

# Notes on the Guidelines of the 2009 Online Workshop on Theories of Consciousness

Alfredo Pereira Júnior\*

## Abstract

The *Online Workshop on Theories of Consciousness* (OWTC), a set of live online (text only) discussions carried out in 2009 at the site *Nature.com*, was an attempt to organize and find minimal agreement on the epistemological bases for scientific and philosophical research on the Theory of Consciousness. This paper covers the guidelines I proposed to the group; some issues discussed after the presentation of the first seminars on “Models of Consciousness”, by Anil Seth, and “Cognitive Theories of Consciousness”, by Vincent de Gardelle; as well as some discussions with the participants about these topics. Following the Concluding Remarks, I added an Epilogue written some years after the Workshop.

Keywords: Consciousness; brain; physiology; epistemology; neurons; astrocytes; explanatory models.

---

\* Universidade Estadual Paulista Júlio de Mesquita Filho/São Paulo State University (UNESP), Professor, Programa de Pós-Graduação em Filosofia, Campus de Marília, <https://orcid.org/0000-0002-5960-041X>

# Notas sobre o *Workshop Online* sobre teorias da consciência realizado em 2009

## Resumo

O Workshop Online sobre Teorias da Consciência (OWTC), um conjunto de discussões online ao vivo (somente texto) realizado em 2009 no site Nature.com, foi uma tentativa de organizar e encontrar um acordo mínimo sobre as bases epistemológicas para a pesquisa científica e filosófica sobre a Teoria da Consciência. Este artigo abrange as diretrizes que eu apresentei para o grupo; algumas questões discutidas após a apresentação dos primeiros seminários sobre “Modelos de Consciência”, de Anil Seth, e “Teorias Cognitivas da Consciência”, de Vincent de Gardelle; bem como algumas discussões com os participantes sobre esses temas. Após as Considerações Finais, acrescentei um Epílogo escrito alguns anos após o Workshop.

Palavras-chave: Consciência; cérebro; fisiologia; epistemologia; neurônios; astrócitos; modelos explicativos.

## Introduction

In 2009 the scientific journal *Nature* opened an interactive online platform, in which public or private discussion groups could be formed and conduct online live meetings, in which text messages could be exchanged. The material was saved for some time in the site, but it was deleted a few years later, when the journal decided to close the interactive platform.

The *Online Workshop on Theories of Consciousness* (OWTC), presided by the Swiss Medical Sociologist Dr. Hans Ricke, having the distinguished neuroscientist Dr. Arnold Trehub and myself as coordinators, and the helping hands of Drs. Chris Nunn and Bernard Baars in both the foreground and the background, took place as a private group in this platform.

The debate had begun a year before, in another public online forum in the same *Nature.com* site, created and administered by me: the *Brain Physiology, Consciousness, and Cognition* group, which became the most active forum in the entire site, with around one thousand members, and even appeared in the *News* section of the main *Nature* printed journal.

On the basis of the discussion carried out in this public forum, I accepted an invitation made by Drs. Ram Vimal and Chris Nunn to contribute, in collaboration with Hans Ricke, to the special edition of the *Journal of Consciousness Studies* (JCS) on the theme of *Defining Consciousness* (Pereira Jr. and Ricke, 2009). Building on this collaboration, the OWTC, of which Hans Ricke was the main organizer and moderator, was created, and continued its activities for more than half a year, producing nearly 1,000 pages of presentations and opinion exchanges. The contributors were leading consciousness

theorists.<sup>1</sup> A partial report, first authored by me (Pereira Jr. et al., 2010), was published in the JCS. A final consensus has not been reached during the workshop, and thus no other report has been published. A subgroup led by me continued the discussion in another private *Nature.com* forum, entitled *Consciousness Researchers Forum*, which produced a valuable book published by Cambridge University Press (Pereira Jr. and Lehmann, 2013).

A complete transcription of the OWTC was kindly made by neuroscientist Dr. Jonathan Edwards, and sent to the organizers, but it was never published. Having both these files and my own notes, I registered the guidelines I proposed to the group, together with epistemological issues that arose after the presentation of the first seminars, on “Models of Consciousness”, by distinguished cognitive scientist Anil Seth, and “Cognitive Theories of Consciousness”, by Vincent de Gardelle, building on his work in collaboration with Sid Kouider.

In the Epilogue below, I am also publishing the sketch of the first dialogue of a theater play that echoes the philosophical issues raised during the OWTC. Writing this piece was suggested to me by Hans Rieke after we met personally some years later, in 2016, in a scientific meeting at Lugano. I wrote only the opening dialogue of the intended play, displaying the conceptual framework of the Theory of Consciousness (called *Triple Aspect Monism*) developed after the end of the OWTC.

---

<sup>1</sup> In alphabetical order: Alfredo Pereira Jr., Alice Kim, Anil Seth, Arnold Trehub, Axel Cleeremans, Bernard Baars, Biyu He, Bjorn Merker, Chris Nunn, David Rosenthal, Dietrich Lehmann, Hans Rieke, Jonathan Edwards, Jean-Pierre Changeux, Lucia Melloni, Max Velmans, Ned Block, Sid Kouider, Stanislas Dehaene, Stuart Hameroff, Sue Pockett, and Vincent de Gardelle

## The Guidelines

In the beginning of the OWTC, I wrote the text below.

“Before discussing a theory of consciousness, it may be convenient to clarify what it addresses. There are at least three classes of theories:

- a) Those that address the *vehicle or system* that supports conscious processing, e.g. brain circuits and mechanisms;
- b) Those that address *properties of* (intersubjectively shared) *conscious contents*, e.g. if they are representations and if so, what kind of representation they are, and
- c) Those that address subjective *aspects* of conscious experiences, the singular “*first-person*” experiences that accompany the instantiation of intersubjectively shared contents.

Based on this distinction (or on a correction or improvement of it), we can make a collective effort to establish an agreed taxonomy of published theories in terms of families of theories that address similar aspects of consciousness.

We may also evaluate the scientific status of theories according to current standards. Assuming that there is a common understanding of what is a scientific measurement, we can divide these theories into four classes:

- a) Those that afford the making of direct measurements (according to current standards) that would gather evidence for or against the claims of the theory;
- b) Those that assume that consciousness is a kind of second-order phenomenon (like entropy) that cannot be directly measured but has values that can be calculated from data obtained by means of direct measurements. These theories

bring with them a set of theoretical assumptions about how to make the calculation, often related to concepts of information and complexity; different conceptions are related to different forms of calculation;

c) Those that would afford a direct or indirect measurement with new techniques, to be invented or improved from already existing technologies; and

d) Those that imply that consciousness (or conscious states and processes) cannot in principle be measured. These theories could be considered mainly philosophical (or metaphysical)”.

I also added the following philosophical remarks: “Explanatory problems faced by consciousness models are not limited to the explanation of phenomenal experiences. On the one hand, considering the complexity of brain activity, it is difficult to define any mechanisms responsible for any cognitive (learning, memory formation, attention) or emotional function. On the other hand, describing properties of consciousness (phenomenal or not) would require a powerful tool, such as the kind of state space that Chris Nunn proposes (Nunn, 2007) or the conceptual space developed by Peter Gardenförs [OBS.: later reviewed in Pereira Jr. and Almada, 2011].

I have agreed with Bernard that the discussion should focus on empirical questions in order to achieve remarkable progress. Here in OWTC we will have an opportunity to discuss in detail important empirical findings with their authors. I will only mention two apparent conflicts as examples of the difficulty of adding the results to find a synthesis:

a) Using similar experimental paradigms, some found that conscious visual perception is supported by 100 ms responses, while others found that a 300 ms ERP is responsible for triggering the (gamma) synchrony that supports it;

b) Most neuroscientists have believed for decades that consciousness is correlated with an increase of firing rates and sustained neuronal excitation, but Tononi's theory predicts that silent neurons have an equal participation in the determination of the qualitative conscious state."

The discussion in the seminars focused on these themes, often bringing new epistemological proposals without achieving an agreement in the end.

## **Discussion with Anil Seth**

After Anil Seth's first seminar on "Models of Consciousness", based on his Scholarpedia entry (Seth, 2007), the following discussion took place. He proposed the definition: "The discussion article reviews various 'models' of consciousness. A model of consciousness is a theoretical description that relates brain properties of consciousness (e.g., fast irregular electrical activity, widespread brain activation) to phenomenal properties of consciousness (e.g., qualia, a first-person perspective, the unity of a conscious scene) through some type of mechanism. One can begin by questioning this assumption and, more generally, by asking about the utility of models in consciousness science. Are there any other useful conceptions of scientific modeling that could apply to consciousness? What kind of new understanding can a model of consciousness generate? Do models need to be predictive in order to be useful? What determines – or will determine – the success of any candidate model, where 'success' is construed very broadly?"

On the basis of the proposed guidelines, I asked: "I cannot identify in most (if not all) reviewed theories:

- a) What are the ‘brain properties’?
- b) What are the ‘phenomenal properties’ of consciousness?

Most theoretical proposals seem to try to capture *cognitive functions* of consciousness, instead of providing a detailed description of brain mechanisms and respective properties of consciousness... This is a problem, because there is an ambiguity with the term “function”. A function may be defined for *the system that supports consciousness* (e.g. integration of neuronal activity as a function carried by synchrony) and can also be defined for *the content* (e.g. conscious attention as a selective function). Both kinds of function do not necessarily overlap”.

Anil also wrote: “The IITC proposes a novel measure of the ‘quantity’ of consciousness generated by a system. This measure, *phi*, is defined as the amount of causally effective information that can be integrated across the weakest link of a system (Tononi and Sporns 2003; Tononi 2004).”

I made some questions, answered by Anil:

(Alfredo) Is this Shannon’s (and Weaver’s) concept of information, a modified conception (as in Dretske) or a totally different one?

(Anil) Information theory in Tononi’s sense is not used in the S-W concept of information *transmission*, but rather as a way of characterizing (multidimensional) variance and shared variance, i.e. in a statistical sense. But the underlying concept of *entropy* is shared.

(Alfredo) Is there a (mathematical) relation between this concept of complexity (*phi*) and Chaitin’s?

(Anil) An important difference is that Kolmogorov-Chaitin proposed an ‘algorithmic’ notion of complexity, which is



maximized for random sequences. The premise of *phi*, neural complexity, and causal density is that complexity should be maximized for systems intermediate between complete order and randomness. According to the IITC, consciousness, as measured by *phi*, is characterized as a “disposition” or “potentiality”. The contents of any given conscious scene are specified by the value, at any given time, of the variables mediating informational interactions within the system. According to Tononi, ‘A distinguishing feature of the IITC is that *phi* is proposed to be a sufficient condition for consciousness, so that any system that has a sufficiently high *phi* – whether biological or non-biological – would be conscious’ (Tononi 2004).”

(Alfredo) Is there a (well-defined) relation between *phi* and negentropy? What is the relation (if any) of causal density and the flux of Shannonian information?

(Anil) Lumping these together, *phi* is defined using entropy as a central concept. Relations to negentropy (or equivalent to negentropy) have not been worked out as far as I know. Intuitively, causal density and *phi* and neural complexity are all closely linked. In my lab, we are currently working out how to express causal density in terms of information theory, to allow formal equivalences to be drawn. This can have a practical benefit, since autoregressive models can be easier to derive from data than multivariate entropies.

At the end of the discussion, Anil made these important remarks: “It is worth noting that some of Tononi’s ideas challenge this assumption (that conscious activity depends on brain activity – APJ). In at least one version of his IITC, the possibility is explicitly noted that a brain with no activity at all could nonetheless be conscious, perhaps of nothing! The logic behind this implication is that, in the IITC, the level of

consciousness (in contrast to content) is a function of the *potential* of the brain to enter this-that-or-the-other state. And one can have potential in buckets, without any activity. Tononi's more recent - 2008 - exegesis is perhaps a little more circumspect on this issue, since his quantity *phi* is now determined by the information generated when a system moves from one state to another. While much information could be generated by a system moving to a completely inactive state, considering this (inactive) state as persisting (in a statistically stationary sense) would lead to low *phi* and a zero conscious ness level...For the record, I myself cleave to the idea that high levels of *phi* (or whatever current measure) may be necessary but not sufficient for non-zero conscious ness levels. Whether it is possible to be conscious without particular conscious contents is a provocative question, but one which I think is ill-posed. If indeed we can be conscious without content, such a content-free state then can be said to *have* content precisely in virtue of this property, on the view that content is what distinguishes one conscious scene from another”.

## **Discussion with Bernard Baars**

Bernard commented about Anil's presentation: “‘Yes or no’ decisions are possible in some cases, as extensive lesion studies. If the cerebellum of a human subject is extensively damaged and the subject does not report any change in conscious experiences, it is possible to conclude that the cerebellum does not directly support consciousness, regardless of any particular theory. The situation becomes exponentially more complex when the conclusion is to be inferred from data obtained with

registering or measuring techniques/technologies. I will focus on fMRI. Logothetis and his group demonstrated that Bold fMRI measures correlate well with Local Field Potentials (registered with intercellular microelectrodes) reflecting dendritic electric activities. However, there is an unnoticed theoretical problem with this correlation: as inhibitory transmission causes hyperpolarization of the membrane (physically speaking, an electric field of greater magnitude than the depolarization caused by excitatory transmission), why don't inhibited brain regions light up in fMRI?"

I replied that: "The question of why BOLD fMRI (after subtraction of baseline/unconscious activity) correlates so well with conscious processing, in spite of all the limitations of this imaging technology, has bothered me for more than 10 years. In the beginning I was a skeptic, but I became converted by the evidence. Only recently did I find a convincing reason. The clue came from astrocyte research. First, it has been well established that astrocytes, not neurons, exert direct vascular control (see e.g. the work of Haydon and Carmignoto, 2006). Second, it was shown – in a revolutionary paper by Schummers et al. (2008) – that BOLD fMRI does reflect astrocyte activity. Since then, my favorite hypothesis became the idea that astrocyte global physiological activity (amplitude-modulated calcium waves) is the analog of feeling patterns. Astrocyte coherent calcium waves are triggered only by neuronal excitatory (release of Glutamate), not inhibitory (release of GABA) activity in tripartite – i.e. neuro-astroglial – connections. Inhibitory transmission generates neuronal electric fields, but these fields do not impact BOLD fMRI because blood flow is controlled by astrocytes, and these cells are not activated by a GABA release in the tripartite synapse intercellular space".

Bernard also commented: “Another important study by Fox et al. (2005) shows that the activation of a brain area is accompanied by the deactivation of others, and most of the pairs they mention correspond to the pairs you mentioned in your post (ventral x dorsal; cerebellum x cortical areas). It seems to me that as the deactivation of one region is correlated with the activation of the other, both belong to the same process, and we therefore cannot infer that only the active one is contributing to consciousness. How would the inactive one contribute? Maybe by increasing a signal-to-noise ratio in a global pattern of activity of the whole brain (this seems to be the solution found by Tononi and Koch)”. This issue was addressed by Biyu He, a young brilliant neuroscientist indicated to the OWTC by Stanislas Dehaene, as reviewed in the next section.

## **Discussion with Biyu He**

Biyu wrote: “I have a few half-baked thoughts on some of the physiological questions that have been under discussion on the involvement of the default network in consciousness. First of all, the pair of task-positive and task-negative regions is not really the ‘ventral x dorsal’ or ‘cerebellum x cortical’ areas. There was one specific pair that was found in the original Fox et al. (2005) paper: dorsal attention network (including FEF, IPS, and MT) and the default network (including BA 39, PCC, MPF, superior frontal regions, and parahippocampal gyrus). Later, it was shown that the executive control network (including DLPFC, dorsal parietal regions (more like BA 40) may also be anti-correlated with the default network. The word “default mode” has been somewhat confusing because there

are many other networks (total of about 8 well described) that are spontaneously coherent in the resting state, and the default network, including the regions above, is only one of them. But the default network is special in the sense that it is more involved in a self-referential type of thought, as Bernard has mentioned. A conjecture on the default network's involvement in consciousness was offered in He and Raichle (2009): Not all brain networks contribute to consciousness equally...We speculate that the anterior cingulate and anterior insular cortices, in addition to the default network, might be more pivotal than the sensory and motor networks, and maybe even the dorsal attention network (including the dorsal visual stream and frontal eye field) in the emergence of consciousness. This conjecture mainly comes from a thought experiment comparing the largely unconscious state – slow-wave sleep (SWS), with the conscious states including wakefulness and rapid-eye-movement (REM) sleep. Whereas the sensory and motor regions and the dorsal attention network are as active in SWS as in wakefulness, the anterior cingulate, anterior insular, and midline regions of the default network are deactivated in SWS and reactivated in both REM sleep and wakefulness. To the best of our knowledge, this conjecture is also consistent with existing data from persistent vegetative patients, blindsight patients, and from manipulations of momentary conscious perception”.

She added: “On the relation between inhibitory neuron's activity and the fMRI BOLD signal, I think Alfredo again raised a very important question. Here I think (speculatively) that part of the confusion in the literature might come from equating the fMRI BOLD signal with energy consumption. It is well established that the inhibitory neuron's activity is accompanied

by cost in glucose metabolism and changes in blood flow (Buzsaki, Kaila and Raichle, 2007; Logothetis, 2008), but the fMRI BOLD signal reflects the extra oxygen delivered over and above that consumed, and is specifically linked to glycolysis, not necessarily the total amount of glucose metabolism (a very nice review is provided by Raichle and Mintun, 2006). I agree with Alfredo that the neurovascular coupling mechanisms might differ between glutamate and GABA signaling, including their different involvement of the astrocytes. One intriguing possibility is that GABA signaling uses, proportionally, more oxidative phosphorylation and less glycolysis than glutamate signaling, but this is a total speculation, and I don't know of any existing data that speaks to this. Lastly, on why the fMRI signal, despite being classically considered the "remotest" form of neuronal signaling, correlates so well with conscious perception, and even better than spiking activity in some instances, especially in V1. My own view on this issue is explained in He and Raichle (2009, p. 305-306): fMRI experiments and single-unit recordings often show discordant results during manipulations of consciousness; this disagreement has been most dramatic in V1. These puzzling results are at least partially illuminated when we bring the SCP (Slow Cortical Potential – APJ) and its underlying physiology into the picture...The top-down effect of attention in early sensory cortex is largely invisible to spike recordings, but it is readily seen in the fMRI signal. Consistent with a close correspondence between the SCP and the fMRI signal, as argued here, top-down influence in V1 can be seen with measurements of the SCP, using either optical imaging or field potential recordings...Using a visual illusion task and depth recording in V1,...perceptual suppression was only associated

with changes in the lowest frequencies in upper cortical layers when the Current Source Density (CSD) method (which has much better localizing power than raw field potentials...) was used. Similar to the SCP, the fMRI signal also tracks perceptual changes, whereas spiking activity was unaffected.”

I replied: “Dear Biyu: Many thanks for your highly clarifying and explanatory speculations...Inhibition has two phases that may have a different impact on fMRI and other measuring technologies: the increase in inhibitory transmission, which is obviously related to the excitation of inhibitory neurons, and the result of inhibitory transmission in the target dendrite (membrane hyperpolarization), not to mention another phenomenon: membrane depression. How do neurobiological theories of consciousness account for them? Are some of these states co-determinant of conscious contents, or do only excitatory activities and/or spikes count?” To which she answered: “In general, I think the activated neuronal groups contribute to the content of conscious awareness. However, the quality of such conscious awareness depends on the life experience of the entire system, which is stored in the fine detailed anatomical and functional architecture of the system (and limited by the genetic blueprint). You might ask what I mean by “functional architecture”, which to me is the spatiotemporal structures present in the spontaneous brain activity, which depend on the anatomical structure but have many more “layers” or “dimensions” to it. In this sense, the silent neuronal groups, by their mere existence, may also contribute to the quality of conscious awareness. And I think that is the exact point that Giulio Tononi is driving at when he keeps talking about the photodiode (vs. thermostat)”.

Anil replied to Biyu: “I also welcome the introduction of default networks as a topic, for here one can envisage another use for models as a kind of ‘negative’ proof. As an example, a recent study by Deco et al. (2009) shows how anti-correlated slow fluctuations can arise simply from underlying physiological constraints rather than serving a specific function. Though this does not exclude a positive role for default networks (and indeed the story is more complicated – see the paper!) the model does serve to sharpen the question...Another paper that might be worth looking at is that of Shulman et al. (2009) which attempts to account for the conscious state in terms of physiology *without* appeal to cognitive/functional models of any kind”.

## **Discussion with Vincent de Gardelle**

In his seminar, Vincent presented three types of cognitive theories of consciousness:

- “1) Theories that emphasize on the architecture of the cognitive system that leads to conscious experience. Here come the Global Workspace Theory (Baars), the Intermediate Level Theory/Attended Intermediate Representations Theory (Jackendoff and Prinz), the Information Integration Theory (Tononi). In these accounts, consciousness occurs in a system that has certain architectural characteristics: global broadcasting, intermediate object representations, information integration;
- 2) Theories that highlight that some features of consciousness are illusory. Dennett’s major point in the Multiple Drafts Model is the rejection of a single narrative stream. Wegner points to the possibility that the causal power of consciousness might be illusory;
- 3) Theories defending that consciousness is the result of



learning in the cognitive system. In the Sensory-Motor Theory (O'Regan and Noë) the subject learns to master sensory-motor contingencies and awareness follows from this expertise. In the Radical Plasticity Thesis (Cleeremans) and in the Higher Order Bayesian Perspective (Lau), higher order representations learn about first order representations”.

He also remarked that the theories may be combined: “For instance, an attempt to reconcile the Higher Order Theories with the Workspace Model can be found in the proposition of ‘Higher Order Global States’ by Robert van Gulick. So, in my opinion, to provide a model of consciousness, one agenda could be to agree on basic properties of the cognitive system (e.g. functional specialization, hierarchical models of perception with different levels of processing, the definition of ‘information’ in the brain, etc.); to envision the different accounts as focusing on different parts of the phenomenon, rather than as exclusive theories; to try to merge the different accounts, maybe by relying on a common ground. As a common basis I would suggest considering the workspace theory, which I find to be the most precise account (to my current and limited knowledge)”.

I replied with a long message: “(In the paper referred in the seminar) You classify the theories in three groups:

1) Those that consider consciousness “as result of specific architectural elements within the cognitive system”. This statement seems to imply that this group of theories refers to *properties of the system or vehicle* that supports consciousness, but they are in fact about *properties of the content or the dynamics of the content*. GWT is essentially about content selection and broadcasting, although in several papers and books his author has covered all aspects of consciousness phenomena. Jackendoff and Prinz’

theories are explicitly about representations, focusing on the grammar of conscious contents. Tononi's theory addresses both the content (the process of integration of information) and possible computational mechanisms operating in the brain;

2) Those that consider consciousness as illusory. These theories, of course, do not explain consciousness, but "explain it away" as somebody once wrote;

3) Those that relate consciousness with learning. There are three different approaches (Sensory-Motor, Radical Plasticity and Bayesian Signal Detecting hypotheses) addressing both the content (e.g. it is non-representational; self-consciousness is meta-representational) and aspects of the supporting system (e.g. interaction with the world by means of sensory-motor operations, operating with Hebbian mechanisms).

I found the three proposals complementary to GWT, emphasizing particular computational or non-computational processes that pose constraints on any causal/neurobiological model. However, these theories do not seem to address any specific brain mechanisms or circuits, not even the default candidates (respectively: motor system and corollary discharges; membrane potentiation – as in LTP studies – and associative cortex – mostly PF – functions)".

Vincent replied: "One important consideration when elaborating models is to focus on the relevant level: is a description of ionic channels relevant for consciousness? Where should we put the link between biology and phenomenology? I would say that the level of cognitive functions is the best candidate...It might be that one major difficulty for most people is to be convinced by an explanation of conscious experience in any biological terms. To sketch out, we continue to see an

“explanatory gap” between these two perspectives, one being subjective, qualitative, human-like, the other being objective, quantitative, information-based and computer-like. Should we say that the good old “hard problem” is just a hidden form of dualism that plagues consciousness research? Should we accept without further discussion that conscious experience can be reduced to the relevant level of information processing? I would say yes.”

Then I replied, again with a long message: “The title of the paper is appropriate: you discuss *cognitive* theories of consciousness. The question is: Are these theories really explanatory? According to current standards in natural sciences, explanation is based on causation. If consciousness science follows these standards, electromagnetic patterns produced by the opening of membrane ion channels and related phenomena are the best candidates for the ‘explanans’. Knowing them, we could not only understand the phenomenon but also predict, control, modify, or artificially reproduce it. If consciousness science does not follow this standard, what do the theories afford? “Information processing” as well as “function” are ambiguous terms, since they can refer to brain activity or to cognitive/conscious activity. Both do not necessarily overlap. Psychology is not supernatural or anti-natural. I would agree that the kind of causation that operates in the domain of Human sciences is different from mechanical physical causation. Probably the relation between brain and conscious states/processes is not of this kind, but to deny any causal connection is a problematic Dualist assumption. Anyway, I do not see the *cognitive* theories of consciousness as Dualist, I am more inclined to consider them as “Magical Functionalist”, in the sense that they look for

*intermediate functions* (mostly computational functions) that create *an illusion of explanation* without really addressing the fundamental issue of the relation of brain/body and conscious states/processes. Epistemologically, we can think of five possible kinds of connection between B (brain activity) and C (conscious activity):

- 1) They are equal;
- 2) C reduces to (is deduced by) B – or vice-versa;
- 3) There is a vague, not quantified correlation between both, as assumed in the NCC approach;
- 4) There is a quantified and statistically relevant correlation between both, as in Wesley Salmon’s classical example: “Smoking cause cancer” (the cause is neither necessary nor sufficient to cause the effect, but increases the probability of occurrence of the effect above the “randomness” threshold); this kind of meaningful correlation affords the making of useful analogies;
- 5) They are “separate” (D. Marr ‘apud’ A. Trehub), statistically independent, or – in Metaphysics – we could say that Substance or Property Dualism characterizes such a relation...

To make a long story short, the crux of the problem is how to treat correlations between brain activities measured with the use of some technological tool (fMRI, etc) and conscious activities/experiences reported by a (human) subject. For instance, just one run of a neurocognitive experiment (e.g. presentation of a visual stimulus and report of the conscious content elicited by the stimulus) does not afford the usage of statistics. How to measure the probability of a brain state and the probability of a conscious state, and then determine the conjoint probability? How many runs with one subject are sufficient,

or what is the required size of a population for a single trial experiment? I have suggested that this task requires sophisticated methodologies, as Physicists have developed in Statistical Mechanics since Ludwig Boltzmann's work in the 19th century. In the Philosophy of Neuroscience and Consciousness Studies, we are still far away from their standards. Experimentally, we cannot prove neither the Identity of brain and conscious states, nor the Dualist views of separation. All we can prove is if their correlation is casual (accidental) or causal (in Salmon's sense = statistically significant). This proof is a very complex one that we are beginning to envisage. In the previous stage of our inquiry (search for the NCC) the issue of measuring the degree of correlation was not on the table. Today – possibly after the results obtained with the “mind reading” paradigm – the controversy has already been raised, and I suggest that we could begin our discussion at this point, instead of returning to metaphysical issues”.

Vincent at this point made a detailed final reply: “Dear Alfredo Thanks for your comment, let me go through it:

1) You say that the ‘architectural theories’ are theories about the properties of the contents. Yes, I agree, but they do more than only that, I would also say that they are theories of the vehicle (‘vehicle’ in the cognitive sense rather than in the biological sense). Indeed, these cognitive theories offer (arguably more than other accounts) a quite precise description of a whole cognitive system which includes different types of units and mechanisms (e.g. the particular role of contexts in the GW). In this complex system, some contents are conscious and others are not. These theories provide propositions for these differences. Of course, this could also be said for other theories: as soon as there is a claim like ‘consciousness is that/conscious contents

are such' there is also an implicit accompanying claim for unconsciousness/unconscious contents. The crucial thing here is that these theories propose the conscious vs. unconscious difference is to be found in the way the cognitive system uses/works the particular content (with vs. without broadcasting/attention/integration). So, my word would be: both content and vehicle,

2) Let me bring some precision here. Neither Dennett nor Wegner say that there is no consciousness. They just remind us that some apparent properties of consciousness (unitary stream of consciousness, causal impact of conscious will) might not be taken as such. In particular, Dennett claims that his theory (*Pandemonium* or *Multiple Drafts Model*) is a theory of consciousness, and that the only problem is to recognize when the explanation is done. You say "These theories of course do not explain consciousness, but 'explain it away' as somebody already wrote". Well, on that issue you seem to disagree with Dennett, while personally I think I would agree with him. This is a very important discussion that won't be settled here. I think we could discuss this in another thread, and importantly rather than having two contradictory claims that are obvious to each of us, could we design a real empirical test?

3) I totally agree that most theories presented here have no particular claim for neural basis. One exception is the GW, for which one correlate (the ERTAS system) has been proposed early on, because of its global connectivity pattern, which was satisfying the constraints set at the cognitive level (broadcasting again). Theories that have been more focused on neural basis are presented by Sid (Kouider – APJ) in the "Neurobiological Theories" chapter, rather than in the present "Cognitive Theories" chapter".

## Discussion on the Conscious Feeling Subject

At some point of the discussion I remarked: “We are finally entering the swampy issue of ‘who is the conscious subject?’. Arnold’s model is clear about who he thinks is the subject (he called it “the I!”) and is compatible with a scientific approach. Damásio’s books indicate another position (that I prefer), but he did not present a scientific model (instead of proposing an experimentally testable model, he calls old Spinoza to support his views). In the Pereira and Ricke (2009) JCS paper, we proposed that the conscious subject is the living individual.

Let me state my disagreement with the popular view that *mapping or representing* events (in physical or phenomenal space) is the fundamental feature of consciousness...This view does not take into account the really fundamental part of consciousness, namely *conscious sensations and feelings* like pain, thirst, fear, pleasure, anger, etc. Mapping or representing a feeling as located in some part of the body map/representation is not the same as describing or explaining the feeling, for two reasons: first, spatial (or temporal) location does not indicate the quality of the phenomenon; and second, maps or representations do not have feelings (it is the individual/body who has them). How could feelings be explained? As in other areas of biological sciences, they can be explained by remote and immediate causes: a) Remote. In the case of pain, there is not much controversy regarding its evolutionary importance for individual survival, as a signal of danger (however, it should be made clear that pain does not *represent* a tissue being damaged or the object that is causing the damage; it is just an icon of danger); b) Immediate. The explanation of pain by means of immediate causes is very

simple, but most theories of consciousness cannot afford it! A nociceptive signal coming from some part of the body goes to two brain circuits, one (somatosensory cortex) where the part of the body that is damaged is detected, and another (e.g. insula) where the feeling is qualitatively generated and then ‘projected’ [according to Max Velmans’ theory, discussed in another seminar - APJ]” back to the place of damage (possibly by means of an integration of somatosensory and insular activities). What occurs in the insula or other region responsible for the qualitative aspect of feelings? Of course, it is not a matter of neural connectivity or synaptic weight only. *What and where* is the activity that is *analog* to the form of the pain? How are different kinds and degrees of toothache determined? It seems to me that the theories of the abovementioned authors lack explanatory tools to answer these questions. Jonathan Edwards’ theory has the right kind of conceptual framework for this task, but places the right mechanism in the wrong place (single cells – APJ)”.

Hans Ricke, my co-author in the 2009 JCS paper, commented: “Dear Alfredo, thank you for bringing up this point. Obviously, the qualities of experience are not addressed by a representational theory that is mainly about location. No one would argue against the importance of that approach. It is just lacking too much! It is the qualities of experiences that are relevant for the “what is it like” as well as for the “does the experience have a consequence”, “does an action follow or not”. I wonder if the argument must go further and include recognition. Local kinds of representation are possibly not required for recognition. This is more applicable to the “inner reality”; obviously where people do not ponder visually or otherwise locally oriented, but



human beings can imagine things like – sorry if I repeat myself – the holocaust, which is a concept that does not seem to fit into the retinoid system and representational concepts of that kind. These are the real concerns for Consciousness Science in my opinion. The representational concept reaches as far as life forms that have already existed hundreds of millions of years before, and we must strive for a theory that is up to date!”

Jonathan Edwards commented: “I obviously agree entirely with Alfredo here, except in that I place experience in the right place!! There is a correspondence between local biophysical dynamics and an experiential integration that can make use of elements that encode things like fear as well as red, half meter, behind, and now. The integration giving the sense of pain in the toe must occur somewhere where signals from insula and somatosensory converge, otherwise you have a floating pain and a neutral toe in different places. I am quite happy to allow floating pain and neutral toe experiences to occur in the respective areas *as well*, but they are not the experiences of customary consciousness”.

I tried to find more agreement: “The qualities are recognized by single cells (neuron membrane and/or astrocyte waves... Locating the qualities would involve an integration of somatosensory and (e.g.) insular activities – maybe with an involvement of the retinoid system (Arnold Trehub’s theory – APJ) to frame things in a triplet of Cartesian coordinates. The experiences involve a broad domain of interaction between the quality instantiated in some cells, the impact of this event on other parts of the brain, and the whole brain interacting with the body and environment”.

## Concluding Remarks

From the OWTC transcription made by Jonathan Edwards and my own notes, I selected the initial discussion topics that influenced my posterior work on the Theory of Consciousness, and my own comments made in the workshop that indicate the pathway I would follow for more than 12 years in several publications. This material reveals a richness of ideas and different approaches still valuable to read and consider 15 years after the discussion took place.

Regarding the intended agreement between the participants, it is clear now that in spite of the qualification of the contributors, we were not philosophically prepared to reinterpret the neurosciences according to the demands of a Theory of Consciousness. The question remains if the intended Theory of Consciousness is philosophical, scientific or a mix of both – as the whole field of Philosophy of Science tends to be, with the corollary of not being well accepted by both philosophers and scientists. Fortunately, there are a few philosophers who are fond of neuroscience, and neuroscientists who are fond of philosophy working hard to bridge this gap.

With regard to the other, deeper gap, the explanatory one, referring to the distance between neural and mental processes, from my point of view was recently reduced with the new concept of Sentience, and its explanatory role bridging life and consciousness (see also Pereira Jr., 2021).

## Epilogue

The “Consciousness Romance” was conceived by me in 2016 as a theater play with three characters, corresponding to

the philosophical theory of *Triple Aspect Monism* (Pereira Jr., 2013), with the intention of bringing this theory to a wider audience. One of the characters (*Natura*), a physicist, represents the material aspect of nature, or “Mother Nature” in the popular image; a second character (*Theo*), a philosopher, represents the ideal aspect of reality, understood as co-existing with the material one, in the context of Aristotle’s *hylomorphism*, or a transcendent *God*, a “Father” in most religions; and the third character (*Pragma*), an androgynous physician, represents the practical ways of feeling and solving problems that emerge from the conflict of the first two characters. The beginning of the play includes the following dialogue:

Pragma - I have not read all these books on consciousness, but I have a basic intuition. Consciousness is a game that requires at least two players, the being that is conscious of an experience and the content that is experienced!

Theo - Yes, your intuition is equivalent to a complex philosophical theory raised by a German philosopher named Edmund Husserl. He wrote that consciousness requires two halves, like the Yin and Yang of ancient Chinese philosophy.

Natura - You may find the conscious being the most mysterious player, but I am more familiar with explanations of things. In physics, we build models of elementary particles and waves that allow us to explain everything that can be observed.

Theo – You’ve got a lot of results about the material world, but what did you find about consciousness?

Natura - If a physicist tries to explain consciousness, she implicitly assumes that consciousness is an aspect of the physical world; otherwise, the explanation would be beyond her reach.

Pragma - In medical sciences, we do not try to explain consciousness, but rather to control it, mostly because we should eliminate or minimize our patient's pains.

Theo - For both of you, the conscious mind and the living body are actually just one reality, right? For the majority of people, however, one is independent of the other, because they believe that the conscious mind (they call it "the soul") can survive the death of the body.

Pragma - Is this the famous Dualist philosophy advanced by Descartes?

Natura - As far as I know the Cartesian approach, the "thinking substance" he mentions is not the soul believed by religious people like Theo.

Theo - You're right, Natura. Descartes' philosophy proposes that the mind and the body are distinct entities that communicate by means of the pineal gland. He probably learned about this gland from esoteric writers of his time.

Natura - I do not agree with an explanation of a phenomenon by means of the invention of an entity that we cannot study scientifically. If I have the right of a choice, I would investigate how a natural phenomenon doubles itself, becoming both the observer and the observed.

Pragma - Natura, your suggestion is intriguing, but to discuss this issue we would need to focus on the whole evolution of the universe, because the existence of these observers seem to be a very late phenomenon, which is known to exist only on our planet!

Theo - There is another possibility: the seeds that grow to the point of becoming conscious people like us may have

been present since the beginning of time, or better, if time has no beginning or end, the seed may be eternal. Conscious beings may appear at every place and time, when and where adequate conditions are fulfilled!

Natura - Good idea! I am happy to agree with you, but I do not think that it implies a supernatural God!

Theo - This is a very subtle issue, Natura. God may be somehow hidden in nature and disclosed only to the human mind, as proposed by another philosopher, Ludwig Feuerbach.

Natura - Yes, but...Even if there is no God at all, there may be a seed of consciousness everywhere; the seed can grow, if local conditions are adequate, or it may remain in a latent state if the environment does not allow it to grow.

Pragma - If you agree on this possibility, the problem is to find what are these conditions. For instance, do we need living cells for the existence of consciousness? Do we need neurons? Or a complete nervous system? Or a brain?

Theo - Good questions, Pragma! However, I'm afraid that this kind of investigation, although interesting to you, is limited if we do not understand what the seed that grows to become consciousness is. Practically all religions assume that this seed is supernatural. I have my own ideas in this regard: just one entity is not enough. To explain the complexity of the world we consciously experience, a plurality of eternal forces and their interactions in time is required.

Natura - Please explain to me why these forces are assumed to be supernatural. Why don't they assume a plurality of physical principles or rules that interact to produce the

complexity of the phenomena we experience?

Theo - To be honest, I don't have an answer to this question, Natura. I just remember that this is the dominant belief at which humanity arrived after thousands of years of evolution on this planet. It cannot be completely wrong – especially since it has been useful to establish social bonds. This belief helps people to build social relations and institutions.

Pragma - I have a challenge for you. We could collectively investigate the conditions that existed on earth for this hypothetical seed to grow, and after an evaluation of these conditions, we could decide if the seed has to contain supernatural powers.

Natura - If you really intend to go on with this kind of project, I can invite some of my university colleagues to give us interviews on their expertise, so that at the end we can figure out the ultimate status of this hypothetical seed!

## References

BUSZÁKI G., KAILA K. and RAICHLE, M. (2007) Inhibition and brain work. **Neuron** 56(5):771-83.

DECO, G., JIRSA, V., McINTOSH, A.R., SPORNS, O., and KOTTER R. (2009) Key role of coupling, delay, and noise in resting brain fluctuations. **Proc Natl Acad Sci USA**. 106(25):10302-7.

FOX, M.D., SNYDER, A., VINCENT, J., CORBETTA, M., VAN ESSEN, D.C. and RAICHLE, M.E. (2005) The human brain is intrinsically organized into dynamic, anticorrelated functional networks. **Proc Natl Acad Sci USA**. 102(27):9673-8.

HAYDON, P.G. and CARMIGNOTO, G. (2006) Astrocyte control of synaptic transmission and neurovascular coupling. **Physiol. Rev.** 86, 1009–1031.

HE B.J. and RAICHLE M.E. (2009) Response to Koch: Elaborations on the SCP hypothesis. **Trends Cogn Sci.** 13(9):368-369.

LOGOTHETIS, N.K. (2008) What we can do and what we cannot do with fMRI. **Nature.** 453(7197):869-78.

NUNN, C. (2007) **From Neurons to Notions: Brains, Mind and Meaning.** Edinburgh: Floris Books.

PEREIRA Jr., A. (2013) Triple-aspect monism: A conceptual framework for the science of human consciousness, in Pereira Jr., A. & Lehmann, D. (eds.) **The Unity of Mind, Brain and World: Current Perspectives on a Science of Consciousness,** Cambridge: Cambridge University Press.

PEREIRA Jr., A. (2014) Triple-aspect monism: Physiological, mental unconscious and conscious aspects of brain activity, **Journal of Integrative Neuroscience,** 13, pp. 201–227.

PEREIRA Jr., A. (2021) The Role of Sentience in the Theory of Consciousness and Medical Practice. **Jnl Cons. St.** 28 (7-8), 22-50.

PEREIRA Jr., A. & ALMADA, L.F. (2011) Conceptual Spaces and Consciousness Research, **International Journal of Machine Consciousness** 3, pp. 1–17.

PEREIRA Jr., A. & LEHMANN, D. (2013) **The Unity of Mind, Brain and World: Current Perspectives on a Science of Consciousness.** Cambridge: Cambridge University Press.

PEREIRA Jr., A. & RICKE, H. (2009) What is consciousness? Towards a preliminary definition, **Journal of Consciousness Studies,** 16 (5), pp. 28–45.

PEREIRA Jr., A., EDWARDS, J.C.W., LEHMANN, D., NUNN, C., TREHUB, A. & VELMANS, M. (2010) Understanding

consciousness: A collaborative attempt to elucidate contemporary theories, **Journal of Consciousness Studies**, **17** (5–6), pp. 213–219.

RAICHLE, M.E. and MINTUN, M.A. (2006) Brain work and brain imaging. **Annu Rev Neurosci**. 29:449-76.

SCHUMMERS, J., YU, H., SUR, M. (2008) Tuned responses of astrocytes and their influence on hemodynamic signals in the visual cortex. **Science** **320**, 1638– 1643.

SETH, A. (2007) Models of Consciousness, **Scholarpedia** **2** (1): 1328. Available at: [http://www.scholarpedia.org/article/Models\\_of\\_consciousness](http://www.scholarpedia.org/article/Models_of_consciousness)

SHULMAN, R.G., HYDER, F., and ROTHMAN, D.L. (2009) Baseline brain energy supports the state of consciousness. **Proc Natl Acad Sci USA**. 106(27):11096-101.

TONONI, G. (2004) An information integration theory of consciousness. **BMC Neurosci** **5**, 42.

TONONI G. & SPORNS, O. (2003) Measuring information integration. **BMC Neurosci** **4**, 31.