Cognitive Critique

CONSCIOUSNESS REVEALED?

ANDREW C. PAPANICOLAOU

Division of Clinical Neurosciences
Department of Pediatrics
University of Tennessee College of Medicine
Memphis, Tennessee

EMAIL: apapanic@uthsc.edu

Accepted March 15, 2012

KEYWORDS

cognition, functional neuroimaging, consciousness, brain activation, vegetative state, coma, MEG, fMRI

ABSTRACT

The purpose of this essay is to address and reconcile the conflicting messages deriving from the functional neuroimaging literature regarding whether the brain mechanism of consciousness and the neuronal correlates of concepts and of transient experiences are visualizable, and to what extent they have been visualized. It is argued, first, that the likelihood of visualization of different aspects of mentation can be deduced unambiguously from the fundamental principles and facts that comprise the formal structure of the functional neuroimaging methods. Second, it is shown that to the degree that such aspects do have neurological validity and the formal structure of the methods holds true, the likelihood of visualizing their neuronal correlates varies from extremely high for all psychological functions, including consciousness, to practically null for conscious experiences constituting the stream of consciousness. Third, the various claims broadcasted through the professional journals regarding visualization of the neuronal networks of consciousness and its products are scrutinized. The results of this scrutiny are that,

thus far, and in spite of tremendous technical achievements on the part of the researchers in the area, none of these claims is correct. The validity of these results and what they bode for future research is left for the reader to evaluate, in the context of the formal structure of the methods and the pragmatic constraints that condition their implementation.

INTRODUCTION

It would not be an exaggeration to say that the greatest technological accomplishment in recent years in the field of cognitive neurosciences has been the development of functional neuroimaging methods. And, much like other revolutionary developments, functional neuroimaging has inspired unbridled enthusiasm and unrealistic expectations, even on the part of the experts themselves. These, in turn, have been countered with dissent and criticism (e.g., Nachev and Hussain 2007; Papanicolaou 2007; Logothetis 2008; Aue et al. 2009; Vul et al. 2009; Beck 2010; Miller 2010; Ropper 2010) leaving the field in a state of confusion. It appears, however, that most often the confusion arises simply out of a peculiar mode of use of language, peculiar in that it often calls to mind commercial advertising, a mode rather unexpected in the professional press. To quote just a couple of examples: *...recent advances in imaging technologies*, three pioneers in the area declare from the pages of a technical publication (Owen et al. 2009); *and in particular the ability of fMRI to detect reliable neural responses in individual participants in real time are beginning to reveal patients' thoughts...* (p.400). And they do so declare even though, as experts, they are assuredly aware of the fact that the fMRI does not record neuronal but hemodynamic responses, that it does not do so in real time and that expressions like *...beginning to reveal thoughts* do not involve less exaggeration than the plainer expression *...reveal thoughts* does. Another example is the widespread practice of calling *functional connections* what are in fact mere correlations of the temporal fluctuations in the hemodynamic activity of different brain regions, and calling *networks* the regions thus correlated. And, it is precisely such practices that tend to mislead the receptive yet unsuspecting reader who is accustomed to the ordinary use of the term *neural network* for denoting structures causally interacting at the level of their constituent neurons and neuronal circuits.

To avoid unnecessary confusion I will first define the terms that occur in this type of discourse, and I will then analyze the principles and facts underlying the functional neuroimaging methods. Once the principles and facts comprising the formal structure of these methods are made clear, it becomes surprisingly easy to gauge the methods' actual range and the proper limits of their application. To articulate that structure is the first aim of this commentary. The second is to deduce the possible range of their application; what aspects of mentation could be, in principle, imaged and what could not, assuming that all practical problems as well as issues of experimental design are satisfactorily settled. The third and final aim is to evaluate claims of actual contributions of these methods to cognitive neurosciences with special emphasis on the most conspicuous one, namely the claim that the mechanism of consciousness is, or is about to be, revealed, which eclipses even Dennett's (1991) confident pronouncement that *consciousness* is *explained*, and the insinuation that *reading*, in the brain's activation patterns, an individual's stream of experiences has been or is about to be accomplished.

THE THEORETICAL RANGE AND LIMITS OF FUNCTIONAL NEUROIMAGING

A.1. THE FORMAL STRUCTURE OF THE FUNCTIONAL NEUROIMAGING METHODS

Functional imaging aims at the visualization of patterns of brain *activity* and *activation*. Brain activity (and activation) has three interrelated aspects: neuronal signaling, giving rise to intracellular currents; changes in local metabolic rates contingent on local rates of signaling; and changes in local blood flow rates contingent on metabolic ones. Each one of these aspects is associated with electromagnetic signals which are collected non-invasively on the head surface and which form the basis for reconstructing images, or models, of the actual intracranial event patterns (e.g., Papanicolaou 1998 for an overview). Being mere models, the reconstructed images approximate, with varying degrees of fidelity, the original patterns they represent, imposing severe practical constraints in recreating the original intracranial patterns. Provided that all three aspects are interrelated, concordance of images based on neuronal signaling and constructed through magnetoencephalography (MEG), images based on metabolic changes and constructed through positron emission tomography (PET), as well as images representing local blood flow rates and created through both PET and functional magnetic resonance imaging (fMRI), obtained under the same circumstances, ought to be compatible, at least in their spatial aspect. Lack of such concordance is one of several signs of questionable validity of the reconstructed patterns and ought to serve as a guide in evaluating the literature.

Brain activity, also called *baseline* or *resting* activity, refers to neuronal signaling and its attendant physiological events (local metabolic and blood flow rates) when the brain is considered to be at rest. The brain, of course, is never resting. On the contrary, it is constantly active, controlling the various basic biological functions including the maintenance of the entire neuronal network that constitutes it. What is, therefore, meant by *resting* is the condition where the person whose brain activity is being imaged is under instructions not to engage explicitly in the performance of any specific task. Lately, the possibility has been considered that such a resting state may be defined by the presence of a *default mode network* (DMN), that is, a set of areas more active when no explicit experimenter-defined task is performed than when any such task is performed, and/or areas the hemodynamic fluctuations of which are correlated during resting conditions.

Activation, in contrast, is the intensification of signaling (and the attendant neurophysiological events) in different brain regions constituting the various functional neuronal networks that mediate the various ongoing behavioral or cognitive functions. Typically the pattern of intensified activity constituting activation is recognized by contrasting it with resting activity, when the brain works at default levels and when people are deemed to be in a condition of rest or *a control condition*. The sum total of neuronal signaling (or the associated metabolic and blood flow rates) that constitutes both activity and activation, may be referred to as *global* activity.

The patterns visualized through imaging (that is, the models of the actual patterns) correspond to three basic types of entities: first, to the activation of *functional networks* specific to each particular function, called forth by tasks, execution of which is presumed to require that, and only that, function; second, to activation putatively specific to particular products of functions, such as different behaviors, percepts, sensations, or thoughts; and third, to patterns of resting activity specific to different diagnostic categories, personality traits, demographic categories and relatively long lasting states (e.g., vigilance, craving, anger). In this work we will be exclusively concerned with the visualization of functional networks and the subset of their putative products, that is, conscious experiences, beginning with a description of the ubiquitous concepts *function* and *mechanism* of a function.

The word *function* stands for the processes or the set of subsidiary operations (conceived in the abstract) deemed to be necessary for generating individual acts or particular experiences. They may be understood as analogous to the algorithms comprising a computer program, designed to generate specific outputs. A brain mechanism on the other hand is the set of neuronal networks through which the algorithms that constitute the function are implemented, along with the particular neural code of the aforementioned algorithms.

The difference between a function and its mechanism is quite clear: a particular output can be generated by any number of alternative sets of algorithms, by any number of alternative functions conceived in the abstract. However, a concrete human visual sensation, for example, is most likely the output of just one among the several theoretically possible alternative visual functions, specifically the one that has been incorporated in the circuitry of the human brain during its evolution. For each function then, it is reasonable to assume the presence of a corresponding brain mechanism; it should be noted that function is not strategy, since strategies may involve a number of alternative functions or operations, and, therefore, alternative brain mechanisms. Also, because a brain mechanism is not a *center,* as the phrenologists and later the *diagram makers* (Head 1926) conceived it to be, but a set of circuits most often distributed widely over different brain regions, with each circuit participating not necessarily in only one but in several mechanisms of several different functions. It should be noted that, at present, functional neuroimaging may only disclose the areas that contain circuits constituting a particular neuronal network. That is, it discloses rather ill-defined regions which are engaged by the respective mechanism but which do not contain the mechanism itself. This is because the functional images do not provide information about the particular neural codes of the algorithms that govern the mode of working of the networks. Current and foreseeable improvements in technology may only disclose models of the networks and not models of the mechanisms.

In order to appreciate the range and the limits of the applicability of functional neuroimaging in visualizing the neural correlates of functions and their products, it is important to first understand the main facts and assumptions on which functional neuroimaging is based. On this basis alone, it is possible to deduce unambiguously the neurophysiological correlates of aspects of behavior and mentation that can and those that cannot be visualized. The possibility of visualizing the brain networks of functions and their products is predicated on the following: when a function is performed, its corresponding brain mechanism is activated, producing an activation pattern; a particular act, or a particular psychological phenomenon (a sensation, an act, a percept or a thought), is also produced. The correspondence between activation patterns and behavioral or psychological phenomena is one of the most fundamental and most widely held assumptions. It may in fact be treated as an axiom since in its absence functional neuroimaging would be pointless. It may be expressed by the following proposition:

1. To each and every single act and to each and every mental phenomenon, whether conscious or not conscious, corresponds a brain activation pattern. Therefore,

2. For every aspect or feature that differentiates any two similar acts or any two similar mental events there is a feature differentiating the corresponding activation patterns.

Also fundamental and widely accepted is this third proposition:

3. Individual acts and individual mental events are unique and non-repeatable in identical form; that they occur as such, only once in the entire history of the universe.

This proposition, although it appears rather occult and metaphysical, actually expresses the simple fact that the second of any two experiences, or the second of any pair of putatively identical acts and their corresponding activation patterns are the products of a slightly older and therefore slightly different system than the one that produced the first act and the first experience in the pair. Though this proposition may be disputed on philosophical grounds, to the degree that one adheres to the principle of correspondence (proposition 1), one must accept this one as well.

Equally plain and uncontroversial are the following two facts that complete the set of the fundamental factors that determine the range of functional neuroimaging:

4. The brain events (whether electromagnetic, metabolic or circulatory) that constitute the different activation patterns are qualitatively identical no matter to which trait, state, function, act or mental event they correspond.

That is to say, the blood flow patterns, for example, that correspond to any two sensations, are not qualitatively distinct no matter how qualitatively distinct the corresponding sensations may be. Rather the patterns corresponding to different acts and experiences, differ only in quantitative terms, namely, in what anatomical areas each transpires, to what extent each area is activated, when each becomes active in relation to others and for how long. And, it is precisely because patterns differ quantitatively that we can visualize them, by extricating them through the use of mathematical tools from the global activity in which they are embedded.

Finally, neuroimaging involves the fact that:

> 5. At any time period, many different mechanisms of many physiological, behavioral and psychological functions are active; therefore at any given time period a large and unknown number of different activation patterns, each at a different phase of its temporal unfolding, are transpiring, and these together form the global brain activity in which they are embedded.

In view of the fact that all patterns (whether of electromagnetic metabolic or circulatory events) are qualitatively homogenous and the fact that they are made up of the same type of brain events, the global activity that we would in fact see evolving, if we were to visualize it directly, would contain all patterns fused together, therefore undistinguishable, without further processing. With these five propositions in mind, the main question of the range and limits of functional neuroimaging may now be addressed: what kinds of patterns can be visualized, albeit in principle?

A.2. ARE THE NETWORKS OF FUNCTIONS VISUALIZABLE?

In light of the facts and principles outlined above, it appears that visualization of function-specific networks is feasible. To begin with, the outlines of many such networks have already been visualized. Moreover, the procedures for extracting them from the ongoing global activity in which they are embedded have already been developed, have already been shown to be efficient and are continuously refined. They are all based on the principle that every time the mechanism of a function is activated the resulting activation patterns, though dissimilar to each other in many ways since they correspond to particular acts and varied experiences (even if they are of the same kind), do have some common features since they are generated by the same mechanism. Therefore, if we subtract activation recorded during a resting state from that obtained during the execution of a function, we are likely to obtain the activation pattern specific to the function, as the residual. This is expected on the basis of the assumption that, whereas resting activity reflects brain events mediating all continuous and automatically performed biological functions, the latter reflects all of those plus the activation pattern specific to the function of interest. Or, if we know when the function-specific mechanism is activated each time and add algebraically the segments of the global activity beginning at the point the mechanism is activated, the invariant features of its successive activations (of the successive patterns) will emerge as an average pattern specific to the function of interest, whereas those of all other concurrently performed functions, being captured at different phases of their temporal unfolding each time, will be eliminated as physiological *noise*. It should be noted here that the two procedures outlined above (the former common to PET and fMRI and the latter to MEG) are examples of several mathematically analogous ones used with the different methods of neuroimaging and that all such procedures have the following two features in common: they require repeated activation of the same mechanism and they result in average patterns.

Using such procedures one may visualize more or less readily the functional networks of that subset of the behavioral and psychological functions that have neurological validity; that is, functions that are associated with particular brain mechanisms. The relative ease with which such networks are identified depends partly on the characteristics of functions that are briefly summarized below.

The first basic characteristic is the experimental accessibility of a function. The accessibility of the function of visual perception is obvious since all it requires for its mechanism to be reliably activated on demand is the presentation of the stimulus objects to be recognized. However, much more sophisticated experimental designs are required to isolate the mechanism of, say, the memory retrieval function and visualize it separately from those of other perceptual and mnemonic functions.

Experimental accessibility of functions is not only defined by the complexity of the system of functions in which the one of interest is typically embedded, but it is also defined by the degree to which it can be represented accurately by the activation tasks, or, to put it differently, on the methodological validity of those tasks. Methodologically valid tasks are those that specifically and exclusively call for a particular function. For example, the function of speech production is readily induced by any activation task that requires production of utterances. And the actual production of utterances during the imaging session is an unequivocal sign that the speech production mechanism was in fact activated. On the other hand, few activation tasks in the imaging laboratory activate the mechanism specific to the emotion of anger, and it is extremely difficult, if not impossible, to verify whether that mechanism has actually been engaged. Finally, the experimental accessibility of a function is defined by the consistency with which it can be repeatedly performed in the course of an imaging session, which is one of the prerequisites for constructing functional images. Once again, though the mechanism of visual object recognition may be repeatedly and consistently activated in the course of an imaging session, the same cannot be said with confidence for the mechanism of the sentiment of lust or of disgust. There, repeated exposure to stimuli that evoke the sentiments may lead to gradual habituation. Therefore the successive activations patterns may not be sufficiently similar to each other as required for constructing adequate images. However, these pragmatic considerations aside, networks specific to all neurologically valid functions are in principle visualizable.

A.3. Can the neuronal network of consciousness, viewed as a function, be visualized?

Is consciousness to be considered a neurologically valid function and is its mechanism visualizable? These are questions that besides being of abiding theoretical interest, have lately acquired practical importance as well, regarding, for example, people in a coma or *vegetative state*, recently renamed unresponsive wakefulness syndrome (Laureys et al. 2010); in such situations the need of an objective sign of the presence of consciousness can be quite urgent. In light of the facts and assumptions of functional neuroimaging, the answer to these questions is positive. The brain networks mediating consciousness, whatever they might be, are ipso facto visualizable, to the degree that the networks of any function that results in conscious experiences are visualizable. But, whether it is possible for the networks that are necessary for conscious experiences to be visualized separately from those of the various psychological functions (visual perception, for example), with which they are typically associated, is a different and more difficult question.

Although this separate visualization has yet to be achieved, it is not daring to predict that it could be, for the following reasons.

We can, and most often do, process information and act as if we had a perceptual experience, without being aware of our percepts and decisions; but even so, we act as though we have processed and utilized the relevant information as much as we do when we are aware of our percepts and decisions. The fact that much of our behavior and mentation is automatic or automated, that is, subliminal, and yet appears to require the operation of information processing mechanisms that conscious behavior also requires, leads to the hypothesis that, in addition to several perceptual and cognitive functions, we possess a function of consciousness which renders some products of our mentation conscious. This function may be conceived in different ways. For instance, it may be conceived as one that involves a single mechanism, separate from the other functions, accepting as its input the output of each of them, as shown in Figure 1.

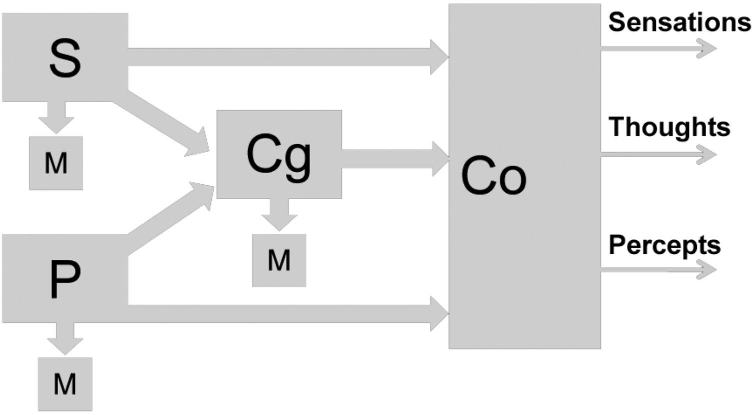

Figure 1: S, P, Cg, Co: mechanisms of Sensory, Perceptual, Cognitive and Consciousness functions, respectively. M: motor mechanisms including those of autonomic and hormonal responses.

Or, consciousness may be conceived as a set of modules each of which constitutes an extension and ramification of the mechanism of each psychological function as shown in Figure 2.

In either case, to the degree that consciousness is a neurologically valid function or a set of modular functions, its network(s) can be visualized using the same procedures that have been shown to be efficient in the case of all other functions. Specifically, the activation pattern associated with a set of conscious experiences is likely to be different from the activation pattern of mental processes that have not cleared the threshold of consciousness. This is because these subliminal events represent the output of the function-specific mechanisms only, whereas the conscious experiences represent the output of both those mechanisms and the mechanism(s) of consciousness.

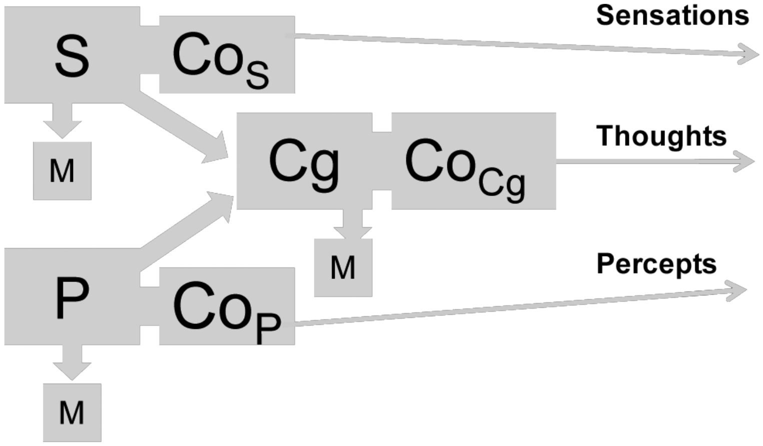

Figure 2: S, P, Cg, M as in Figure 1. Co_S, Co_P, Co_{Cg}, mechanisms of Sensory, Perceptual and Cognitive Consciousness modules.

What remains, of course, is the actual conduct of the relevant experiments that will produce reliably differing images of brain activation associated with subliminal and conscious perception of the same events. Once this requirement is fulfilled, the same procedures may be applied for assessing consciousness in people who are unable to verbally or symbolically communicate its presence, that is, to people in a coma or unresponsive wakefulness state.

But to really predict who in a coma will regain consciousness, one has to answer a different question, namely, whether it is possible to determine what kinds of ongoing, resting brain activity are indicative of the presence of intact, although not necessarily activated, mechanism(s) of consciousness. This question appears to also be empirically answerable: once patterns of ongoing *resting state* brain activity are classified and ranked in terms of the probability of their association with *waking* from a coma, they may then be prospectively applied to predict re-emergence of consciousness. Needless to say, before the requisite classification and ranking is completed, it would be premature and dangerous to divine the presence or absence of the ability of a comatose patient to regain consciousness. At present, hints or explicit statements to the contrary notwithstanding, no such separate visualization has been achieved. What has been achieved so far are glimpses, more or less tenuous yet promising, of some aspects or distinguishing features of neurophysiological events that might be part of the still occult network(s) of consciousness. We will deal with these in Part B of this paper.

A.4. ARE PATTERNS CORRESPONDING TO CONSCIOUS PRODUCTS OF PSYCHOLOGICAL FUNCTIONS VISUALIZABLE? THE CASE OF CONCEPTS.

As mentioned above, the second broad category of activation patterns corresponds to the products of functions, that is, percepts, sensations, sentiments, thoughts, intentions, in short, all the members of the sequence of transient experiences that constitute the stream of human consciousness. So, does the assertion made earlier that the network mediating conscious experiences is visualizable imply that single activation patterns corresponding to single, individual acts and experiences are also discernible, even only in principle?

In the first place, to *read* or to recognize or to interpret, in the ongoing flow of global brain activity, the successive thoughts, percepts, sentiments, and volitions that comprise a person's stream of consciousness, requires, much like the reading of a book, that one has already established the correspondence between sets of letters and particular word meanings, or, in the case under discussion, the correspondence between particular activation patterns and particular experiences. Only then could one discern in the stream of global activity the embedded patterns of the successive experiences and be able to announce, *now this feeling is experienced, now that intention is followed by that anticipation, etc*. However, as noted above, the principle of non-repeatability of particular transient experiences would render the establishment of such a dictionary of individual experience-specific patterns problematic. Yet, it has always been assumed from classical antiquity to present times (e.g., Hebb 1949; Pulvermüller 1999; 2003) that the recognition of particular experiences and the meaningfulness of the stream of consciousness requires the existence of concepts, that is, structures containing the invariant features common to sets of particular experiences. These are said to be encoded and stored in terms of cell assemblies or networks that, when retrieved, form patterns of brain activation (reverberating circuits, in Hebbian terminology). So, the initial question becomes: could such patterns, specific to concepts, be visualizable?

At present, and contrary to widespread rumors, not only do we not have a dictionary of such patterns, but we have yet to come up with one that corresponds to any one concept whatsoever. As for the aforementioned rumors, they are based on results of questionable specificity, and are contaminated by the repeated error of identifying the correspondence between external stimulus features and activation patterns with that between concepts and activation patterns. These issues will be dealt with again in part B of this commentary. At present, the following summary illustrates the statements above. Imaging experiments conducted for the purpose of visualizing brain activation patterns corresponding to any concept presuppose that such patterns, in the form of engrams or cell assemblies, exist somewhere in the brain where they remain nascent until a thought process or an external stimulus activates them and converts them into reverberating circuits (e.g., Pulvermüller 1999 for a modern rendition of the standard Hebbian model). Consequently, in a typical imaging experiment an appropriate stimulus is used to activate the corresponding concept (e.g., a picture of an object in the case of concrete concepts, and printed words in the case of abstract concepts). The stimulus is repeated several times. Each time it is presented it gives rise to a sensory pattern that codes its physical characteristics. This pattern is expected to be isomorphic to the stimulus and to be produced by the sensory cortex of the corresponding modality of the stimulus. Immediately after its production this pattern activates the corresponding cell assembly and the combination of both constitutes the pattern which corresponds to the specific unique and transient experience or the token of the thing that the concept represents.

Each repetition of the process by means of successive presentations of the same stimulus results in such a combined pattern embedded into the global activity representing all processes that contribute to rendering each token unique and non-repeatable, whereas the purpose of the experiment is to visualize its concept-specific component. That component is expected to emerge through the process of signal averaging, or a mathematically equivalent one; all components of activation that represent other concurrent processes not related to stimulus-specific or concept-specific patterns that are presumed invariant across presentations, are removed. What remains is a pattern consisting presumably of two components: the component due to stimulus-specific activity and the one specific to the concept. Provided that what is known to remain invariant across repetitions is the stimulus, the stimulus-specific component is highly likely to be represented in the pattern. Whether the concept-specific component is represented depends on whether the same stimulus activated the same cell assembly with the same pattern of interactions each time.

Besides considering the combined pattern or the stimulus-specific pattern as the concept-specific one, people frequently think that they have discovered the activation patterns corresponding, say, to the concept *table* or *rose* because having repeatedly presented a picture of a rose they can extract, using an averaging routine or an equivalent one, an average pattern. And, especially when the same pattern is obtained reliably, it is claimed to be the brain sign of the concept *rose*, since it has been evoked by the presentation of the picture of a rose. Instead, what may have been visualized could be the sign of the color red or the sign of small pictures, or of round objects, or of round red objects, or of objects that bring to mind particular smells, and so on, to infinity. At best, if we were to expend energy and resources to conduct of long series of experiments, through which we would successively eliminate alternative possibilities of the sign we think is that of a rose, we could increase the probability — without ever reaching certainty — that we have, in fact, visualized the brain sign of a rose.

A much longer series of experiments would be necessary to determine to the same degree of certainty that we have visualized the sign of abstract concepts like that of *justice,* and an even longer set of experiments for the sign of the conjunction *and* or *the square root of -3,* as opposed to that of *-4* or *-5*. In this case, an additional difficulty arises: unlike the case of concrete concepts, like those of roses and tables where their evocation is made highly probable by the presentation of the same perceptual stimulus, abstract concepts may not be as consistently evoked by the repeated presentation of their respective symbol (e.g., the sign √*-3* or the word *justice*), that the same notion springs to mind and the same activation pattern is lurking, in the ever-flowing stream of global activity every time the symbol is presented. This difficulty goes a long way to explain the fact that concepts whose activation patterns have been typically sought (see part B, below) are concrete ones. Moreover, as it will become clear when the relevant literature is reviewed in part B of this commentary, use of pictures to evoke the concept-specific activation patterns leads almost invariably to misattribution of the patterns to the concepts instead of to the actual physical stimuli. Nevertheless, the theoretical possibility of creating a dictionary of a reasonable length, consisting of patterns, the meanings of which are more or less accurate, though it is an extremely remote one, is not precluded by the formal structure of neuroimaging.

A.5. ARE PATTERNS CORRESPONDING TO CONSCIOUS PRODUCTS OF PSYCHOLOGICAL FUNCTIONS VISUALIZABLE? THE CASE OF TRANSIENT EXPERIENCES.

The student of cognitive neuroscience, not to mention the interested non-specialist, encountering in the scientific literature titles like *Decoding mental states from brain activity in humans* (Haynes and Rees 2006) or *Reading hidden intentions in the human brain* (Haynes et al. 2007) could hardly be blamed for believing that such a feat has either been accomplished or is imminent. In fact, it is neither. In this section we will demonstrate why the possibility of *reading* the stream of consciousness is infinitely small if not impossible, in view of the formal structure of neuroimaging; and, in part B of this commentary, we will relate why claims to the contrary are factually baseless.

Before proceeding, however, it may be useful to make absolutely clear what is meant by *reading* the stream of consciousness. It means that it is possible for the neuroscientist (of the future, at any rate) to be announcing while watching the unfolding of global activity of someone's brain, either off-line or in real time, such things as: Helen (the subject of the imaging experiment) has just felt an itch and an urge to scratch her nose but as she is intending to move her hand she experiences a flashback to the time her mother reprimanded her for doing just that…. and while the itch is passing away another memory from the same episode surfaces — the memory of the pleasure she felt from smelling the jasmine that was blooming in the yard — and now that experience of pleasure is also fading away and Helen wonders whether I (the neuroscientist) have really been able to see the memory of her pleasure on the screen of my neuroimaging device, or else, why do I seem to be so amused?...

To begin with, recognizing conscious experiences in the stream of unconsciousness by interpreting their signs (the corresponding activation patterns) in the flow of global activity requires the equivalent of a sizeable library of patterns corresponding to concepts which, as we have just shown, is unlikely to ever be forthcoming. Then we must assume that the stream of consciousness consists of a series of distinct, serially arranged, non-overlapping experiences. But this assumption is also untenable: serial order of behavior, motor, linguistic or otherwise does not presuppose serial order of the mental operations that produce it because such an order would exclude plans and intentions, and other anticipatory experiences. In fact, as Lashley (1951) argued elegantly many decades ago, serial order in verbal behavior would be impossible if the operations arranging the order of the behavioral units (e.g., the articulatory gestures necessary for articulate speech) were themselves serially arranged.

But even assuming that a concept library of the requisite size existed and that the stream of consciousness were a train consisting of distinct juxtaposed units — experiences strung out in time like beads in a necklace rather than a stream of always overlapping fluid elements — reading it would still remain a untenable technological ideal. The degree of similarity between patterns corresponding to concepts *jasmine* or *pleasure* or *smell* we would presumably already have, and the patterns corresponding to the transient present memory of the then transient percepts of that particular jasmine or the memory of Helen's pleasure — not any pleasure, but the pleasure associated with that aroma at that time in Helen's history in the context of that particular stream of her experiences — that degree of similarity is indeterminate. First, it is indeterminate because patterns associated with transient mental events count for an extremely small proportion of the global activity (proposition 5), and second, because they are made of the same material, that is, the very same brain events as all other simultaneously evolving or overlapping patterns (proposition 5). Consequently the odds that a sufficient number of features of any one transient experience-specific activation pattern would be sufficiently visible in the mercurial flow of the global activity for it to be recognized, and be sufficiently similar to the corresponding concept-specific pattern we presumably already have, to thus establish that it is, in fact, representing a token of that concept, are infinitesimal. More importantly, what renders such determination impossible is the requirement that those fragments of the experience-specific discernible (if discernible at all) pattern be compared with several concept-specific ones in our dictionary to establish which matches best. But even if, with our presumably perfect neuroimaging device, we were able to match the visible shred of the pattern embedded in the stream of global activity to all patterns in the dictionary, there would be no way to confirm the reliability of the match for the very simple reason that individual acts or experiences never repeat (proposition 3). We must therefore conclude that we will not obtain patterns specific to individual unique, transient experiences, let alone *read* the unfolding stream of conscious experiences in the flow of global activity.

TOWARDS VISUALIZATION OF THE NETWORK OF COUNSCIOUSNESS AND ITS PRODUCTS

B.1. THE DEFAULT MODE AND THE RESTING FUNCTIONAL NETWORKS.

It was mentioned in A.5 that to the degree that the neuronal networks of cognitive functions, like perception, that produce conscious experiences are visualized, the networks of consciousness must be visualized as well, albeit not separately from those of the cognitive functions. Also mentioned is the fact that, although in principle separate visualization of the networks of consciousness is not precluded by the formal structure of the neuroimaging methods, such a feat has yet to be accomplished. In this unit, efforts towards the accomplishment of this goal will be reviewed.

The main approach that characterizes all these efforts is based on the rationale that the networks active while the person is in a state of awareness but not engaged in any particular task (therefore not engaged in the execution of any cognitive function) may be said to be the networks mediating consciousness. Although this rationale appeals to common sense, it is, as will be shown, rather problematic, rendering the outcome of all research in the area inconclusive at best.

It is assumed that the *resting* or *control* condition where the subject is instructed not to engage in any particular cognitive task is most suitable for activating the mechanism of the function of consciousness alone. During that condition it has been repeatedly found that a set of brain structures display correlated temporal fluctuations in their hemodynamic response and the very same structures display reduced activity when any of a large number of cognitive tasks is deliberately performed. This network has been named the default mode network (DMN) (e.g., Shulman et al. 1997; Gusnard and Raichle 2001; Raichle and Snyder 2007; Buckner et al. 2008). In addition, during the resting condition a number of other networks, namely the various resting functional networks, are identified on the basis of the coherence of their hemodynamic fluctuations, each associated with a set of functions in that they become activated when the particular functions are deliberately executed. In other words, they display a mode of functioning opposite to that of the DMN. These are called resting functional networks. So, the first question emerges: is the DMN the network mediating consciousness, or does the DMN, along with the rest of resting functional networks, constitute the consciousness network?

If the DMN were the consciousness network, one would expect that during production of conscious experiences it would display more, not less, activation much like all other functional networks engaged in the production of conscious acts and experiences. The fact that the DMN does not behave this way has led to the hypothesis that it represents the outlines of activated mechanisms of several alternative, *intrinsic* processes being continuously and automatically, that is unintentionally, performed (Shulman et al. 1997; Gusnard and Raichle 2001). The reason for attributing to the functional networks passive and to the DMN active status during the *resting* condition is that the latter is inhibited during the deliberate execution of *extrinsic* functions whereas the former are enhanced under such circumstances.

To be certain that the DMN is indeed part of the mechanism of intrinsic functions it would be necessary to observe it in an enhanced state when one or more of these functions are unfolding in earnest. Yet that would be an extremely difficult situation to create experimentally, especially since the functions allegedly mediated by the DMN remain conjectural, and since all of them are believed to be automatic and unintentional, and therefore hard to deliberately manipulate in the context of standard experimental paradigms. And, to perform the automatic, non-intentional intrinsic functions intentionally is, by definition, impossible.

One would be tempted here to overcome the dilemma by considering the state of meditation, deliberately entered into, as the closest substitute for the state of automatic and unintentional enhancement of the *intrinsic* functions. So, what happens during that state? Are the DMN constituent structures more active than usual? Do they exhibit tighter connectivity with each other, and enhanced connectivity among their constituent voxels? Either of the above or both of the above? The evidence here is fragmentary and inadequate to settle the issue, but provocative nevertheless. Hölzel et al. (2007) reported that at least one constituent DMN region, the medial prefrontal cortex (and also a region not normally associated with the DMN, the anterior cingulate cortex) showed more activation although no higher internal connectivity, or connectivity to any other structure, in meditation practitioners than in control subjects who had no meditation experiences during a resting state. This being the case, the relation of the DMN to the mechanisms of consciousness remains inconclusive. Equally inconclusive have been the results of studies in which changing levels of conscious awareness were related to alterations in the DMN and the resting functional networks, such as the one described below.

The DMN, a pair of resting executive-control networks, and an auditory and a visual resting network, were studied in a group of normal subjects during an awaking normal consciousness state and during states of light and deep propofol-induced anesthesia associated with varying degrees of disruption of awareness (Boveroux et al. 2010). During the normal state of consciousness, the DMN consisted of the structures that define it most consistently, namely the precuneus and the posterior cingulate cortex, the medial prefrontal cortex, the temporo-parietal junction region, plus structures reported as parts of the DMN but less consistently, such as the superior frontal sulci, the parahippocampal cortex, the lateral temporal areas and, finally, the brainstem and the thalamus.

The left and right hemisphere executive networks were found to consist of the left and right, middle, inferior and superior cortical regions, the anterior cingulate cortex, the temporo-occipital and the posterior parietal regions. They also shared with the DMN the thalamus but, rather unaccountably, they did not share the brainstem, a fact raising doubts regarding the nature of any network, especially an executive one that can function in the absence of brainstem contribution.

At the other extreme, during the state of unconsciousness, the connectivity in all five networks was reduced and a linear relation was found between the degree of network integrity (i.e. strength of correlations of the hemodynamic fluctuations of the constituent structures) and the level of consciousness. Conspicuously missing from that account were changes in the behavior of the brainstem as a function of the level of consciousness. Also paradoxical were the negative correlations between the fluctuations in the activity of the thalamus to the rest of the DMN structures, and the two executive networks during the state of unconsciousness, as compared to the reduced yet positive correlations of the thalamic activity fluctuations with the two sensory networks. Moreover, the connectivity of the two sensory networks did not decrease during unconsciousness as the other networks did. In spite of the fact that this study raises a number of questions regarding the extent of each network and the nature of the connectivity among constituent cortical and subcortical areas, it does answer the question, however tentatively, whether only the DMN is to be considered part of the mechanism of consciousness. And the answer is negative. To the degree that the executive networks degrade with decrements in consciousness, they should also be considered parts of that mechanism. Moreover, replication of the finding that the sensory networks do not follow the same trend may point to a means of separating consciousness-specific from sensory function-specific networks and, ultimately, from mechanisms.

Less easily interpretable are the data from studies of changes in the resting state networks in deep slow wave sleep (SWS). The main difficulty with these studies was the fact that the amount of activation rather than the degree of connectivity was used to define the networks. For example, Maquet et al. (1997) report a decrease in activity during SWS as compared to the awake resting state in a number of areas including the precuneus but also in areas other than those implicated in any resting functional networks or the DMN. If reduction involved the entire DMN then one could conceivably make the argument that the DMN is representing (part of) the consciousness mechanism, but reduction was reported in only one constituent DMN structure plus other structures outside the DMN. Additional studies of brain activity during sleep further frustrate any attempts at integrating and interpreting the findings since sometimes reduction of the precuneous activity is reported (Maquet et al. 1996, 1997), and sometimes not (Braun et al. 1997), during deep SWS.

In view of these observations it appears reasonable to conclude that, first, thus far, the definition of the consciousness network(s) has not become any clearer; second, the entire set of resting networks may be part of the mechanism of consciousness; and third, changes in any one of the networks may be suggestive, at best, of the level of consciousness. This third conclusion is important to keep in mind when evaluating the correlations of different levels of disordered consciousness with different degrees of coherence of the DMN network, or any other of the set of resting functional networks, reviewed in the following section of this commentary.

B. 2. THE EVIDENCE FROM COMA STUDIES

In the studies to be reviewed, two agendas have been pursued in tandem: first, the discovery of the cerebral mechanism of consciousness; and second, a way to judge objectively on the basis of brain activity patterns (a) what is the subjective state of consciousness in a particular individual and (b) what are the prospects for a given individual to regain normal consciousness.

Lesions in the brain, whether they are produced by traumatic injury or other causes (e.g., anoxia) often result in transient, persistent or permanent alterations to the state of consciousness. Although these changes in level of consciousness are gradual they have been conveniently segregated into three discrete states (e.g., Monti et al. 2010): *coma*, *vegetative state*, and *minimally conscious state*. Complete absence of any signs of consciousness for more than one hour defines coma. Comatose patients are unresponsive even to painful stimulation and remain with eyes closed, showing no evidence of awareness of their environment or themselves. A less severe consciousness disorder is the vegetative state characterized by opening and closing the eyes without any evidence of awareness of self or environment, or the ability to communicate verbally, thus the new term *unresponsive wakefulness syndrome* (Laureys et al. 2010). Signs of consciousness are apparent in the minimally conscious state where patients exhibit, albeit intermittently, purposeful behavior, awareness of the environment and rudimentary ability to communicate and comprehend language. Beyond these three states of impaired consciousness there is a state characterized by perfectly normal consciousness yet nearly complete inability to communicate via any form of behavior requiring the voluntary musculature except vertical eye movements, called the *locked-in* syndrome. This state is often confused with the vegetative state, although it does not involve any known deviations from the normal state of consciousness.

The first attempts to establish relations between brain physiology and the various states of aberrant consciousness (Vevy et al. 1987; Rudolf et al. 1999; Schiff et al. 2002) focused on the global and regional amount of brain activity, initially assessed through blood flow and glucose utilization measurements using PET, and found, reliably, substantial reduction of that activity in the vegetative state as compared to the resting state of normal subjects. Besides pervasive reductions in overall activity levels, De Volder et al. (1990) reported substantial reductions in particular regions, specifically the parieto-occipital regions and the medial frontal regions near those later identified as components of the DMN. Regional decreases in vegetative patients as compared to normal volunteers involving DMN regions were recently found by Kim et al. (2010) through glucose utilization measurements. Specifically, reductions were found in the posterior cingulate, the left precuneus (but not in the medial prefrontal cortex or the temporo-parietal junction area, also parts of the DMN). Once the presence of resting networks, especially the DMN, defined in terms of correlated fluctuations in the activity of constituent brain regions, was demonstrated, the possibility that the degree of correlations (somewhat prematurely named *connectivity strength*) in the fluctuating activity of the areas comprising such networks may reflect the level of consciousness, was examined. The first such study (Laureys et al. 1999) involving a series of four patients in the vegetative state and a relatively large sample of control subjects in a resting state, and using glucose metabolism measurements through PET, addressed both the possibility that regional reduction and the reductions in degree of *connectivity* of DMN areas would identify deviations from the norm specific to the vegetative state. They found that in the patients, a subset of areas typically considered part of the DMN showed reduced activity. In addition, they found that the posterior cingulate cortex, which almost invariably features as one of the main DMN component structures, was less connected with premotor and prefrontal areas, thus establishing the utility of the DMN connectivity level as a sign of the level of consciousness, in addition to the degree of activity reduction.

The utility of this measure was indeed successfully exploited in a series of studies to differentiate among normal resting activity conscious states and states varying in degree of aberration from normal. Boly et al. (2009) demonstrated that fluctuations in the fMRI signal could be used in the same capacity. They reported that for one patient in the vegetative state, the degree of connectivity between the posterior cingulate and the precuneus regions with the thalamus was lower than what was found in a control group of 41 normal subjects. Yet, the degree of connectivity among cortical regions of the DMN did not differ from that observed in normal individuals. These data suggest that a possible distinctive feature between activity patterns indicative of normal versus compromised consciousness is reduction of the functional connectivity (literally, in correlated fluctuations of activity) in the cortex and the thalamus.

But a rather different suggestion emerges from a more recent and larger study of the same kind, by the same group of investigators (Vanhaudenhuyse et al. 2010), involving 14 patients in various states of compromised consciousness and an equal number of control subjects. In this study, the existence of the normal resting DMN was identified and the expected integrity of that network among (fully conscious) patients in the *locked-in* state was verified, as was the anticipated negative relation between the degree of connectivity of the DMN and the different states of compromised consciousness from a minimally conscious state to a comatose state. Yet, this time, reduction in connectivity between the thalamus and the cortical components of the DMN was not recognized as the distinctive feature differentiating conscious from non-conscious states. Rather, degree of connectivity of the precuneus was found to be the aspect of the DMN that differentiated normally conscious from minimally conscious or unconscious individuals.

In spite of differences in findings, these studies have established that aspects of the DMN differ in conscious and unconscious individuals. Obviously, they by no means carry the implication that the DMN is the consciousness network, because the degree of integrity or connectivity of other resting functional networks may also co-vary with level of consciousness, as earlier studies, reviewed in the previous section of this work, have shown.

Moreover, in view of the fact that features of the DMN (or other resting functional networks for that matter) reliably differ from one conscious state to the next, at the level of groups rather than individual patients, it is premature to claim that it is now possible to judge, on the basis of brain activity patterns, the level of consciousness of a particular individual. In the future, such a feat may become reality.

It could be urged that that day is near because, as the next study to be described shows, we are now in a position to specify brain activity features with sufficient accuracy for a machine to differentiate and accurately classify activity patterns as belonging to conscious as opposed to non-conscious individuals. It is indeed true that such automatic classifications have been accomplished (Phillips et al. 2010). But such a post hoc classification is as practically insignificant as it is technically impressive. Patterns consist of multiple features, thus groups may be separated, even reliably, on the basis of some such features. But, unless these features are specified and the constraint is imposed that classification of future instances are to be based on those and not some other features, machine prediction is not a very reassuring index of progress in specifying the networks of consciousness and in classifying individuals as conscious or non-conscious.

The pioneering studies reviewed have all employed the resting activity state pattern of normal subjects as a standard from which to assess deviations associated with each aberrant consciousness state. The fact that they have mainly utilized only one of the several networks that characterize that pattern, limits the scope of their conclusions regarding the nature of the mechanism of consciousness and the prognostic efficacy of the disrupted DMN. But a far more serious limiting factor of these studies is the absence of a consistent definition of the consciousness networks, i.e., whether degree of connectivity or level of activity of the constituent structures, or both, define such networks.

To enlarge the scope of these conclusions, future studies on these patient samples would need to consider the fate of all discernible resting networks, not only the DMN, and to define them consistently. If the current research efforts in this area continue unabated, one may safely predict that this feat will be accomplished and a more supple picture of what brain activity patterns correlate with the state of normal consciousness and what deviations of that pattern correspond to each of the various states of disordered consciousness will be specified.

B.3. CLAIMS OF VISUALIZATION OF CONSCIOUS EXPERIENCES: THE CASE OF CONCEPTS

All functions unfold unconsciously. We are never aware of the neural events that control the syntactic order of our utterances, any more that we are aware of the algorithms that turn patterns of energy impinging on our retinae into visual objects. We are however conscious of some products of some functions, especially the products of those psychological functions we call cognitive. These products are experiences: percepts, sensations, intentions, sentiments, thoughts, and mental images. Experiences are by definition conscious. There is no such thing as a sensation not sensed, a percept not perceived or an intention not intended. Nevertheless we often talk about unconscious experiences. In such cases we surmise that something similar to an experience must have transpired because we behave as if it actually did. We stop, for instance, at an intersection and then we realize that the light is red. The examples abound. Their presence has led to the hypothesis that those functions that produce conscious experiences may also, at times, produce an unconscious variant of them. This hypothesis has not been addressed with imaging studies. Although such studies are, as was mentioned earlier in section A.5, perfectly feasible, they have not been undertaken, so it is premature to declare that we can recognize which activation patterns correspond to conscious experiences and which to their unconscious equivalent. Thus far, all studies have attempted to demonstrate only the correspondence of activation patterns to conscious experiences.

As alluded to in section A, all transient experiences are tokens of particular kinds. This is why we are able to recognize them. Becoming aware that I see a tree is, fundamentally, the realization that the stimulus-specific pattern produced by the stimulus activates the concept-specific one which determines that the former represents an example of a kind of object, namely, the kind *tree* (similarly, with the preposition *before*, the noun *love*, or the notion *square root*). In all cases, the unique, specific and, according to the proposition in section A.2, unrepeatable experience is known only by reference to a concept.

Are these concepts neurologically valid enough to justify attempts to disclose the activation patterns to which they correspond? The principle of correspondence that applies to tokens, that is to transient experiences and to some functions, may not apply to hypothetical or abstract entities. Those, being hypothetical, may or may not have neurological validity. It is of course possible that there are nascent circuits in the brain which are codes for each concept, and recognition is indeed the matching of stimulus-specific activation to activated cell assemblies, that is to say, to reverberating concept-specific circuitry as Hebb (1949) taught and as many believe today (e.g., Pulvermüller 1999, 2003).

Unlike the case of functions, whose validity is supported by lesion data, and unlike tokens, that is, transient experiences, the validity of which is enshrined in the principle of correspondence, no concept has ever been selectively obliterated as a result of a lesion, and no reverberating, concept-specific activation pattern has been clearly identified.

Nonetheless, there is reason to believe that concepts are neurologically valid. Tokens of a particular concept, though unique and unrepeatable in all their details, must have some invariant features in common if they are to be classified together and be recognized as instances of a particular concept. And, it is precisely the aspects of the token-specific activation patterns that correspond to those invariant features that repeat across occurrences of the same kind of token, that are extractable through the technique of averaging or mathematically-equivalent techniques used in neuroimaging. But if this assumption is true, why have there not been any discoveries of concept-specific patterns? Most likely, it is because of the difficulties inherent to the methods of imaging and those specific to the nature of the concepts previously commented on in section A.6. Nevertheless, as also mentioned in that section, visualizing concept-specific patterns is not incompatible with the formal structure of neuroimaging.

Claims are often made to the effect that specific correspondences of concepts and patterns have been established within the acceptable margin of uncertainty common to all inductive inferences. However, as discussed below, the truly ingenious experiments supporting these claims are problematic. They have uniformly failed to establish sufficiently specific correspondences between patterns and concepts, and whatever degree of specificity they have established is not between activation patterns and concepts but between patterns and stimuli. That is, they have simply demonstrated that the isomorphism known to exist between stimulus features and patterns of peripheral sensory activity is also to be found between the stimulus features and activity patterns in the sensory cortex. Nevertheless, these findings, though they fall short of supporting the claims made, pave the way for the future discovery of higher order isomorphisms between concepts, at least concrete ones, and activation patterns.

In a recent review of the relevant literature, Haynes and Rees (2006) state: *Here we will review new and emerging approaches that directly assess how well a mental state can be reconstructed from non-invasive measurements of brain activity in humans* (p. 523). What they in fact do is provide a review of sophisticated methods that directly assess how well a stimulus pattern can be reconstructed from recordings of sensory brain responses in humans. Yet, the difference between a stimulus and a mental state is considerable, one being a pattern of physical energy impinging on a biological system of sensors, and the other a subjective experience. Haynes must have been aware of the difference, unless he only became aware of it three years later (Haynes 2009) when he stated that *...the perceived properties of objects are often different from the physical properties, as is the case in contextual interactions in brightness and color perception. For example, the encoding of chromatic signals in the* *retina and in V1 does not match the subjects' conscious perception…* (p. 198). This being the case, the question becomes: is there any evidence that concept-specific patterns have been forthcoming such that by looking at the pattern of one's brain activity an observer may be able to tell what that person is experiencing?

As mentioned previously, perceiving in the sense of becoming aware of the nature of an object entails at least two distinct processes each associated with its own activation pattern, namely, registering of the stimulus and activation of the engram that corresponds to it, that is, turning the concept-specific cell assembly into a reverberating circuit. Possibly it involves a third process, which may be some composite of the two. But of the three, if any has in fact been reconstructed, it is the first, not the second or the third. And it is the second that would properly be named *concept-specific*, since the first is stimulus specific and therefore possibly different from the second and the third, as suggested by Haynes' above quoted admission.

It is common knowledge that concepts corresponding to objects are of different types. There are concrete or abstract concepts corresponding to objects rendered by nouns (e.g., tree, value), those corresponding to object attributes rendered by adjectives (e.g., red, big, profound) and those corresponding to activities, actual or intended, rendered by verbs, adverbs, prepositions and other function words. Of the above types, only concrete concepts corresponding to (usually) visual objects and to sensory attributes of objects are alleged to be predictable on the basis of their corresponding patterns. In evaluating the truth of these allegations, two basic criteria ought to be consulted: first, the criterion of specificity, that is, when looking at a pattern of symbols we can tell what concept it represents. For example, when looking at the pattern of symbols, *TABLE* we can tell that they correspond to the concept of a table and to no other concept. Second, the criterion of appropriate correspondence, namely that the pattern is specific to the particular concept and not to the stimulus used to evoke the concept. To make sure that a recorded pattern corresponds to the concept and not the stimulus, we should obtain it in response to some cue other than the stimulus (e.g., to the spoken word *table*) or spontaneously, in the absence of any cue or stimulus. These would be the proper conditions for judging the correspondence with certainty and, as will soon be obvious, these conditions are far from being met.

Thus far, investigators have managed to derive, mostly using fMRI, patterns specific to broad concept categories such as *faces*, *cats* and *man-made objects* (e.g., Haxby et al. 2001; Carlson et al. 2003; O'Toole et al. 2005; Kamitani and Tong 2005), but always too broad and too few to establish a reasonable degree of specificity. Moreover, they have not managed to derive concept-specific but stimulus-specific activation patterns, as the latter have been defined previously. Exceptions to this general trend are studies like the one conducted by O'Craven and Kanwisher (2000) who endeavored to derive patterns specific to the mental images of faces and of places and, separately, to stimuli of these two types of objects. Here again, the degree of specificity achieved was considerably lower than what would be required to justify the claim that concept-specific patterns have in fact been specified.

B.4. Claims of visualization of conscious experiences: the case of intention

What has been said of concepts can also be said of intentions. Haynes et al. (2007) thought it possible to image intention-specific patterns by instructing subjects to experience, several times in a row during an fMRI session, one of two intentions: the intention to add together two numbers or the intention to subtract them. The authors contend that this possibility materialized, since they were able to record from a group of eight normal subjects, two distinct patterns of activation, one specific to each of the two types of intention, namely, the intention to add and the intention to subtract. In fact, they obtained four distinct patterns, one pair associated with the two intentions and another pair associated with their implementation, the actual cognitive act of subtracting or adding. All four patterns involved the region of the medial prefrontal cortex, a region that features predominantly in the DMN (e.g., Shulman et al. 1997; Gusnard and Raichle 2001; Raichle and Snyder 2007). The fact that the subjects could choose by themselves whether to intend to add or to subtract a pair of numbers that were about to appear on a screen, rendered, according to the authors, the intention-forming process an *internally generated* task of the type that is said to be carried out during rest and which results in the formation of the DMN. In view of the demonstrated success of the application of the sophisticated pattern classification procedure to identify the four distinct patterns, the authors felt justified to declare that they had demonstrated that reading the hidden intentions forming in their subject's brains is a feasible project. However, what was actually accomplished appears to fall far short of what was claimed.

In the first place, it is not at all clear what it means to *intend to subtract* and keep on doing so continuously, for periods ranging up to 10.8 seconds. It could mean that subjects decided what to do right after the cue and then engaged in rehearsal of the decision for the rest of the period (e.g., thinking to themselves, *add*, *add*, *add…*). And, *rehearsal* is certainly not *intending*. This being the case, the obtained patterns would be specific to what was rehearsed or to the operation of rehearsal itself, a constituent operation of the function of memory. But it could also have been possible that the subjects engaged in a number of alternative cognitive operations. Nothing in the experimental design controls for these possibilities. Therefore the obtained patterns could be said to correspond to a number of different functions and experiences besides the intention to add or the intention to subtract. But, there is another curious point about this study. The region whose activation discriminated the four patterns is part of the DMN; it should be deactivated during a structured and deliberate forced choice task, according to the very definition of the DMN. Yet, in this experiment the region was found to be activated instead. For these reasons, the authors conclusion that they *have demonstrated that regions of both medial and lateral prefrontal cortex contain localizable task-specific representations of freely chosen intentions* (p. 324) is clearly unjustified. Moreover, and more generally, only if it were possible for one to recognize which patterns, among several, is the *intention to add* or the *intention to subtract*, in the absence of prior knowledge of the particular activation conditions, would the conclusions drawn in this type of study be justified.

Although no concept-specific or intention-specific patterns have been established, tremendous progress in multivariate classification algorithms for classifying and recognizing stimulus-specific (but not concept-specific) patterns has been made thus far (e.g., Cox and Savoy 2003; Kamitani and Tong 2005; Kriegeskorte et al. 2006; Haynes and Rees 2006; Miyawaki et al. 2008; Haynes 2009). In view of such progress, the prediction that the sought-after library of patterns may, in the future, become the repository of some concept-specific ones is not unrealistic. Whether enough such patterns can be accumulated to make some types of *mind-reading* possible is a question best answered by the reader, given what has been said thus far and is to be said in the concluding section of this commentary.

B.5. CLAIMS OF VISUALIZATION OF CONSCIOUS EXPERIENCES: THE CASE OF TRANSIENT EXPERIENCE IN THE STREAM OF CONSCIOUSNESS.

We have concluded, on theoretical grounds, that we may not expect to obtain patterns specific to individual unique, transient experiences, that such a feat is beyond the capabilities of current functional neuroimaging methods. As was stated in section A.2, the non-repeatability principle along with the fact that patterns corresponding to transient experiences unfolding in real time are embedded in the flow of the global activity, preclude the identification of transient experience-specific patterns. Yet a recent demonstration that brain responses to single stimuli (Jung et al. 2001) can be discerned in the raw recording would, in the minds of some readers, challenge these assumptions. This possibility creates an additional motive for considering closely the efforts of investigators to match transient experiences to activation patterns, best represented by the work of Kay and his associates.

Kay et al. (2008) first exposed viewers to hundreds of different natural images while recording their hemodynamic response from visual areas VI, V2, and V3 using fMRI. For each of the pictures viewed they derived a distinct pattern of activation extending over the voxels into which the above named areas were divided. Next, they showed each viewer a series of 120 new pictures that were not part of the initial set of 1,750 pictures for which activation patterns were established. While viewing each of these new pictures the hemodynamic response was again recorded and 120 new activation patterns were computed, extending over areas V1-V3. Next, the investigators did what is generally believed that the brain does in order for us to recognize a particular experience as being a token of a particular type. Namely, they compared each new pattern to those already in the library. Specifically, the activation patterns of the old pictures were used to predict the nature of the new picture. And, *...the image whose predicted voxel activity pattern was most correlated (Pearson) with the measured voxel activity pattern was selected* (p. 353). Two subjects were used, and correct identification of the pictures that the subjects had seen was 92% and 72%.

Between this achievement and the possibility of reading even isolated percepts (let alone abstract thoughts) lies a veritable chasm. Once again the relationship established in this study was between stimuli and activation patterns, not necessarily between experiences and patterns, which means that, conceivably, the same results could have been obtained by using as a subject a suitably equipped robot or a patient in a vegetative state, neither of which may be said to have experiences.

Nevertheless, it is possible that percepts and pictures involve similar patterns. If the same pattern obtains when one thinks of a pencil and when one sees a pencil, then one may claim that single *words* in the brain's book have been read, that individual contents of consciousness have been discerned in the flow of brain activity. But that the equivalent of such single *word* reading will eventually be expanded to reading of the entire stream of consciousness (with or without having first trained the algorithms with samples of the activity of the individual person whose consciousness contents are to be *read*) is a virtually impossible scenario for reasons that have already been discussed.

CONCLUSIONS

It is certainly theoretically possible that functional neuroimaging will be instrumental in revealing the brain networks of behavioral and psychological functions including the network of consciousness. It is also possible, although practically difficult, to visualize patterns corresponding to some concepts. And, it is possible only because concepts consist in the invariant, therefore repeating features common to sets of individual transient experiences, features that define the kind to which each set belongs, that define their time-invariant *structure*. However, the a priori probability that transient patterns can be isolated, which specify individual experience, is practically null, and the feat of reading in the ever-flowing stream of global brain activity, an individual's streams of consciousness appears to be well beyond the capacity of neuroimaging methods as we know them, even if they were refined to the point of perfection. The empirical data reviewed concur with the theoretical expectations. Aspects of the networks of consciousness, albeit alloyed with those of other cognitive functions, are beginning to be discerned. However, no evidence that concept-specific or transient, experience-specific patterns are about to be identified, is to be found in the literature. As for the odds of such patterns emerging from future research, it is for the reader to estimate.

ACKNOWLEDGEMENTS

A version of part A of this paper was presented in a conference at the Gaia Center sponsored by the Goulandris Foundation in October of 2009. I wish to thank Professors Fotis Kafatos and Athanasios Focas for their kind invitation and Mrs. Goulandris and the staff at the Gaia Center for their hospitality and interest. The present version of this paper was written while I was supported by NIH and Department of Defense grants (NIH P50HD052117-03, NIH P501NS046588-05, and W81XWH-08-2-0135). I wish to express my deepest gratitude to Professor Bruno Breitmeyer of the University of Houston for helping me clarify several obscure points in a previous version of this manuscript and for prompting me to consider others that I had overlooked.

REFERENCES

Aue T, Lavelle LA, Cacioppo JT (2009) Great expectations: what can fMRI research tell us about psychological phenomena? Inter J Psychophysiol 73:10-16

Beck DM (2010) The appeal of the brain in the popular press. Persp Psychol Sci 5:762-766

Boly M, Tshibanda L, Vanhaudenhuyse A, Noirhomme Q, Schnakers C, Ledoux D, Boveroux P, Garweg C, Lambermont B, Phillips C, Luxen A, Moonen G, Bassetti C, Maquet P, Laureys S (2009) Functional connectivity in the default network during resting state is preserved in a vegetative but not in a brain dead patient. Human Brain Mapping 30:2393-2400

Boveroux P, Vanhaudenhuyse A, Bruno M, Noirhomme Q, Lauwick S, Luxen A, Degueldre C, Plenevaux A, Schnakers C, Phillips C, Brichant J, Bonhomme V, Maquet P, Greicius MD, Laureys S, Boly M (2010) Breakdown of within- and between-network resting state functional magnetic resonance imaging connectivity during propofol-induced loss of consciousness. Anesthesiology 113:1038-1053

Braun A, Balkin T, Wesensten N, Carson R, Varga M, Baldwin P, Selbie S, Belenky G, Herscovitch P (1997) Regional cerebral blood flow throughout the sleep-wake cycle: an H215O PET study. Brain 120:1173-1197

Buckner RL, Andrews-Hanna JR, Schacter DL (2008) The brain's default network: anatomy, function, and relevance to disease. Ann NY Acad Sci 1124:1-38

Carlson TA, Schrater P, He S (2003) Patterns of activity in the categorical representations of objects. J Cogn Neurosci 15:704-717

Cox DD, Savoy RL (2003) Functional magnetic resonance imaging (fMRI) brain reading: detecting and classifying distributed patterns of fMRI activity in human visual cortex. Neuroimage 19:261-270

De Volder A, Goffinet A, Bol A, Michel C, de Barsy T, Laterre C (1990) Brain glucose metabolism in postanoxic syndrome. Archives Neurol 47:197-204

Dennett, DC (1991) Consciousness explained. Little Brown and Company, Boston, MA

Gusnard DA, Raichle ME (2001) Searching for a baseline: functional imaging and the resting human brain. Nature Rev Neurosci 2:685-694

Haynes J (2009) Decoding visual consciousness from human brain signals. Trends Cogn Sci 13:194-202

Haynes J, Rees G (2006) Decoding mental states from brain activity in humans. Nature Rev Neurosci 7:523-534

Haynes J, Sakai K, Rees G, Gilbert S, Frith C, Passingham R (2007) Reading hidden intentions in the human brain. Curr Biol 17:323-328

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293:2425-2430

Head H (1926) Aphasia and kindred disorders of speech. Vol 1. Cambridge University Press, Cambridge, UK

Hebb DO (1949) The organization of behavior: a neuropsychological theory. Psychology Press, East Sussex, UK

Hölzel BK, Ott U, Hempel H, Hackl A, Wolf K, Stark R, Vaitl D (2007) Differential engagement of anterior cingulate and adjacent medial frontal cortex in adept meditators and non-meditators. Neurosci Letters 421:16-21

Jung TP, Makeig S, Westerfield M, Townsend J, Courchesne E, Sejnowski TJ (2001) Analysis and visualization of single-trial event-related potentials. Human Brain Mapping 14:166-185

Kay KN, Naselaris T, Prenger RJ, Gallant JL (2008) Identifying natural images from human brain activity. Nature 452:352-355

Kamitani Y, Tong F (2005) Decoding the visual and subjective contents of the human brain. Nature Neurosci 8:679-685

Kim YW, Kim HS, An YS, Im SH (2010) Voxel-based statistical analysis of cerebral glucose metabolism in patients with permanent vegetative state after acquired brain injury. Chinese Medical J (Engl) 123:2853-2857

Kriegeskorte N, Goebal R, Bandettini P (2006) Information-based functional brain mapping. Proc Nat Acad Sci 103:3863-3868

Lashley KS (1951) The problem of serial order in behavior. In: Jeffress LA (ed) Cerebral mechanisms in behavior. Wiley, New York, NY, pp 112-136

Laureys S, Goldman S, Phillips C, Van Bogaert P, Aerts J, Luxen A, Franck G, Maquet P (1999) Impaired effective cortical connectivity in vegetative state: preliminary investigation using PET. Neuroimage 9:377-382

Laureys S, Celesia G, Cohadon F, Lavrijsen J, Leon-Carrion J, Sannita W, Sazbon L, Schmutzhard E, von Wild KR, Zeman A, Dolce G, European Task Force on Disorders of Consciousness. (2010) Unresponsive wakefulness syndrome: a new name for the vegetative state or apallic syndrome. BMC Med 8:68

Levy DE, Sidtis JJ, Rottenberg DA, Jarden JO, Strother SC, Dhawan V, Ginos JZ, Tramo MJ, Evans AC, Plum F (1987) Differences in cerebral blood flow and glucose utilization in vegetative versus locked-in patients. Ann Neurol 22:673-682

Logothetis NK (2008) What we can do and what we cannot do with fMRI. Nature 453:869-878

Maquet P, Degueldre C, Delfiore G, Aerts J, Peters J, Luxen A, Franck G (1997) Functional neuroanatomy of human slow wave sleep. J Neurosci 17:2807-2812

Maquet P, Peters J, Aerts J, Delfiore G, Degueldre C, Luxen A, Franck G (1996) Functional neuroanatomy of human rapid-eye-movement sleep and dreaming. Nature 383:163-166

Miller GA (2010) Mistreating psychology in the decades of the brain. Persp Psychol Sci 5:716-743

Miyawaki Y, Uchida H, Yamashita O, Sato M, Morito Y, Tanabe H, Sadato N, Kamitani Y (2008) Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. Neuron 60:915-929

Monti M, Laureys S, Owen A (2010) The vegetative state. Clinical Reviews 341:292-296

Nachev P, Husain M (2007) Comment on "Detecting Awareness in the Vegetative State". Science 315:1221

O'Craven KM, Kanwisher N (2000) Mental imagery of faces and places activates corresponding stimulus-specific brain regions. J Cogn Neurosci 12:1013-1023

O'Toole AJ, Jiang F, Abdi H, Haxby JV (2005) Partially distributed representations of objects and faces in ventral temporal cortex. J Cogn Neurosci 17:580-590

Owen AM, Schiff ND, Laureys S (2009) A new era of coma and consciousness science. Progr Brain Res 177:399-411

Papanicolaou AC (1998) Fundamentals of functional brain imaging: a guide to the methods and their applications to psychology and behavioral neurosciences. Swets and Zeitlinger, Lisse, The Netherlands

Papanicolaou AC (2007) What aspects of experience can functional neuroimaging be expected to reveal? Int J Psychophysiol 64:101-105

Phillips C, Bruno A, Maquet P, Boly M, Noirhomme Q, Schnakers C, Vanhaudenhuyse A, Bonjean M, Hustinx R, Moonen G, Luxen A, Laureys S (2010) Relevance vector machine consciousness classifier applied to cerebral metabolism of vegetative and locked-in patients. Neuroimage 53:58-64

Pulvermüller F (1999) Words in the brain's language. Behav Brain Sci 22:253-336

Pulvermüller F (2003) The neuroscience of language: on brain circuits of words and serial order. Cambridge University Press, Cambridge, UK

Raichle ME, Snyder AZ (2007) A default mode of brain function: a brief history of an evolving idea. Neuroimage 37:1083-1090

Ropper AH (2010) Cogito ergo sum by MRI. New Engl J Med 362:648-649

Rudolf J, Ghaemi M, Ghaemi M, Haupt WF, Szelies B, Heiss WD (1999) Cerebral glucose metabolism in acute and persistent vegetative state. J Neurosurgical Anesthesiology 11:17-24

Schiff N, Ribary U, Moreno D, Beattie B, Kronberg E, Blasberg R, Giacino J, McCagg C, Fins JJ, Llinas R, Plum F (2002) Residual cerebral activity and behavioural fragments can remain in the persistently vegetative brain. Brain 125:1210-1234

Shulman GL, Fiez SA, Corbetta M, Buckner RL, Miezin FM, Raichle ME, Petersen SE (1997) Common blood flow changes across visual tasks: II. Decreases in cerebral cortex. J Cogn Neurosci 9:648-663

Vanhaudenhuyse A, Noirhomme Q, Tshibanda L, Bruno M, Boveroux P, Schnakers C, Soddu A, Perlbarg V, Ledoux D, Brichant JF, Moonen G, Maquet P, Greicius MD, Laureys S, Boly M (2010) Default network connectivity reflects the level of consciousness in non-communicative brain-damaged patients. Brain 133:161-171

Vul E, Harris C, Winkielman P, Pashler H (2009) Puzzlingly high correlations in fMRI studies of emotion, personality and social cognition. Persp Psychol Sci 4:274-290