackboard architecture* artificial intelligence paradigm for solving problems using the cooperation among a set of specialist knowledge holders through a centralized knowledge called *blackboard*. The connection between this paradigm and empirical evidence from the brain was later made by B.Baars [10] and presented as a theory of consciousness, namely Global workspace theory (GWT). As a theory of consciousness, it relies on a structure-mapping, which consists in an analogy between the mind and a

[8]We will see below that these axioms differ on other points from mathematical ones, such as their construction history. I believe that in general we should be careful when calling axioms some propositions. Regarding this it may be fit to recall another excerpt from the Critique of pure reason: "*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*". On the other hand it is possible to abandon the foundational role of axioms, leaving them with the one of *regulating idea* used in order to create concepts and a formal apparatus for investigating the relation between mind and body. However this leaves room for many other potential formal approaches.

theater. B.Baars describes it in the following way:

"*Consciousness in this metaphor resembles a bright spot on the stage of immediate memory, directed there by a spotlight of attention under executive guidance. Only the bright spot is conscious, while the rest of the theater is dark and unconscious.*"

where the executive guidance corresponds to the director, scriptwriter and similar roles holders, who take decisions behind the scene. This metaphor has been developed further, including in particular the idea of competition among the audience of the theater sitting in the dark, for having the occasion to appear on the stage. In the words of B.Baars, "*this analogy leads to specific neural hypotheses*" about the involvement, in the theater's dynamics, of some parts and processes of the brain and interactions between them. This led to a conceptualisation of the brain architecture and its dynamics, centered around a *global workspace* which may correspond to the cortico-thalamic complex. This global workspace is used by other parts of the brain acting like specialized and autonomous processors - specialized in processing spatial or temporal information for instance - for sharing information globally in order to coordinate their functioning, some of them having a singular role such as directing the spotlight of attention (sensory projection areas of the cortex), and decisions about the play of the actors (prefrontal cortex). According to the model, an information becomes conscious when it is shared across the global workspace, as a result of competition of local processors cooperating in groups, in order to access the global workspace and share the information they hold, a process which was called *ignition* by B.Baars. On the other hand the global availability of information would explain that conscious information is easily accessible by the subject and in principle reportable. Later, J.-P. Changeux and S.Dehaene, building on GWT, searched for more precise mechanisms of this architecture, in particular proposing a model, known as Changeux-Dehaene model, which is centered on self-organized criticality in neural networks. This is the refinement of GWT which has received most attention in the past decades.

(ii) An analogy is a structure-mapping. — Overall it is the theater analogy which makes GWT a theory about *consciousness*, and one may see that, besides the functional coherence of all the elements of the theater, it is the use of structure-mapping, between a mind structure identified metaphorically and a brain structure, which makes it convincing. In order evaluate the strength of this analogy, it is worth taking a closer look at how this structure-mapping is constructed. The correspondence between the elements of the structures in question is often made explicitly (for instance the 'stage' corresponds to the cortico-thalamic complex). On the other hand the analogy relies implicitly on relations between these elements. These relations are of topological and functional nature: the centrality of the stage (resp. cortico-thalamic complex) and periphery of the audience, director, scripwriters, etc (resp. the prefrontal cortex, and other specialized local processors); the dark and light, refering to the unaccessibility and accessibility of information (resp. shared information through

the workspace and local one); the uniqueness of the stage and multiplicity of the audience (resp. multiplicity of local processors and uniqueness of the global workspace); the competition between memories in the mind to access conscious experience; the dynamics between the conscious and unconscious (resp. global workspace and local processors) in particular how a memory or a signal becomes conscious (ignition).

(iii) Discovering mind structures requires theoretician's introspection. — The topological and functional structures on which the theater analogy relies are identified and recognized through introspection. However they can hardly be considered as evident in the sense that they would be readily observable by anyone with a fresh mind about this matter. In other words their identification and recognition require a certain effort. It is in principle possible these structures are only mind constructions and are not *real*, in the sense that they must be found by any introspecting subject, even if they do *seem* evident to theoricians in the field, especially so when they were constructed a long time ago and passed through culture.[9] This is a common reason in cognitive sciences to doubt from theoretical apparatus which result from introspection. However the search of a relation between the mind and the physical requires, by definition, introspection. Although S.Dehaene defends a rehabilitation of introspection in the methodology of cognitive sciences, it is restricted to interview with subjects according to methods which are meant to be objective, such as for instance the micro-phenomenological interview [13], the role of introspection in the foundational analogy of GWT is left unquestioned. I think a proper examination of these foundations would reveal that the *theoretician's introspection* (by opposition to subjects introspection and report), that the theorician uses in order to develop a theory, is necessary.[10] In place of rejecting the theorician's introspection as a method because it can not be relied upon, we should search for conditions of reliability of this form of introspection and determine its use more clearly. That is, not to search for objectivity immediately but instead *visibility* on the mind structure, even if this visibility is achieved with imperfect tools.

II. The structure of judgements and its axiomatisation

In Section **I**, we have identified three main difficulties in the use of structure-mapping: 1. the absence of a method for discovering mind structures; 2. the fact that mind constructions interfere with this strategy; 3. counter-intuitively, the property of independence of mind structures from the particular experience

[9] For instance H.Berson [12] was writing about the competing memories for appearing in the conscious experience.

[10] One of the reasons is, B.Baars wrote, that the study of consciousness necessitates a global approach. Another is that a theory, like GWT, is not really a theory of consciousness but only an interpretative framework for investigating the architecture of the brain and the dynamics of its functioning.

does not make it easier. I will begin this section with presenting the idea of considering judgements in order to find structure in the mind. We will put into evidence that their (complex) structure can be introspected with examples, and see along the way how the other difficulties are addressed in this setting. Other reasons for considering judgements is that it is possible in principle, through the structure, to find relations between judgements about consciousness and ones about the physical via physical judgements. Another reason is that explaining consciousness comprises explaining judgements.

The structure of judgements. — Before all, I call *judgement* a proposition of an experiencing subject about their experience that they believe in.[11] Let us see what structure can be found in the set of judgements of a subject.

Before turning to general judgements, consider mathematical propositions. Such a proposition is a judgement about mathematical objects, where these objects are mental contents resulting from instanciation in the moment of experience of abstract concepts. For example, one can instanciate a number n by thinking about a series of n vertical bars. When I add n to m and obtain $n+m$, I operate a concatenation of n bars and m bars. A proposition such as "for all n positive integers, the sum of the first n positive integers is $\frac{n(n+1)}{2}$" is a judgement which posits a causal relation between mental contents corresponding to integers and operations on them: if on one hand I consider an integer n, then enumerate all the numbers between 1 and n and add them progressively, and on the other hand I add n (resp. $n+1$) times the number $\frac{n+1}{2}$ (resp. $\frac{n}{2}$) if n is odd (resp. n is even), then the two resulting numbers are equal. Let us denote by P_n the proposition that the sum of the n first positive integers is $\frac{n(n+1)}{2}$. I believe that it is the case that P_n for all positive integer n because it is the case that P_1 and that if P_n then P_{n+1}.[12] In other words, the belief in the judgement that all P_n is true results from the *composition* of the beliefs in other judgements.

Each such a composition is a relation between judgements, and one can see that this type of relation can be found in a more general setting than only mathematical propositions, and all these relations provide a structure to the set of all judgements.

Judgements which irreducibly involve a point of view. — This provides structure within the mind of any subject capable of judgement. However this structure is too general to teach us anything about the consciousness of such subject. For this we need to restrict to judgements which are *about* consciousness. As a working criterion, I will say that a judgement is about consciousness when it involves irreducibly the point of view of a particular subject. For instance, any statement about '*what it is like*' to be a certain subject or a subject in a category of subjects involves irreducibly a point of view[13], precisely because

[11] I shall refrain from making this definition more precise. The reason is that the justification of additional precision with examples would engage me into exhibiting their structure.

[12] This is called a proof by recursion.

[13] For instance T.Nagel, who defined[15] a subject of experience as whatever of which can be

the point of view is its object. In contrast with this, judgements about physical activity are - one could say by definition - judgements which are independent from the subject who is judging and thus independent from any point of view. This is also the case for judgements about what can be found only within a particular point of view. One example is '*I understand Chinese*'. Of course another subject may have hints that I do or do not understand Chinese, but only me have access to the fact that I do not. The reason is that understanding a language is the possibility to form images in one's mind of mental contents corresponding the words, and after forming a mental response to these mental contents, forming words which reflect this response. Only I can judge if I have this possibility.[14]

Why is the hard problem hard ? — The hard problem of consciousness [4] is the one of explaining why and how human beings have phenomenal experience. As the hard problem is hard, D.Chalmers [3][15] formulated recently the meta-problem of consciousness which consists in an approach of the hard problem. This consists in explaining why we think that the hard problem is hard. Answering this would else dissolve or clarify it.

As a side note to the current paper, one can use similar ideas as the ones above in order to attempt an answer of the meta-problem. For this I will define phenomenal experience as what in a point of view is not accessible from the point of view of another subject. I believe that the hard problem is hard for the reason that the sciences have thus far focused on parts of experience which are *objective*. In other words, judgements on these parts of experience are independent on the point of view. This excludes by definition judgements which involve irreducibly a point of view. Therefore, one can hardly assume that the current methods of science can be effective in order to explain phenomenal experience. The hard problem is hard for the reason that our thought mechanisms are not adapted to it.[16]

Axiomatisation. — In a short unpublished text [2], the mathematician

---

said that there is '*something it is like*' to be it, proposed the bat as an example of subject of experience whose experience is not accessible by a human being in his or her own experience, for it has a sense (sonar) that humans can't conceive what if feels like to have it.

[14] I believe that it is what J.Searle's *Chinese room argument* is meant to put into evidence. This argument consists in picturing oneself in a room which has only two openings for inputs and outputs and, following predefined rules, performing a translation from one language to another without understanding any of these languages. From the outside it would seem that the room understands both languages (for it to be able to translate from one to the other), while it is clear that nothing contained in the room actually understands them.

[15] Later he extended this idea to other judgements about consciousness, including debunking arguments for illusionism [5], tweaking their epistemological role and ontological reach (becoming more neutral), as well as other philosophical positions, such as the possibility of philosophical zombies, or that machines are not conscious. It makes this line of research close to the one proposed in the current paper.

[16] J.Searle wrote in [14] (p. 114), another methodology is required: "*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*".

M.Gromov asked: '*what kind of mathematics we need to speak about the mind ?*'. In a sense G.Tononi and C.Koch addressed this question before it was formulated by M.Gromov, by formulating what they called axioms (of conscious experience) and defining mathematical concepts in order to formulate properties that neural networks should have in order to be correlates of consciousness. I have already argued that the possibility to make use of such mathematical concepts in order to understand consciousness is tied to the strength of the theory's foundations (G.Tononi's and C.Koch's axioms). The strength of axioms in mathematics lies in the fact that they are self-evident, but also in the fact that from the combination of these axioms one can derive in principle all mathematical theorems. While T.Bayne has put into evidence that G.Tononi's and C.Koch's axioms fail to be self-evident, they also differ from the second aspect. One way to get around this is to find inspiration in the historical development of mathematics: the axioms of mathematics, such as Peano axioms for instance, come historically after theorems, but reducing progressively the set of all proven mathematical statements to the smallest possible set from which all can be proven by composing ones from this small set. The statements in this small set are called axioms, and they were a posteriori characterized as self-evident because they can't be decomposed or refuted. What allowed mathematicians to find them is the process I just described, that I shall call *axiomatization*, rather than the search by introspection using only the idea of self-evident judgements about conscious experience.

By axiomatizing the set of judgements of a subject about its experience involving irreducibly a point of view, we would have an understanding of what conscious experience is, and how it is structured. Furthermore, we could reasonably focus a set of axioms identified this way in order to find relations between them and parts of the brain. Then we could hope to find mappings between structures of the mind and of the brain.

Natural judgements. — Continuing the analogy with mathematics as a field, axiomatization should begin with '*natural*' judgements. By this I mean judgements about the conscious experience of a subject by itself which are shared collectively and are the result of a pretheoretical natural selection by intersubjective disagreement. In mathematics, these judgements are called *conjectures*. The belief in a conjecture results from continuous consideration of mathematical objects. For instance consider the function from the set of positive integers to itself which to any $n \geq 1$ associates $n/2$ if n is even and $3n+1$ if n is odd. Let's say I applied repeatedly this function to an integer and I observe that whatever the integer I take, I always fall on the integer 1. Then I will believe that this is true for all integer chosen at the beginning. If I cannot prove that it is false, I will put forward this belief to other mathematicians.[17] Known conjecture are the ones which stayed conjectures for a long time, selected by their resistance to proof or disproof.

Examples of natural judgements involving irreducibly a point of view could

[17]The one I just mentioned is called Collatz conjecture.

be: I live (or am) *in* a space'; 'this space is three-dimensional'[18]; 'all computations I can do can be done by a Turing machine and reciprocally' (Church-Turing thesis); 'I see *through* my eyes'; '*This* is an illusion', 'I distinguish *this* pattern in my experience'.[19]

Decomposing some natural judgements. — Let us consider two of these judgements and see how they can be decomposed into other more elementary ones. *(i) The space I live in is three-dimensional* – In order to see how this judgement can be decomposed into other judgements, I can consider how I may *verify* that it is indeed the case. This leads me to the idea that the space I live in is three-dimensional because the visual experience I have is left invariant by certain actions. For instance let's say I see a tree in front of me. If I turn around myself I will see this tree again at some point. Furthermore, this fact does not depend on the experience I have at the beginning: the tree could be a bear, or something else. Of course the tree may not be in the same exact position, if the wind is blowing for instance. However the experiences are the same relatively to a certain conceptualisation of them, like "a tree with nothing around": the two experience have both this same conceptualisation. Another instance: if I move closer to the tree and then farther, then I will see at some point the tree is a similar way as I saw it at the beginning. From there it should not be hard to see that the judgement that I live in a three-dimensional space derives from a collection of more elementary judgements as the ones above, which have the form of a causal relation as follows:

$$x \ \& \ \text{action} \Rightarrow x$$

for a certain fixed action and x is any experience. As a matter of fact these relations are what one uses in order to navigate in space. *(ii) 'This is an illusion'* - Let us say for instance that I see a mountain floating above the horizon in the desert. I believe this is an illusion. How do I come to believe so ? Suspending all I know, I can imagine two possible worlds. One in which if I go in the direction of the mountain until I can touch it, put my hand below it to verify that it is not held by anything invisible and is indeed floating in the air. That would be hard to believe but if this was the case, I would ultimately change my belief about what is possible or not in the world and I would not call this experience an illusion. The other world is the one we live in, in which I see the mountain disappear on the way or see it further away. Therefore the mountain does not exist, or it does so but not where I thought it would be. The vision I had was an illusion. In other words, the reason why I think this is an illusion is that I have experienced an image of a mountain and believed that by moving towards what I saw for an amount of time that I estimated at this moment, I have another experienced that I pictured in my mind at the same moment. Then I judged that my experience is different from the one I pictured.

[18] As observed by G.Tononi and A.Haun [1], spatial aspects of experience are more easily introspectable, which makes them a good departing point.

[19] Note that these judgements involve a point of view, under the form of I, in an irreducible way.

Detour on illusionism. — I believe that this simple thought experiment shed light on what an illusion actually is. From this I would like to provide some criticism about Illusionism. Illusionism about consciousness is the view according to which conscious experience is partially illusory (*weak* form) or entirely so (*strong* form). Proponents of this position are D.Dennett and K.Frankish [18] for instance. I believe that weak illusionism is trivial, for the reason that (common) illusions are experiences therefore conscious experience is at least partially illusory. As for illusionism, I believe that it is based on misplaced doubt. In the simple thought experiment I presented before, there is illusion. However one has no reason to doubt that the experience I have at the beginning (the image of the mountain) is real. What is not real is the object I think I will see if I move towards what I see at this moment. In other words, the idea that '*if I move towards what I see, the experience I am having will change in the same way as if there is a mountain there*' is not correct. Furthermore, strong illusionism is not consistent: if one doubts about all that conscious experience contains, should'nt one also doubt about judgements such as '*this experience is an illusion*' or '*conscious experience is illusory*' ? I conjecture, contrary to illusionism, that only *expected* experiences can be unreal, and that when it is the case, it is for the reason that a causal relation which is part of the model of the world one has in mind is being contradicted by experience.

The causal structure of the subject's world. — One may observe that the judgements considered before are decomposed into more elementary judgements about *causal relations*, some of which are positive (certain events, including actions, cause others), some of which are negative (certain events do not cause others). I conjecture that this is general, meaning that all judgement that one can decompose can be decomposed into causal judgements. I also conjecture the construction of a judgement from causal ones is determined by the structure of the set of these causal judgements.

Furthermore, I believe that the method used in order to put into evidence the role of causation in the illusion judgement, which is to imagine how the world could be different than it is, is applicable to a wide set of judgements: the ones which are about the world of the subject, meaning the set of what this subject considers as *possible* experiences. The reason is that this world admits a description in terms of causal relations between *instant* experiences.[20]

[20]Such thought experiments are similar to the ones that D.Dennett explores in '*Where am I ?*' [19]. In this short story, D.Dennett has presented in 1978 a thought experiment in which he imagined that, for the purpose of an experiment, he had his brain disconnected from his body and reconnected via radio links. This thought experiments breaks the conception that we have of our personal worlds and consequently some unthought intuitions such as '*I am in my head*', or '*I occupy a connected area of space*', '*I is unique*'. I have not found a use of this story as an argument for illusionism, but it could perfectly be so. Although it is not clear how precisely the conceptualisation of the world that one holds would be changed in this situation, the reflexion of D.Dennett in the story puts into evidence that it would indeed change, because there would be reasons to believe that *I am where my brain is* (here not in my head) or *I am where my body is*. It could be both, but then the area of space I occupy would not be connected, and I would multiple: the one which determines my reactions to

Differences with IIT. — The approach of consciousness proposed in the present article is close to the one of IIT, for they both rely on the core concepts of causation and axioms. However there are major differences that I shall highlight. *(i)* While IIT proposes a relation between the structure of concepts and the causal structure of neural networks in the brain, here causation is involved in elementary judgements into which judgements are decomposed, in other words this causation is in the mind and not in the physical activity of the brain. I believe that there is a theoretical gap between these two domains of causation that IIT jumps uncarefully. *(ii)* In terms of method, the proposed approach is expected to arrive at some axioms (of consciousness). Contrary to the ones of IIT, their importance (preventing the effect of artificial mind constructions that we have discussed before) and explanatory power are ensured by the axiomatisation process.[21] Furhermore, IIT relies on the idea that axioms should be independent of any particular experience, while axioms are here expected to be on the contrary *dependent* on the particular world of the subject as a whole.[22]

References

[1] **G.Tononi, A.Haun**, *Why Does Space Feel the Way it Does? Towards a Principled Account of Spatial Experience*. Entropy 2019, 21 (12), 1160

[2] **M.Gromov**, *Great Circle of Mysteries: Mathematics, the World, the Mind*. Unpublished

[3] **D.Chalmers**, *The meta-problem of consciousness*. Journal of Consciousness Studies, 25, No. 9–10, 2018, pp. 6–61

[4] **D.Chalmers,** *Facing up to the problem of consciousness*. Journal of Consciousness Studies, 2, No. 3, 1995, pp. 200–219.

[5] **D.Chalmers** *Debunking arguments for illusionism on consciousness*. Journal of Consciousness Studies 27 (5-6):258-281 (2020)

[6] **T.Bayne**, *On the axiomatic foundations of the integrated information theory of consciousness*. Neuroscience of Consciousness; 18, No. 1, 2018 Jun 29;2018(1)

events in the world and my personality, and the one which corresponds to the point of *view* I have on the world. D.Dennett explores extensively along the story the conceptualisation switches that could possibly occur. What he does not do, and would be interesting to do, is analysing *how* we would reform similar judgements in this situation.

[21] It is possible though for the axiomatisation process to result in some reformulation of the axioms of IIT. The other way around, one could wonder how IIT axioms may be recasted in a way that are similar the ones found by aximatisation.

[22] In a similar way, to begin with simple objects in order to compare the physical and the mind is natural, however not very effective. The reason is that more complex objects are more constraining.

[7] **G Tononi, O Sporns, and G M Edelman,** *A measure for brain complexity: relating functional segregation and integration in the nervous system.* Proc Natl Acad Sci U S A. 1994, 91(11): pp. 5033—5037.

[8] **G.Tononi et al.** *Integrated information theory (IIT) 4.0: Formulating the properties of phenomenal existence in physical terms.* Unpublished

[9] **G.Tononi, C.Koch**, *Consciousness: here, there and everywhere?* Phil. Trans. R. Soc. B 370.

[10] **B.Baars** *A Cognitive Theory of Consciousness* Cambridge University Press, 1998, Cambridge.

[11] **B.Baars** *Global workspace theory of consciousness: toward a cognitive neuroscience of human experience?* in Progress in Brain Research, Vol. 150.

[12] **H.Bergson** *Dreams.* T.F. Unwin, 1914

[13] **M.Bitbol, C.Petitmengin** *Neurophenomenology and the Microphenomenological Interview.* In Susan Schneider & Max Velmans (eds.), The Blackwell Companion to Consciousness. Chichester, UK: Wiley. pp. 726–739 (2017)

[14] **J.Searle** *The mystery of consciousness.* Granta Books (1990)

[15] **T.Nagel** *What Is It Like to Be a Bat?* The Philosophical Review, Vol. 83, No. 4 (Oct., 1974), pp. 435–450

[16] **J.Searle** *Minds, Brains and programs.* The Behavioral and Brain Sciences, vol. 3, Cambridge University Press, 1980.

[17] **A.Newell** *Unified Theories of Cognition.* Harvard University Press, 1990, Cambridge, Massachusetts.

[18] **K.Frankish** *Illusionism as a theory of consciousness.* Journal of Consciousness Studies 23 (11-12):11-39 (2016)

[19] **D.Dennett** *Where am I ?* In Brainstorms. MIT Press (1978)

[20] **N.H.A.Newman** *Mr. Russell's causal theory of perception* Mind 37 (146):26-43 (1928)