

## Consciousness Monograph

CITATION: De Sousa A. Towards An Integrative Theory Of Consciousness: Part 2 (An Anthology Of Various Other Models). Mens Sana Monogr 2013;11:151-209

# Towards An Integrative Theory Of Consciousness: Part 2 (An Anthology Of Various Other Models)

Avinash De Sousa\*

### ABSTRACT

*The study of consciousness has today moved beyond neurobiology and cognitive models. In the past few years, there has been a surge of research into various newer areas. The present article looks at the non-neurobiological and non-cognitive theories regarding this complex phenomenon, especially ones that self-psychology, self-theory, artificial intelligence, quantum physics, visual cognitive science and philosophy have to offer. Self-psychology has proposed the need to understand the self and its development, and the ramifications of the self for morality and empathy, which will help us understand consciousness better. There have been inroads made from the fields of computer science, machine technology and artificial intelligence, including robotics, into understanding the consciousness of these machines and their implications for human consciousness. These areas are explored. Visual cortex and emotional theories along with their implications are discussed. The phylogeny and evolution of the phenomenon of consciousness is also highlighted, with theories on the emergence of consciousness in fetal and neonatal life. Quantum physics and its insights into the mind, along with the implications of consciousness and physics and their interface are debated. The role of neurophilosophy to understand human consciousness, the*

\*M.D. (Psych Med), D.P.M., M.S. (Psychotherapy and Counselling), D (Clin Psy) (UK), M.B.A. (Human Resource Development), Consultant Psychiatrist and Psychotherapist, Mumbai, India.

Address correspondence to: Dr. Avinash De Sousa, Carmel, 18, St. Francis Avenue (Road), Off S.V. Road, Santacruz West, Mumbai - 400 - 054, Mumbai, India. E-mail: avinashdes999@yahoo.co.uk

Received 6 Mar 2012. Revised 1 May, 28, 30 Sept. Accepted 30 Dec 2012. Further revised 11, 13 Dec 2012. Final Acceptance 13 Dec 2012.

#### Access this article online

##### Quick Response Code:



##### Website:

[www.msmonographs.org](http://www.msmonographs.org)

##### DOI:

10.4103/0973-1229.109341

*functions of such a concept, embodiment, the dark side of consciousness, future research needs and limitations of a scientific theory of consciousness complete the review. The importance and salient features of each theory are discussed along with certain pitfalls, if present. A need for the integration of various theories to understand consciousness from a holistic perspective is stressed.*

**Key Words:** *Artificial Intelligence; Consciousness; Philosophy; Quantum physics; Self*

## Outline of the Article

1. The self and consciousness
  - 1.1 Morality, self-psychology and consciousness
  - 1.2 Spatial issues in bodily self-consciousness
  - 1.3 The somatic marker hypothesis of consciousness
  - 1.4 Consciousness and social cognition
2. Consciousness and artificial intelligence
  - 2.1 Does consciousness have a role in artificial intelligence
  - 2.2 Types of artificial consciousness
  - 2.3 Machine consciousness
  - 2.4 Research arenas in machine consciousness
  - 2.5 Ongoing projects in machine consciousness
  - 2.6 Synthetic phenomenology
  - 2.7 Social, legal and ethical issues in machine consciousness
  - 2.8 Conclusions
3. Miscellaneous facets and approaches to study consciousness
  - 3.1 Visual consciousness
  - 3.2 Computational neuroscience approaches to consciousness
  - 3.3 Emotional consciousness
  - 3.4 Neural models of emotional consciousness
  - 3.5 The phylogeny of consciousness
  - 3.6 The emergence of consciousness in fetal and neonatal life
4. Quantum physics and consciousness
  - 4.1 The interface of physics and consciousness
  - 4.2 Time and the riddle of consciousness
  - 4.3 Electromagnetic theories of consciousness
  - 4.4 Dynamic geometry, brain function and consciousness
  - 4.5 The role of gravity in consciousness
  - 4.6 Three worlds and three mysteries
  - 4.7 The anthropic principle and consciousness
  - 4.8 Quantum physics, computers and neuroscience
  - 4.9 Conclusions

5. Philosophical approaches to consciousness
  - 5.1 General issues in the philosophy of consciousness
  - 5.2 Philosophical approaches to consciousness
  - 5.3 Mind, matter and consciousness in Indian philosophy
  - 5.4 Philosophy of a scientific theory of consciousness
  - 5.5 The function of consciousness
  - 5.6 Embodiment and consciousness
6. The dark side of consciousness
7. The limitations of a scientific theory of consciousness
8. Final conclusions

## **Introduction**

The previous article (Desousa, 2013<sup>[68]</sup>) reviewed studies on consciousness in neurobiology and applied psychology sciences, but the study of consciousness is not restricted to these branches alone. Today we have a new array of theories at our disposal from very distinct fields. Various subspecialties like self-psychology, artificial intelligence, mechanics, quantum physics, computational neuroscience, visual neuroscience and religious studies have contributed in the development of an integrated theory of consciousness. They deserve mention, and the purpose of this review is to provide an overview of these theories, for a better understanding of consciousness.

## **1. The Self and Consciousness**

### **1.1 Morality, self-psychology and consciousness**

To have the status of moral personhood means two things: First, that one has the ability to take responsibility for one's actions, and second, that one ought to be treated in a certain way. Moral personhood, therefore, involves the ability to take responsibility on the one hand and moral rights and obligations on the other. One must distinguish between the 'moral agent' who must be capable of taking responsibility for his or her own action, and the 'moral subject' who has rights and is owed respect (Malloy, 2011<sup>[157]</sup>).

Dennett, in an influential essay (1976)<sup>[73]</sup>, proposed that six conditions must be met to attain moral personhood. First, the entity to whom we would attribute the moral agency must have rationality. Second, we must be able to take the intentional stance towards it, that is, we must be able to attribute states of consciousness or intentions to it. Third, it must be the target of a certain kind of attitude (we have to treat it as a person, for example, with respect or, as the

case may be, hostility). Fourth, it must be capable of reciprocity, and thereby, return that attitude. Fifth, it must be capable of communicating with others.

The second, third, fourth and fifth conditions explicitly and importantly involve social dimensions, although, for Dennett, the precise nature of these social dimensions is still an open question.

Finally, these first five conditions are necessary for the sixth: the entity must be capable of self-consciousness. Self-consciousness is here understood to be a higher order reflective mental process, of which, as Dennett and others (Wilkes, 1988)<sup>[242]</sup>, suggest, young children are incapable. In a variety of other contexts, however, it is suggested that a brain in a vat or a computer might be able to have this kind of self-consciousness (Dennett, 1991<sup>[73]</sup>). This implies that these conditions do not depend on embodiment in any strong sense, and at the same time it raises questions about the social dimensions that are involved in some of these conditions.

If we think of acquiring practical reason, as involving action and the imitation of action, then a recent study on brain imaging has shown that we have a clear neural basis for gaining practical knowledge. Specific brain areas (the pre-frontal, pre-motor areas, the inferior parietal cortex and other areas) have been shown to be activated not only when a subject acts, but also when a subject perceives another person doing an intentional action. These overlapping areas of 'shared neural representations' are also activated when the subject imagines doing an action and when he/she prepares to imitate the action presented by another (Decety and Grézes, 2006<sup>[72]</sup>; Decety and Sommerville, 2003<sup>[71]</sup>). These and similar studies supplement and expand the research on mirror neurons – neurons found in the premotor cortex of the macaque monkey and the human, that are activated both when we perform certain intentional actions (e.g., reaching, grasping) and when we observe others engaging in such actions (Gallese, 2001<sup>[96]</sup>).

Dennett's final condition as mentioned earlier is self-consciousness. Self-consciousness in Dennett's sense, involves the ability to take a second-order volitional attitude towards oneself, as if from the outside, that is, as if I were acting upon another person (Dennett, 1976<sup>[73]</sup>). Phenomenologists suggest that intentional action is always accompanied by a pre-reflective self-consciousness – a self-awareness that is implicit to experience itself. This kind of situated self-consciousness develops within the dimensions defined by primary and secondary inter-subjective interaction, where our motor systems reverberate with the actions of others, and the right or appropriate thing to do is reinforced in narratives that we begin to hear and understand at a very early age (Gallagher, 2006<sup>[93]</sup>; Cohen and Dennett, 2011<sup>[49]</sup>).

What gives self-consciousness its moral significance is its function in moral deliberation. It allows us to stand back from our proposed action and

ask whether this is appropriate or not. It gives us a perspective on ourselves that allows us to deliberate about our planned actions. In contrast to this functional understanding of self-consciousness, Gallagher (1996),<sup>[91]</sup> has argued in a way that suggests that self-consciousness may have intrinsic moral significance.

Bermúdez (2005)<sup>[19]</sup> employs what he terms the ‘principle of derived moral significance’, which states that, ‘if a particular feature or property is deemed to confer moral significance upon a life that has it, then any primitive form of that feature or property will also confer moral significance, although not necessarily to the same degree’. On this basis he argues that a kind of self-consciousness that is something less than the sort described by Dennett must still have moral significance. This minimal form of self-consciousness is characterised by three features: first, a primitive proprioceptive sense of one’s body; second, the capacity to differentiate between self and non-self; and third, a recognition that the other is of the same sort as oneself. Bermúdez cites evidence from experiments on neonatal imitation to show that this sort of self-consciousness can be found in very young infants. Whatever moral significance this minimal self-consciousness has, however, it is not due to the sort of function that Dennett is interested in. Hence, Bermúdez seems to be suggesting that it has some kind of intrinsic moral significance simply because it is a form of self-consciousness (Farhenfort and Lamme, 2012<sup>[84]</sup>).

## 1.2 Spatial issues in bodily self-consciousness

In our daily life body and self are unified at one single location in space. What are the crucial sensory cues the brain takes into account in the creation of this apparently stable and embodied self-representation? Do we localise ourselves according to where we feel our body to be (somato-sensory cues), where we see our body to be (visual cues), or at the origin of our visual perspective? The empirical study of bodily self-consciousness has proven difficult, because the body is always there (James, 1890<sup>[131]</sup>) and is never a discrete object of perception.

Data from neurological patients may be useful here, as neurological interference enables the study of instances of spatially dissociated bodily representations (Blanke *et al.*, 2002<sup>[22]</sup>). In certain pathological conditions, as during an out-of-body experience, the self can be localised at the origin of the visual perspective, even though this location is different from the location seen of one’s body (Blanke *et al.*, 2004<sup>[23]</sup>). In other neurological cases, the self can be experienced as being at the location of the felt body, although this location does not correspond with that of the origin of visual perspective or the seen body (De Ridder *et al.*, 2007<sup>[66]</sup>). Furthermore, patients with heautoscopy may experience two rapidly alternating perspectives, often leaving them confused about where their self is localised (Blanke *et al.*, 2005<sup>[24]</sup>).

Systematic studies were necessary due to the small sample sizes of the clinical studies, the difficulties in generalising these findings to normal functions and other methodological concerns. By exposing participants to conflicting multi-sensory cues by means of mirrors or simple virtual reality devices, these authors developed experimental strategies to manipulate the spatial unity between body and self in healthy subjects (Altschuler and Ramachandran, 2007<sup>[7]</sup>; Ehrsson, 2007<sup>[80]</sup>; Lenggenhager *et al.*, 2007<sup>[147]</sup>).

Recent philosophical and neurological theories converge on the relevance of bodily processes in self-consciousness (Gallagher, 2005<sup>[92]</sup>). Studying the role of various bodily cues in self-representations, in a rigorous scientific set-up, is important to further evolve these theories. Researchers have investigated where participants experience and localise their self, given conflicting information about the seen and the felt body, as well as the visual perspective. The data suggest that participants localise their 'self' where they perceive to be touched, even if this tactile perception is mislocalised through visual capture leading in the present setup to predictable up- or downward shifts in self-location. The former is associated with a feeling of floating, as typically found in neurologically caused cases of disturbed self-location. Disentangling the contribution of different bodily cues to self-location may help to better understand normal and abnormal embodiment and self-consciousness (Zahavi, 2005<sup>[245]</sup>; Di Francesco, 2008<sup>[77]</sup>).

### 1.3 The somatic marker hypothesis of consciousness

Researchers have proposed that background feelings are essential for the emergence of core consciousness or the sense of self, which is generated moment to moment in a pulse-like fashion, reflecting the ongoing interaction of the human being with the environment and the minute-to-minute changes in homeostatic body states (Damasio, 2000<sup>[60]</sup>; Damasio, 2012<sup>[61]</sup>). Subjective perception of the body state requires right insula activity (Craig, 2009<sup>[55]</sup>), and activation of the anterior cingulate cortex follows for the motivational component that accompanies feelings (Rushworth and Behrens, 2008<sup>[203]</sup>).

Extended consciousness refers to that same moment-to-moment self, extended by connections to both past experiences and an anticipated future creating an autobiographical memory. The simultaneous holding of images from autobiographical memory and images of objects, for a substantial amount of time, results in a unified experience of knowing (Cabeza and Jacques, 2007<sup>[34]</sup>). Another model that is similar is the comparator model (Gray, 1995<sup>[107]</sup>). This holds that the contents of the consciousness generated correspond with the outputs of a comparator that, on a moment-to-moment basis, compare the current state of the world with a predicted state, internally generated from the past world inputs, on the basis of planning or predictor systems. The area devoted to predictions is conjectured to be the hippocampus, which, along with

the amygdale, plays an important role in reward memory for inputs (Le Doux, 2007<sup>[146]</sup>; Bird and Burgess, 2008<sup>[21]</sup>).

#### 1.4 Consciousness and social cognition

Consciousness and social cognition are connected in two ways. First, social cognition is responsible for an important slice of our conscious experience. Second, the ability to represent the conscious experience of other people, a particular facet of social cognition, is responsible for some of the most important features of our social functioning. In describing the social character of consciousness, we must note that consciousness is to some extent a social phenomenon. Although each individual has his own distinctive point of view on the world, a good deal of the content of individual experience is picked up from his/her contact with others. This tendency of consciousness to spread from person to person appears to play a causal role in phenomenal mind reading, the process by which we understand the conscious states of others (Marsh and Robbins, 2008<sup>[159]</sup>; Winkielman and Schooler, 2011<sup>[243]</sup>).

Probably the best evidence to this effect comes from the research on face-based emotion recognition (Eimer and Holmes, 2007<sup>[81]</sup>). There is a whole crop of neuropsychological studies on individuals who have lost the ability to recognise a specific emotion (e.g., fear or disgust) as well as the capacity to experience that emotion, but who can recognise and experience other emotions normally (Goldman, 2006<sup>[104]</sup>). A natural way to explain these cases of selective paired deficits in emotion, recognition and experience is according to the 'mental resonance' model of phenomenal (low-level) mind reading: We recognise others' emotions by mirroring their state, then introspectively classifying the result of that mirroring, and finally attributing the introspected state to the target. On this model, mechanisms of affective contagion are part of the machinery of mind reading. This model is consistent with clinical studies of autism, which shows deficits in both emotion recognition and emotional mimicry (Panksepp, 1998<sup>[175]</sup>; McIntosh *et al.*, 2006<sup>[163]</sup>).

In terms of thinking about the relation between consciousness and the social mind, this literally brings us full circle. We know that consciousness is shaped by the social mind, in virtue of phenomena like social pain and affective contagion. We know that the ability to represent the conscious states of others contributes something vital to social competence. Finally, we have just seen that this representational ability may turn on the ability to mirror the conscious states of others in oneself. Hence, just as consciousness depends on the wirings of the social mind, social mindedness may depend on the wirings of consciousness (Turiel 1983<sup>[236]</sup>; Singer, 2006<sup>[215]</sup>; Robbins and Jack, 2006<sup>[192]</sup>; Amadio and Ratner, 2011<sup>[9]</sup>).

## 2. Consciousness and Artificial Intelligence

### 2.1 Does consciousness have a role in artificial intelligence?

Consciousness is considered that part of the brain that is least amenable or reproducible by artificial intelligence (AI). It is assumed that the very nature and essence of consciousness may not be explainable by computations, algorithms and processes or the function methods of AI (Chrisley, 2008<sup>[46]</sup>).

First, consciousness and intelligence are not so clearly distinguishable. For example, in most cases where we would say that a task requires intelligence, we would also say that it requires consciousness. Second, the field of AI, in its broadest sense, is poorly served by the name 'artificial intelligence'. This hides the fact that despite an early emphasis on problem solving, the field has always had more than just intelligence in its sight. AI is an attempt to create artefacts that have mental properties or exhibit characteristic aspects of systems that have such properties, and such properties include not just intelligence, but also perception, action, emotion, creativity and consciousness. In this sense, artificial consciousness (AC) is a subfield of AI (Antonov, 2011<sup>[10]</sup>).

Some AC researchers in robot navigation and planning are as concerned with exploring the extent to which the processes the robot employs can be usefully viewed as instances of imagination and the existence of an inner world, as they are actually getting the robot to avoid collisions, find its way to a goal location and so on (Hesslow 2002<sup>[116]</sup>; Holland and Goodman, 2003<sup>[119]</sup>; Hesslow 2007<sup>[117]</sup>). Other AC research is concerned with the phenomenological aspects of imaginative and counterfactual reasoning includes Aleksander, 2000<sup>[3]</sup>; Carruthers, 2000<sup>[39]</sup>; Shahnanan, 2006<sup>[210]</sup>; Chrisley and Parthermore, 2007<sup>[45]</sup>.

### 2.2 Types of artificial consciousness

Engineering AC is primarily concerned with creating artefacts that can do things that previously only naturally conscious agents could do; whether or not such artificial systems perform these functions or behaviour in the way that natural systems do is not considered a matter of primary importance. Of central concern are the functional capabilities of the developed technology: What functional benefits can accrue from making a system behave more like a conscious organism? Whether or not the system so developed is really conscious is not an issue. Scientific AC, however, is primarily concerned with understanding the processes underlying consciousness, and the technologies provided by engineering AC, however impressive, are only considered of theoretical relevance to the extent that they resemble or otherwise illuminate the processes underlying consciousness (Searle, 1980<sup>[207]</sup>).



Weak AC is any approach that makes no claim of a relation between technology and consciousness. This would be the use of technology for understanding consciousness in a way similar to the use of computational simulations of hurricanes in meteorology, that is, understanding can be facilitated by such simulations, but no one supposes that this is because hurricanes are themselves computational in any substantive sense. At the other extreme, Strong AC is any approach whose ultimate goal is the design of systems that, when implemented, are thereby instantiations of consciousness. Between these two extremes is a neglected zone of possibility that might be termed Lagom AC. ‘Lagom’ is a Swedish word with no direct English equivalent, which means something like ‘perfection through moderation’. The Lagom AC view, unlike weak AC, claims that the modelling relation holds as a result of deeper, explanatory properties being shared by the technology and conscious mental phenomena. However, unlike Strong AC, Lagom AC does not go on to claim that instantiating these common properties is alone sufficient for instantiating consciousness; something else might be required. Unlike Lagom AC, the necessity involved is not a constitutive matter of what properties the artificial agent must have for it to be conscious, but rather a practical matter of what tools and concepts we must have to be able to build it (Phillips, 2011<sup>[183]</sup>).

These practical requirements are themselves of two types: Causal and conceptual. Causal requirements have to do with the kinds of software, hardware, user interfaces and so on that we will need to help us achieve AC. No doubt sophisticated, intelligent, computational technology not yet invented will be needed to help us collect, mine and process the enormous quantities of data we can anticipate to acquire over the next decades, with regard to the operations of the brain and body that underlie consciousness. Similar advances in technological AI will also likely be needed to assist in the design of any system complex enough to be a candidate for artificial consciousness. Conceptual requirements have to do with the types of systems we will need to have the experience of building, and the types of learning/creativity/perceptual/performance enhancing technologies we will need to develop, in order to get ourselves into the right conceptual/knowledge state for achieving AC (Chrisley, 2000<sup>[44]</sup>; Clarks and Chalmers, 1998<sup>[48]</sup>).

Most AC research is autonomous, in that it aims to create a self-standing, individual artificial consciousness. Much less frequently discussed is the possibility of prosthetic AC: ‘Artificial consciousness’ is a phrase parallel in construction to ‘artificial leg’ rather than ‘artificial light’. Prosthetic AC would seek to extend, alter or augment the already existing consciousness rather than create it anew. It is often a misunderstanding to think that creating or discovering more instances of a phenomenon to be explained only increases the problem; rather, the consideration of new instances is a way to increase the robustness of one’s models and theories (Aleksander, 2000<sup>[3]</sup>; Aleksander and Gamez, 2011<sup>[6]</sup>).

There are two subtypes of a prosthetic AC, that is, *conservative* and radical. The former seeks to create alternative material bases for extant kinds of experience; the latter seeks to create alternative material bases that result in novel kinds of experience (Chrisley, 2000<sup>[44]</sup>; Hurley, 1998<sup>[125]</sup>; Dennett, 1987<sup>[74]</sup>).

An important distinction within AC has to do with whether one is attempting to reproduce/explain human (or biological) consciousness in particular, or whether one is attempting to reproduce/explain consciousness more generally. The quest for generality can either be more ambitious or more modest. For example, in the case of constitutive scientific AC, generality would be more ambitious if one were attempting to explain, for any possible conscious agent, why that agent is conscious. A more modest form of generality would be merely attempting to explain how it is that some particular non-human (non-biological) physical thing is also a conscious thing. Even if one's ultimate goal is to explain human consciousness, it is important to understand how knowledge of the functioning at a mechanical level of some particular artificial system can help us understand what it feels like to be conscious (Dennett, 1991<sup>[75]</sup>; Searle, 1992<sup>[208]</sup>; Penrose, 1994<sup>[180]</sup>; Chalmers, 1996<sup>[41]</sup>; Sloman and Chrisley, 2003<sup>[219]</sup>; Chrisley, 2008<sup>[46]</sup>).

## 2.3 Machine consciousness

Over the last ten years, there has been a resurgence of interest in human consciousness and a large number of philosophers, psychologists and neuroscientists are now working on this area. Researchers have also started to test theories of consciousness using computer models and there has been some speculation that this could eventually lead to more intelligent machines that might actually have phenomenal states. This type of research is gradually becoming known as 'machine consciousness', although 'artificial consciousness' and occasionally 'digital sentience' have also been used to describe it (Aleksander, 2005<sup>[4]</sup>). Each of these terms has their own merits, but 'machine consciousness' has become the standard name for the field.

Machine consciousness is currently a heterogeneous research area that includes a number of different research programs. For example, some people are working on behaviours associated with consciousness, others are modelling the cognitive characteristics of consciousness and some others are interested in creating phenomenal states in machines (Starzyk and Prasad, 2010<sup>[223]</sup>). To make sense of this diverse subject, the first part of this article identifies four different areas of machine consciousness research, namely (Gamez, 2008<sup>[98]</sup>):

- Machines with the external behaviour associated with consciousness.
- Machines with the cognitive characteristics associated with consciousness.
- Machines with an architecture that is a correlate of human consciousness.
- Phenomenally conscious machines.

This classification starts with systems that replicate aspects of human behaviour and moves on to systems that are attempting to create real artificial consciousness.

## 2.4 Research arenas in machine consciousness

One research arena in machine consciousness is on systems that replicate conscious human behaviour. Although this type of research can be based on cognitive models or on architecture associated with consciousness, it is not necessary that it works on machines, which can also use large lookup tables or first-order logic to generate the behaviour. Although certain external behaviours are associated with phenomenal states in humans, this is not necessarily important to people working on machine consciousness, as it has often been claimed that a zombie robot could replicate conscious human behaviour without experiencing the phenomenal states (Ito, Miyashita and Rolls, 1997<sup>[130]</sup>; Wallach *et al.*, 2011<sup>[240]</sup>).

The modelling of cognitive characteristics associated with consciousness has been a strong theme in machine consciousness, where it has been carried out in a wide variety of ways, ranging from simple computer programs to systems based on simulated neurons. Cognitive characteristics that are frequently covered by this work include imagination, emotions, global workspace architecture and internal models of the system's body and environment. In some cases the modelling of cognitive states has aimed at more realistic conscious behaviour or used an architecture associated with consciousness (Franklin *et al.* 2005<sup>[87]</sup>; Wallach *et al.*, 2011<sup>[240]</sup>).

Many researchers are working on the simulation of architectures that have been linked to human consciousness, such as global workspace (Baars, 1998<sup>[13]</sup>), neural synchronisation (Crick, 1994<sup>[57]</sup>) or systems with high information integration (Tononi, 2004<sup>[234]</sup>). This type of research often arises from the desire to model and test neural or cognitive theories of consciousness, and it is one of the most characteristic areas of machine consciousness.

The first three approaches to machine consciousness (described above) are all relatively uncontroversial, as they are modelling phenomena linked to consciousness without any claims about real phenomenal states. The fourth area of machine consciousness is more problematic philosophically, as it is concerned with machines that have real phenomenal experiences – machines that are not just tools in consciousness research, but actually conscious themselves. In some cases it may be hypothesised that the reproduction of human behaviour, cognitive states, or internal architecture leads to real phenomenal experiences. It may also be possible to create a system based on biological neurons that is capable of phenomenal states, but lacks the architecture of human consciousness and any of its associated cognitive states or behaviours (Duch, 2005<sup>[78]</sup>).

Furthermore, it has been claimed that even thermostats may have simple conscious states. If this is correct, the presence of phenomenal states in a machine will be largely independent of the higher level functions that it is carrying out (Chalmers, 1996<sup>[41]</sup>).

Chalmers (1996)<sup>[41]</sup> distinguishes between the easy problem of explaining how we can discriminate, integrate information, report mental states, focus attention and so on and the difficult problem of explaining phenomenal experience. Although many theories have been put forward about the difficult problem, it can be argued that we have no real idea about how to solve it, and if we do not understand how human consciousness is produced, then it makes little sense to attempt to make a robot phenomenally conscious (Levy, 2011<sup>[149]</sup>).

There are a number of reasons why the difficult problem of consciousness may not be devastating for the work on machine consciousness. To begin with, it can be argued that asking questions about phenomenal consciousness in machines and building models can improve our understanding of human consciousness and take us closer to a solution to the difficult problem. Second, there are the arguments that suggest that it may be indeterminable whether a machine is conscious or not (Holland, 2003<sup>[120]</sup>). This may force us to acknowledge the possibility of consciousness in a machine, even if we cannot tell for certain whether this is the case, by solving the difficult problem of consciousness. Third, it may be possible to create conditions that allow consciousness to emerge in a system, without understanding the causes of the phenomenal states (Cotterill, 2000<sup>[51]</sup>). Finally, the future replacement of brain parts with silicon will force us to tackle machine consciousness in humans, even if we abandon the study on this area in machines.

Machine consciousness has also been criticised by researchers who claim that the processing of an algorithm is not enough to evoke phenomenal awareness, because subtle and largely unknown physical principles are needed to perform the non-computational actions that lie at the root of consciousness (Penrose 1990<sup>[179]</sup>, 1994<sup>[180]</sup>, 2005<sup>[181]</sup>, 2010<sup>[182]</sup>). Electronic computers have their undoubted importance in clarifying many of the issues that relate to mental phenomena (perhaps, to a large extent, by teaching us what genuine mental phenomena are not). Computers, we conclude, do something very different from what *we* are doing when we bring our awareness to bear upon some problem.

The most straightforward response to Penrose is to reject his theory of consciousness, which is far from convincing and has been heavily criticised by some (Grush and Churchland, 1995<sup>[109]</sup>).

Some authors have developed an approach to machine consciousness based around five axioms, which they believe are minimally necessary for consciousness, namely (Aleksander, 2005<sup>[4]</sup>):

- *Depiction*: The system has perceptual states that ‘represent’ elements of the world and their location.
- *Imagination*: The system can recall parts of the world or create sensations that are like parts of the world.
- *Attention*: The system is capable of selecting which part of the world to depict or imagine.
- *Planning*: The system has control over sequences of states to plan actions.
- *Emotion*: The system has affective states that evaluate the planned actions and determine the ensuing action.

These axioms link cognitive attributes, such as imagination and emotions, to phenomenal consciousness. Aleksander is careful to state that this is a preliminary list of mechanisms that could make a system conscious, which must be revised as our knowledge of consciousness develops – a useful starting point that can be used to test ideas and develop the field. These axioms have been incorporated by him into a kernel architecture, which includes a perceptual module that depicts sensory input, a memory module that implements non-perceptual thought for planning and recall of experience, an emotion module that evaluates the ‘thoughts’ in the memory module and an action module that causes the best plan to be carried out.

## 2.5 Some interesting projects in machine consciousness

CRONOS is one of the few large projects that has been explicitly funded to work on machine consciousness. It consists of CRONOS, a hardware robot, closely based on the human musculoskeletal system, SIMNOS, a soft, real-time, physics-based simulation of this robot in its environment, a biologically inspired visual system and a spiking neural simulator called Spike Stream. The main focus of this project is on the cognitive, architectural and phenomenal aspects of machine consciousness.

Cog was a humanoid robot developed that consisted of a torso, head and arms under the control of a heterogeneous network of programs written in L, a multi-threaded version of Lisp (Nehaniv, 1998<sup>[171]</sup>). Cog was equipped with four cameras providing stereo foveated vision, microphones on each side of its head and a number of piezoelectric touch sensors. This robot also had a simple emotional system to guide learning and a number of hard-wired, ‘innate’ reflexes, which formed a starting point for the acquisition of more complex behaviours. The processors controlling Cog were organised into a control hierarchy, ranging from small microcontrollers for joint-level control to digital signal processor networks for audio and visual processing.

CyberChild is a simulated infant controlled by a biologically-inspired neural system, based on a particular theory of consciousness (Cotterill, 2000<sup>[51]</sup>). This virtual infant has rudimentary muscles controlling the voice and limbs,

a stomach, a bladder, pain receptors, touch receptors, sound receptors and muscle spindles. It also has a blood glucose measurement, which is depleted by energy expenditure and increased by consuming milk. As the consumed milk is metabolised, it is converted into simulated urine, which accumulates in the infant's bladder and increases its discomfort level. The simulated infant is deemed to have died when its blood glucose level reaches zero. CyberChild also has drives that direct it towards acquiring sustenance and avoiding discomfort and it is able to raise a feeding bottle to its mouth and control urination by tensing its bladder muscle. However, these mechanisms are not enough on their own to ensure the survival of the simulated infant, which ultimately depends on its ability to communicate its state to a human operator.

Other researchers are developing a system that is intended to display cognitive characteristics associated with consciousness, such as emotion, transparency, imagination and inner speech, using detailed neural simulation. This cognitive architecture starts with sensory modules that process visual, auditory and tactile information into a large number of on/off signals that carry information about different features of the stimulus (Haikonen, 2003<sup>[110]</sup>). Perceived entities are represented using combinations of these signals, which are transmitted by modulating a carrier signal. There is extensive feedback within the system, and cross-connections between different sensory modalities integrate the qualitative characteristics carried by the signal with its location in motor space. Haikonen's architecture also includes emotions, for example, there is an analogue of pain, which uses information about physical damage to initiate withdrawal and redirect attention. In this architecture, language is part of the auditory system and the association of words with representations from other modalities enables sequences of the percepts to be linguistically described. He claims that the percepts become conscious when different modules cooperate in unison and focus on the same entity, which involves a wealth of cross-connections and the forming of associative memories (Haikonen, 2006<sup>[111]</sup>).

Cicerobot is a robot that has sonar, a laser rangefinder and a video camera, and works as a museum tour guide in the Archaeological Museum of Agrigento. The cognitive architecture of this robot is based around an internal 3D simulation, which is updated as the robot navigates around its environment. When the robot moves it sends a copy of its motor commands to the 3D simulator, which calculates expectations about the next location and camera image. Once the movement has been executed, the robot compares its expected image with the 2D output from its camera and uses discrepancies between the real and expected images to update its 3D model. Cicerobot uses this 3D simulation to plan actions, by exploring different scenarios in a way that is analogous to human imagination (Chella and Macaluso, 2006<sup>[42]</sup>).

Other studies on machine consciousness include scientists who have used physics, computer science and information theory to outline how

consciousness and a conscious self-model may be implemented in a machine (Mulhauser, 1998<sup>[169]</sup>).

## 2.6 Synthetic phenomenology

Synthetic phenomenology is a new area of research that has emerged out of the studies on machine consciousness. This term was first coined to refer to the synthesising of phenomenal states (Jordan, 1998<sup>[133]</sup>). Within the machine consciousness community, ‘synthetic phenomenology’ is now more generally used to refer to the determination of whether artificial systems are capable of conscious states and the description of their phenomenology if and when this occurs, and it is in this sense that I will be using it here. It is also related to synthetic epistemology, which is defined as the creation and analysis of artificial systems in order to clarify philosophical issues that arise in the explanation of how agents, both natural and artificial, represent the world (Chrisley and Holland, 1994<sup>[43]</sup>). One of history’s phenomenological projects was the description of human consciousness; the synthetic phenomenological project is the description of machine consciousness – a way in which people working on machine consciousness can measure the extent to which they have succeeded in realising consciousness in a machine (Husserl, 1960<sup>[126]</sup>; 1964<sup>[127]</sup>; Damasio, 2012<sup>[61]</sup>; Arico *et al.*, 2011<sup>[11]</sup>).

It is impossible to describe the phenomenology of a system that is not capable of consciousness, and so the first challenge faced by synthetic phenomenology is to identify the systems that are capable of phenomenal states. One approach to this problem is to use a theory of consciousness to distinguish between systems that are and are not phenomenological. For example, there have been two proposed axioms that a system must conform to if it is to be a candidate for synthetic phenomenology. To be synthetically phenomenological, a system *S* must contain machinery that represents what the world and the system *S* within it seems like, from the point of view of *S* (Aleksander and Morton, 2008<sup>[5]</sup>). An unpacked version of this definition is used by Aleksander and Morton to argue that their kernel architecture is synthetically phenomenological, whereas, the global workspace architecture is not.

Synthetic phenomenology has a number of overlaps with the description of human phenomenology from a third-person perspective. This type of research is commonly called ‘neurophenomenology,’ although this term is subject to two conflicting interpretations. The first interpretation of neurophenomenology has been put forward by scientists who have used it to describe a reciprocal dialogue between the accounts of the mind offered by science and phenomenology. This type of neurophenomenology emphasises the first person human perspective and has little in common with synthetic phenomenology (Varela, 1996<sup>[238]</sup>). However, neurophenomenology can also be interpreted as the description of human phenomenology from a third person perspective, using measurements of



brain activity gathered using techniques, such as Functional magnetic resonance imaging (fMRI), Electroencephalography (EEG) or electrodes. A good example of this type of study is a research that used the patterns of intensity in fMRI voxels to make predictions about the phenomenal states of their subjects (Kamitani and Tong, 2005<sup>[135]</sup>). In some ways neurophenomenology is easier than synthetic phenomenology because it does not have to decide whether its subjects are capable of consciousness, and the description of non-conceptual states is considerably easier in humans. However, both disciplines are attempting to use external data to identify the phenomenal states in a system, and there is considerable potential for a future collaboration between them (Aleksander and Gamez, 2011<sup>[6]</sup>).

## 2.7 Social, legal and ethical issues in machine consciousness

Many people believe that studies on machine consciousness will eventually lead to machines taking over and enslaving humans in a Terminator- or Matrix-style future world. This is the position of authors who believe that we will increasingly pass responsibility to intelligent machines until we are unable to do without them – in the same way that we are increasingly unable to live without the Internet today. This will eventually leave us at the mercy of potentially super-intelligent machines that may use their power against us (Buttazzo, 2008<sup>[30]</sup>).

Many argue that it is very unlikely that intelligent machines can possibly produce more dreadful behaviour towards humans than humans already produce towards each other, all round the world, even in the supposedly most civilised and advanced countries, both at individual levels and at social or national levels (Goertzel and Pennachin, 2007<sup>[105]</sup>).

At present our machines fall far short of many aspects of human intelligence, and we may have hundreds of years to consider the matter before either the apocalyptic or optimistic scenarios come to pass. It is also the case that science fiction predictions tell us more about our present concerns than about a future that is likely to happen, and our attitudes towards ourselves and machines will change substantially over the next century, as they have changed over the last. As machines become more human and humans become more mechanical, the barriers will increasingly break down between them until the notion of a takeover by machines makes little sense (Kurzweil, 2000<sup>[140]</sup>).

A second ethical dimension to study machine consciousness is to learn how we should treat conscious machines. We will eventually be able to build systems that are not just instruments for us, but participants with us in our social existence. However, this can only be done through experiments that cause conscious machines a considerable amount of confusion and pain, which has led to comparing studies on machine consciousness to the development of a race of retarded infants for experimentation (Metzinger, 2003<sup>[165]</sup>).



A final aspect of the social and ethical issues surrounding machine consciousness is the legal status of conscious machines. When traditional software fails, responsibility is usually allocated to the people who developed it, but the case is much less clear with autonomous systems that learn from their environment. A conscious machine may malfunction because it has been maltreated, and not because it is badly designed, and so its behaviour can be blamed on its caretakers or owners, rather than on its manufacturers. Conscious machines can also be held responsible for their own actions and punished appropriately (Calverley, 2008<sup>[35]</sup>).

## 2.8 Conclusions of this section

Machine consciousness is a relatively new research area that has gained considerable momentum over the last few years, and there are a growing number of research projects in this field. Although it shares some common ground with philosophy, psychology, neuroscience, computer science and even physics, machine consciousness is rapidly developing an identity, and problems, of its own. The benefits of machine consciousness are only starting to be realised, but studies on machine consciousness are already proving to be a promising way of producing more intelligent machines, testing theories about consciousness and cognition and deepening our understanding of consciousness in the brain. As machine consciousness matures it is also starting to raise some novel social and ethical issues.

## 3. Miscellaneous Facets and Approaches to the Study of Consciousness

### 3.1 The concept of visual consciousness

The visual brain consists of several parallel, functionally specialised processing systems, each having several stages (nodes) that terminate their tasks at different times; consequently, simultaneously presented attributes are perceived at the same time, if processed by the same node, and at different times if processed by different nodes (Von der Hydt, 1987<sup>[239]</sup>). Clinical evidence shows that these processing systems can act fairly autonomously. Damage restricted to one system specifically compromises the perception of the attribute that that system is specialised for; damage to a given node of a processing system that leaves the earlier nodes intact results in a degraded perceptual capacity for the relevant attribute, which is directly related to the physiological capacities of the cells left intact by the damage. By contrast, a system that is spared when all others are damaged can function more or less normally (Tootell and Taylor, 1995<sup>[232]</sup>).

Moreover, internally created visual percepts—illusions, after images, imagery and hallucinations specifically activate the nodes specialised for

the attribute perceived. Finally, anatomical evidence shows that there is no final integrator station in the brain, one which receives the input from all visual areas; instead, each node has multiple outputs and no node is only a recipient. Taken together, the above evidence would lead us to propose that each node of a processing-perceptual system creates its own micro-consciousness (Logothetis, 1998<sup>[154]</sup>). It has also been proposed that, if any binding occurs to give us our integrated image of the visual world, it must be a binding between the micro-consciousnesses generated at different nodes. As any two micro-consciousnesses generated at any two nodes can be bound together, perceptual integration is not hierarchical, but parallel and post-conscious. By contrast, the neural machinery conferring properties on those cells whose activity has a conscious correlate is hierarchical, and is referred to as generative binding, to distinguish it from the binding that may occur between the micro-consciousnesses (Zeki and Bartels, 1999<sup>[250]</sup>).

Researchers have put forward a few propositions while moving towards an integrated theory of visual consciousness. They are:

- The visual brain consists of parallel distributed and functionally specialised processing systems (Zeki, 1993<sup>[248]</sup>).
- Forward connections within a processing system of like with like variety leads to cells of increasing receptive field size, follow a hierarchal pattern (Zeki, 1978<sup>[246]</sup>).
- The lateral interconnections that anatomically link different processing systems can be like with like, like with unlike and diffuse variety, and they are not exclusively hierarchal (Bartels and Zeki, 2000<sup>[17]</sup>).
- There is no terminal station in the cortex for a given processing system and there is no common terminal area to which different processing systems connect (Zeki and Bartels, 2003<sup>[252]</sup>).
- Generative or preconscious integration is hierarchal and limited to a given processing system (Zeki, 2003<sup>[253]</sup>).
- Parallel integration is non-hierarchal and can occur between the nodes of different processing systems as well within a single node (Zeki, 2004<sup>[254]</sup>).
- The visual cortex possesses the property of temporal asynchrony and visual perception is modular in nature (Zeki and Marini, 1998<sup>[249]</sup>).
- When two visual events occur together, they need not be integrated for each to be perceived and mutual integration is not a must for conscious perception (Zeki, 1999<sup>[251]</sup>).
- Activities in each node possess a micro-consciousness of their own. Thus we have several micro-consciousnesses, as there are different nodes within different processing systems (Zeki, 2009<sup>[255]</sup>).
- Two visual events occur at the same time and are perceived at the same time because they are processed at the same site. When two visual events occur at the same time and are perceived at different times, it is because they are processed at two different sites (Tong *et al.*, 1998<sup>[233]</sup>).
- Micro-consciousness of a functionally specialised area is a reflection of

the activity in a specialised specific processing node (Arnold and Clifford, 2002<sup>[12]</sup>).

- Damage to the prestriate component of a specific processing system does not lead to the total loss of the relevant visual faculty and some residual functional form of that visual faculty is always present in the subject (Zeki, 1990<sup>[247]</sup>).
- Activity at the node of a processing system can have a conscious orientation even in the presence of a lesion, provided there is some input to that node (Kawabata and Zeki, 2004<sup>[136]</sup>).
- Processing and perceptual sites in the visual cortex are generally one and the same (Moutoussis and Zeki, 2002<sup>[167]</sup>).
- Processing systems are autonomous in function (Gallagher and Frith, 2003<sup>[95]</sup>).

The propositions that have been given herewith form a chain, which leads us towards our theory of visual consciousness. Some have been proven beyond doubt, while others are under research study. However, all of them are so consistent with each other and with the known facts that, when considered as a whole, they are able to lead us towards a theory of visual consciousness, which might be applicable to other parts of the brain (Treisman, 1996<sup>[235]</sup>).

Visual consciousness consists of many, functionally specialised, micro-consciousnesses that are spatially and temporally distributed if they are the result of activity at spatially distributed sites. This is a direct consequence of the fact that the several, parallel, multi-nodal, functionally specialised and autonomous processing systems are also perceptual ones, and that the activity at each node of each processing-perceptual system can become perceptually explicit (Hubel and Weisel, 1977<sup>[122]</sup>; Prinz, 2000<sup>[187]</sup>).

Activity at each node, therefore, has a micro-conscious correlate that is functionally specialised and asynchronous with the micro-conscious correlate generated by that at other nodes. If integration occurs between different nodes, the communication between them must influence the micro-consciousness that each creates in a consistent way, leading to consistent, integrated percepts. It is, therefore, not surprising that there is no terminal station in the cortex, as activity at each node represents, in a sense, a terminal stage of its own specialised process, when it becomes perceptually explicit and acquires a conscious correlate. This leaves us with the grand problem of how, in physiological terms, the micro-consciousnesses are bound together. Indeed, it raises the question of whether they are bound at all, given what appears to be the non-unitary nature of conscious experience (LaRock, 2007<sup>[145]</sup>; Cardin *et al.*, 2011<sup>[36]</sup>; Fftyche and Zeki, 2011<sup>[85]</sup>).

### 3.2 Computational neuroscience studies and consciousness

The theory described by computational neuroscience suggests that it feels like something to be an organism or machine that can think about its own thoughts.

It is suggested that qualia, raw sensory and emotional subjective feelings arise secondary to having evolved into a higher order of thought system, and that sensory and emotional processing feels like something because it would be unparsimonious for it to enter the planning and higher order thought system and not feel like something (Matzke, 2010<sup>[160]</sup>). Raw sensory feels and subjective states associated with emotional and motivational states, may not necessarily arise first in evolution (Rolls, 2007<sup>[196]</sup>).

Some issues that arise in relation to this theory, such as, reasons why the ventral visual system is more closely related to the explicit than implicit processing (because reasoning about objects may be important) and why explicit, conscious processing may have a higher threshold in sensory processing than implicit processing, have been dealt with by researchers (Rolls, 2000<sup>[193]</sup>; 2004<sup>[194]</sup>; 2005<sup>[195]</sup>).

This theory explains what the underlying computational problem is (how syntactic operations are performed in the system, and how they are corrected), and argues that when there are thoughts about the system, that is, higher order syntactic thoughts (HOSTs) and the system is reflecting on its first-order thoughts (Weiskrantz, 1997<sup>[241]</sup>), then it is a property of the system that it feels conscious. The theory is also different from some other theories of consciousness (Carruthers, 1996<sup>[38]</sup>; Gennaro, 2004<sup>[102]</sup>; Rosenthal, 2005<sup>[200]</sup>), in that, it provides an account of the evolutionary, adaptive value of a higher order thought system in helping to solve a credit assignment problem that arises in a multistep syntactic plan, links this type of processing to consciousness, and therefore, emphasises a role for syntactic processing in consciousness.

The current theory holds that it is the HOSTs that are closely associated with consciousness, and this may differ from Rosenthal's higher order thoughts (HOTs) theory (Rosenthal, 1986<sup>[199]</sup>; 2005<sup>[200]</sup>), with the emphasis in the current theory on language. Language in the current theory is defined by syntactic manipulation of symbols, and does not necessarily imply verbal or 'natural' language.

Very complex mappings in a multilayer network can be learnt if hundreds of learning trials are provided. However, once these complex mappings are learnt, their success or failure in a new situation on a given trial cannot be evaluated and corrected by the network. Indeed, the complex mappings achieved by such networks (e.g., back propagation nets) mean that after training they operate according to fixed rules, and are often quite impenetrable and inflexible (Rolls and Deco, 2002<sup>[198]</sup>).

### 3.3 The search for emotional consciousness

Consciousness and emotion feature prominently in our personal lives. Emotion consists of an emotion state (functional aspects, including emotional response) as

well as feelings (the conscious experience of the emotion), and that consciousness consists of levels (coma, vegetative state and wakefulness) and content (what it is we are conscious of). Not only is consciousness important to aspects of emotion, but also the structures that are important for emotion, such as brainstem nuclei and midline cortices, overlap with structures that regulate the level of consciousness. Detailed theoretical accounts of emotion (Russell, 2003<sup>[205]</sup>; Prinz, 2004<sup>[189]</sup>; Rolls, 2005<sup>[195]</sup>; Panksepp, 2005<sup>[176]</sup>) and consciousness (Chalmers, 1996<sup>[41]</sup>; Koch, 2004<sup>[137]</sup>) together with recent data from functional imaging (Dalgeish, 2004<sup>[58]</sup>) and clinical populations (Russell, 1991<sup>[204]</sup>), provide an unprecedented opportunity for the progress on these topics. Research in emotional consciousness is based on the idea that affective processes are supported by brain structures that have appeared earlier on the phylogenetic scale (such as the periaqueductal grey area), and they run in parallel with cognitive processes, and can influence behaviour independently of cognitive judgements. This new kind of approach contrasts with the hegemonic concept of conscious processing in cognitive neurosciences, which is based on the identification of brain circuits responsible for the processing of (cognitive) representations (Lewis *et al.*, 2008<sup>[150]</sup>).

Both emotion and consciousness depend on neural representations of the subject's own body, arising from structures in the brainstem and medial telencephalon, which receive interoceptive information. We need, not only more data, but also further theoretical development of the concepts that are under investigation, which make the intersection of emotion and consciousness a fruitful domain for collaborations among neuroscientists, psychologists and philosophers (Yang, 2011<sup>[244]</sup>).

One needs to examine, across development and across phylogeny, the correlation between an elaborated self-representation and the capacity for a rich conscious experience. Do all invertebrates have explicit central interoceptive representations and can this criterion be used to determine which species might be capable of conscious experience? One needs to explore the relationship of moods to consciousness. Neuronal activity during sleep, under anaesthesia and during wakefulness has been studied at the neurochemical, electrophysiological and computational level. Positive or negative mood alters the distribution of neurochemical modulators (Damasio, 2000<sup>[59]</sup>). Can one model the effects of mood in computational models of consciousness? What is the role of language and symbolic thought in emotion experience? The cognitive interpretation of a situation influences the emotion experience, but how and at what stage of processing? Is there an aspect of emotion experience that is relatively independent of thought and reflection, and an aspect that depends on it? (Gazzaniga, 2012<sup>[101]</sup>).

Charles Darwin (Darwin, 1889<sup>[62]</sup>) has described in detail the aspects of emotional expression that serve as social communicative functions, and phenomena such as emotional contagion and empathy demonstrate that emotional expression in others can induce emotions in us. This topic has received

considerable attention from simulation theories and from discoveries of mirror neurons and mirror systems in the brain. Should a broader conception of an emotion state, or even an emotion experience, be transpersonal and include more than one brain? These suggestions could be exciting future directions for social neuroscience (Bernat, 2006<sup>[20]</sup>).

If basic emotion processing was necessary for consciousness, severe impairment in the basic emotion processes should lead to compromised consciousness. A more neuroanatomically specific hypothesis would be that alteration to structures that represent physiological changes in one's own body would alter or destroy conscious experience. We see more psychological theories of emotion and biological ones as being commensurate and complementary of each other (Damasio, 2003<sup>[60]</sup>).

### 3.4 A neural model of emotional consciousness

The key phenomena that a theory of emotional consciousness should explain include differentiation, integration, intensity, valence and change. Each of these aspects provides a set of explanation targets in the form of questions that a theory should answer. A mechanism is a structure performing a function in virtue of the operations, interactions and organisation of its component parts (Bechtel and Abrahamsen, 2005<sup>[18]</sup>; Thagard, 2006<sup>[230]</sup>). Candidates for explaining emotional phenomena include, for example, neural mechanisms in which the parts are neurons and the operations are electrical excitation and inhibition; biochemical mechanisms in which the parts are molecules and the operations are chemical reactions organised into functional pathways; and social mechanisms in which the parts are people and the operations are social interactions.

By *differentiation* we mean that people experience and distinguish a wide variety of emotions. The English language has hundreds of words for different emotions, ranging from the commonplace 'happy' and 'sad' to the more esoteric and extreme 'euphoric' and 'dejected'. Some emotions, such as happiness, sadness, fear, anger and disgust, seem to be universal across human cultures (Ekman, 2003<sup>[82]</sup>), while others may vary with different languages and cultures. Some emotions such as fear and anger appear to be shared by non-human animals, whereas, others such as shame, guilt and pride seem to depend on human social representations. A theory of emotional consciousness should be able to explain how each of these different experiences is generated by neural operations.

By *integration* we mean that emotions occur in interaction with other mental processes, including perception, memory, judgement and inference. Many emotions are invoked by perceptual inputs, for example, seeing a scary monster or smelling a favourite food. Perceptions stored in memory can also have strong

emotional associations, for example, the mental image of a sadistic third-grade teacher. Hence, a theory of emotional consciousness needs to explain how perception and memory can produce emotional responses. A theory of emotional consciousness must explain how we combine our awareness of an object with an associated emotion and must account for how different interpretations of a situation can lead to very different emotional reactions to it, as when a tap on the shoulder is construed as an affectionate pat or an aggressive gesture (Humphrey, 2006<sup>[123]</sup>).

A theory of emotional consciousness must provide a mechanism for explaining such differences in intensity. It must also provide a mechanism for valence, the positive or negative character of emotions. Positive emotions like happiness and pride have a very different qualitative feel from negative ones like fear, anger and disgust. We need to identify the neural underpinnings of experiences with these different valences (Winkielman and Schooler, 2011<sup>[243]</sup>).

The last set of emotional phenomena that a theory of emotional consciousness must be able to explain concern change. Emotional changes include shifts from one emotion to another as the result of shifts in attention to different objects or situations, but can also stem from a reinterpretation of a single object or situation, as when a person goes from feeling positive about a delicious food to feeling negative when its caloric consequences are appreciated. Emotional changes can also be more diffuse, as when a generally positive mood shifts to a more negative one as a frustrating day unfolds (Churchland, 2002<sup>[47]</sup>; Marian and Shimamura, 2012<sup>[158]</sup>).

It should not be surprising that explanations of phenomena that involve both emotions and consciousness require a wide range of neurological and physiological mechanisms, of which one model is the EMOCON model. The EMOCON model is largely consistent with sophisticated frameworks for naturalising consciousness proposed by various researchers (Baars, 2005<sup>[14]</sup>; Edelman, 2003<sup>[79]</sup>; Koch, 2004<sup>[137]</sup>; Panksepp, 2005<sup>[176]</sup>).

Despite EMOCON's comprehensiveness, there are notable elements not included in the model because they do not seem to increase its explanatory power. It has not included any special non-material substance of the sort that theologians and other dualists have postulated as the basis of consciousness. The EMOCON theory has not assigned any special role to neural synchronisation accomplished by a 40-Hz (40 cycles per second) brain wave that various theorists have speculated may contribute to binding representations together (Engel *et al.* 1999<sup>[83]</sup>). If there is such synchronisation, it is probably an effect of neural processing rather than a causal factor.

If the EMOCON account of emotional consciousness is correct, it has implications for other interesting psychological phenomena including intuition



and ethical judgement. Intuitions, or gut feelings, are conscious judgements that can be viewed as originating from the same interconnected processes. Similarly, ethical judgements are always emotional and conscious, but they can also have a cognitive-appraisal component that complements the somatic signalling that is also part of the account. Thus, the identification of some of the neurophysiological mechanisms responsible for emotional consciousness must help to illuminate many other aspects of human thinking. First, it must provide a new theoretical account of the neural mechanisms of emotion that synthesise previously disjointed accounts based on somatic perception and cognitive appraisal, and second, it must show how these mechanisms can give rise to central aspects of emotional experience, including integration, differentiation, valence, intensity, and change (Thagard and Aubie, 2008<sup>[231]</sup>).

### 3.5 Towards a phylogeny of consciousness

Consciousness was long considered a human privilege, all other animals being merely machine-like beings (Cabanac, 1996<sup>[31]</sup>). This view was challenged when Darwin pointed out that other mammals could express emotion (Darwin, 1889<sup>[62]</sup>). The question then faded into the background, largely because of the excesses of psychoanalysis and the efforts of the behaviourist school to make behaviour the only object of study, to the exclusion of the underlying thought processes. In the 1990s, there was a renewal of interest in animal consciousness (Griffin, 1992<sup>[108]</sup>; Dawkins, 1993<sup>[65]</sup>) and a growing acceptance that humans were not the only thinkers. Indeed, if we accepted indirect evidence for the existence of human consciousness in other people, that is, from the verbal and behavioural signs that they provide, why should similar indirect evidence be rejected when it came to animals?

Although less direct than that provided through verbal communication, such evidence is available (Burghardt, 1999<sup>[29]</sup>). Yet, one must be prudent and always remain aware that the evidence is always indirect. For example, many fishes display complex behaviours such as cheating, altruism, species recognition, individual recognition and cleaning symbiosis (Heyes, 1994<sup>[118]</sup>) that we would be tempted to consider signs of consciousness, but these can be explained on the basis of simple reflexes. Also, the complex foraging and social communication behaviour of bees is often considered intelligent and 'conscious;' however, Gould and Grant-Gould (1995)<sup>[106]</sup> have shown that it is purely reflexive, in the same way as a computer can be artificially intelligent.

If we exclude self-consciousness — a human property — from the private model of reality that consciousness is, we may ask the question as to which animals are conscious? And which are not? At what point in evolution did nervous systems cease to operate only on a reflexive basis? Before the apes? Mammals? Or Vertebrates? (McFarland, 1991<sup>[162]</sup>).



The existence of consciousness in an animal does not imply that the behavioural responses are rational in those animals that possess a mental space. On the contrary, this mental space may simulate several possible lines of action and use the feelings they evoke to decide which response is best. Dictionaries provide no precise term for this kind of non-rational mental modelling when the response is purely reflexive. It may be appropriate to use mentalist terminology, that is, emotions, feelings and so on., but only for animals. Animal behavioural responses that can be mimicked in artificial models must be described in a way that does not imply consciousness (Denton, 1994<sup>[76]</sup>; Damasio, 2012<sup>[61]</sup>).

Consciousness could have emerged because of the increasing complexity of life in a terrestrial environment. In this new adaptive landscape, existence required many more stimulus-response pathways; eventually, a point was reached where it became more efficient, in terms of speed and flexibility, to route all decision-making through a single mental space (Koch, 2012<sup>[138]</sup>). Within this space, different possible responses would be compared and judged according to the degree of pleasure they evoked, the aim being to maximise pleasure and minimise displeasure. The hedonic dimension of consciousness thus became a common currency in decision-making, to select the final behavioural path in animals (Cabanac 1999<sup>[32]</sup>; Cabanac, Cabanac and Parent, 2009<sup>[33]</sup>).

### 3.6 The emergence of consciousness in fetal and neonatal life

A simple definition of consciousness is sensory awareness of the body, the self and the world. The foetus may be aware of the body, for example, by perceiving pain. It reacts to touch, smell and sound, and shows facial expressions responding to external stimuli. However, these reactions are probably preprogrammed and have a subcortical non-conscious origin. Furthermore, the foetus is almost continuously asleep and unconscious, partially due to endogenous sedation. Conversely, the newborn infant can be awake, exhibit sensory awareness and process memorised mental representations. It is also able to differentiate between self and non-self touch, express emotions and show signs of shared feelings. Yet, it is unreflective, present oriented and makes little reference to the concept of him/herself. Newborn infants display features characteristic of what may be referred to as basic consciousness and they still have to undergo considerable maturation to reach the level of adult consciousness. The preterm infant, *ex utero*, may open its eyes and establish minimal eye contact with its mother. It also shows avoidance reactions to harmful stimuli. However, the thalamo-cortical connections are not yet fully established, which is why it can only reach a minimal level of consciousness (Lagercrantz and Changeux, 2009<sup>[143]</sup>).

At birth, the newborn brain is in a 'transitional' stage of development, with an almost adult number of neurons (with the exception of adult neurogenesis), but an immature set of connections (Nowakowski, 2006<sup>[174]</sup>). During the few months after birth, there is an overproduction of synapses accompanied by a process of

synaptic elimination and stabilisation, which lasts until adolescence (Bourgeois, 1997<sup>[281]</sup>). Myelination begins prenatally, but is not completed until the third decade in the frontal cortex, where the highest executive functions and conscious thoughts take place (Koch, 2004<sup>[137]</sup>). Thalamic afferents to the cortex develop from approximately 12–16 weeks of gestation, reach the cortical subplate, but wait until they grow into the cortical plate (Kostovic and Milosevic, 2006<sup>[139]</sup>).

From approximately 34 weeks, a synchrony of the EEG rhythm of the two hemispheres becomes detectable at the same time as long-range callosal connections, and thus, the global neuronal workspace (GNW) circuits are established (Vanhatalo and Kaila, 2006<sup>[237]</sup>). From the twenty-sixth week, the pyramidal neurons in the primary visual cortex of humans develop dendritic spines. At birth, the dendritic spines would not have reached adult density, but are sufficient for the detection of visually evoked potentials. Connectivity of the cerebral cortex, particularly in the prefrontal area, matures later than the subcortical structures. However, the fusiform area for face recognition (Johnson, 2005<sup>[132]</sup>) and the left hemispheric temporal lobe cortices for processing speech stimuli (Lagercrantz *et al.*, 2002<sup>[141]</sup>) already function in the newborn.

Moreover, the main fascicles of myelinated long-range connections such as the corpus callosum, cerebellar peduncles, corticospinal tract and the spinothalamic tract are unambiguously identified at the age of one to four months in the newborn (Searle, 2000<sup>[209]</sup>). The neurochemistry of the developing brain reveals that gamma amino-butyric acid (GABA) is the dominant excitatory neurotransmitter during fetal life (Letinic *et al.*, 2002<sup>[148]</sup>). Right before or around the birth, depending on the brain area, GABA becomes the main inhibitory neurotransmitter. Then glutamate and aspartate become the major excitatory amino acids. In addition, a transient switch in GABA signalling from fetal excitatory to inhibitory is elicited by maternal oxytocin release upon delivery (Cote *et al.*, 2007<sup>[50]</sup>). The rich dopaminergic innervation of the prefrontal cortex (Nijhuis, 2003<sup>[172]</sup>) accompanies the cognitive advances in infants between six and twelve months. Well-defined sleep states appear at approximately 32 gestational weeks in the human foetus or in preterm infants.

By ultrasound recordings, active sleep can be identified by rapid eye movements, breathing, swallowing and atonia, whereas, apnea, absence of eye movements and tonic muscle activity occur during quiet or non-rapid eye movement sleep. This spontaneous activity is interpreted as an early 'inner stimulation,' which could anticipate the sensorimotor experience of the newborn with the outside world and regulate thalamocortical development (Prechtel, 1985<sup>[186]</sup>; Lewis *et al.*, 2008<sup>[150]</sup>; Light and Zahn-Waxler, 2012<sup>[152]</sup>).

A first conclusion of this ongoing research is that the foetus *in utero* is almost continuously asleep and unconscious, partially due to endogenous sedation. In particular, it would not consciously experience nociceptive inputs like pain.

Conversely, the newborn infant exhibits, in addition to sensory awareness, especially to painful stimuli, the ability to differentiate between self and non-self touch, sense that their bodies are separate from the world, express emotions and show signs of shared feelings. Newborn infants display features characteristic of what may be referred to as basic or minimal consciousness (Zelazo, 2004<sup>[256]</sup>). They still have to undergo considerable maturation to reach the level of adult consciousness. The preterm infant *ex utero* may open its eyes and establish a minimal eye contact with its mother. It also shows avoidance reactions to harmful stimuli. A pending question is the status of the preterm foetus born before 26 weeks (700 g), who has closed eyes and seems constantly asleep. The immaturity of its brain networks is such that they may not even reach the level of minimal consciousness. The postnatal maturation of the brain may be delayed and there are indications that the connectivity within the brain will be suboptimal in some cases as indicated by deficient executive functions (Lagercrantz, 2007<sup>[142]</sup>). Therefore, the timing of the emergence of minimal consciousness has been proposed as an ethical limit of human viability, and it may be possible to withhold or withdraw intensive care if these infants are severely brain damaged (Gazzaniga, 2006<sup>[100]</sup>).

## 4. Contributions of Quantum Physics to Consciousness

### 4.1 The interface of physics and consciousness

From the advent of quantum mechanics and quantum physics to the present day, physicists worldwide have struggled with proposing and disposing physical theories to explain the workings of the human brain and the nature of human consciousness. We shall in this section look at the various contributions made by quantum physics and other modern physics theories to our understanding of consciousness. For a neuroscientist the conceptual basis of these theories has more to offer than the mathematical jargon and formulae on which they stand. I shall stay away from physics formulae and mathematical operations and shall try to convey to the reader the essence of theories which renowned physicists have put forward.

Physicists have regarded consciousness as fundamental. Everything we talk about existing, postulates a consciousness and one cannot shy away from the same. Some scientists regard consciousness as a taboo and exclude it while proposing a theory of everything; but consciousness, being transcendental and immanent, has been encountered whenever any theory has tried to explain the universe and all its phenomena. It is worthwhile to mention that the mind and the objective world are inter-related and non-separable. They may even at times be regarded to be the same or reflections of one another. Space and time often regarded as physical entities are in fact mental constructs, although they are studied more by physicists than by psychologists worldwide. It is also

important that physicists realise that science can account for and explain only a small part of reality, that is, the part that we see and perceive. The nominal or hidden part is often not accessible to any form of scientific research and always remains a mystery. Mind and matter are thus equitable and non-separable. For further discussions on the mind and matter problem, brain-mind dyad as well the functions and structure of the mind and the brain, readers are referred to excellent reviews on the same in previous editions of this journal (Singh and Singh, 2011<sup>[217]</sup>).

#### 4.2 The nature of time and the riddle of consciousness

The Oxford English Dictionary (Simpson and Weiner, 2009<sup>[213]</sup>) defines time as: “The successive states of the universe regarded as a whole in which every state is either before or after every other duration, indefinitely continued existence, the progress of which is viewed as affecting persons and things”. Although dictionary meanings are insufficient to define entities, they are useful starting points. As expected this meaning sheds little light on the nature of time, and inadvertently, makes things more confusing by introducing other concepts such as duration. The human brain has always been fascinated by the mystery of time. Humans have reflected on the nature, origin and flow of time from antiquity and continue to refine their understanding of time. They have used religion, mythology, philosophy, mathematics and science to unravel the mysteries of time (Carroll, 2010<sup>[37]</sup>).

Physical time speaks about the run of clocks in space, while space in turn is regarded by scientists as timeless. The universe is itself a timeless entity. The time we use and the running of clocks is merely a duration of material change. It is a reference point for various facets of life, while the real nature of time remains elusive. Time is a fundamental dimension of life. Time, as it is perceived in our brains, is based on the pacemaker accumulator model, where, in the brain, different populations of neurons fire in a distributed manner. Coincidental activation of different neural populations across the brain helps us to perceive time differences and help us to develop the notion of past, present and future (Barbour, 1999<sup>[16]</sup>).

Yet, despite the centrality of time in our life, it may not be a fundamental element of the universe. It appears that time is a way in which we have learnt to organise the universe. As Ernest Mach (1960)<sup>[155]</sup>, the famous Austrian physicist and philosopher put it, “Time is an abstraction at which we arrive by means of the changes of things.” This conception of time may appear surprising and counter-intuitive to everyday life; however, a number of developments in many diverse fields tend to support this conclusion (Frank, 2011<sup>[86]</sup>).

The idea of time as a ‘mere illusion’ has been adopted by modern physics. Time becomes even more counter-intuitive in quantum mechanics, where it may

simply be indeterminate in the quantum superposition phase events, and there is even a possibility that quantum information may be sent 'backwards in time', as exemplified by Aharonov's 'dual vector' theory (Aharonov and Bohm, 1958<sup>[8]</sup>). This effect has been experimentally verified in the most common case, called Aharonov-Bohm solenoid, in that, knowledge of the classical electromagnetic field acting *locally* on a particle is not sufficient to predict the quantum-mechanical behaviour. More interestingly, all laws of fundamental physics (i.e., the Dirac equation, Schrödinger's equation, Maxwell's equations, Einstein's field equations of gravity, Feynman diagrams) are time reversible (Barbour, 1999<sup>[16]</sup>). This is to say that, at the most fundamental level, there is no preference for direction (past). Physics provides no objective reason to believe that our present is in any way special, or more real than any other instant of time.

However, at the macro level, the laws of physics, chemistry and biology are irreversible. This is most clearly exemplified in the second law of thermodynamics, which states that the levels of entropy (disorder) increase in the universe as a whole. Thus, the arrow of time flows from the direction of less order to more disorder. However, even the second law of thermodynamics does not always guarantee a progression from the past to the future. If we look closely, it is the entropy of any closed system (and the whole universe can be considered a closed system) that increases in the direction of disorder *on an average*. For a single system, the entropy can either increase or decrease; thus the orientation of time is not absolute; and for small systems (such as neurochemical processes) it may become nebulous and difficult to resolve the direction in time of one (future) over the other.

Freud emphasised the timelessness of unconscious processes. He showed how the unconscious ignored time and temporal progression. For example, in dreams and fantasies, where past, present and future were united in one representation, he showed that certain aspects of psychopathology were also essentially atemporal. In a note added in 1907 to *The Psychopathology of Everyday Life* (1907)<sup>[88]</sup>, concerning the indestructibility of memory traces, Freud wrote that "the unconscious is completely atemporal." In his essay on the metapsychology of the unconscious, he further noted that the processes of the unconscious system were, "timeless, that is.. they are not ordered temporally, are not altered by the passage of time; they have no reference to time at all." (Freud, 1907<sup>[88]</sup>). Yet Freud struggled to reconcile his notion of unconscious time with his Kantian and Newtonian view of the psyche. He wrote, "If the philosophers maintain that the concepts of time and space are the necessary forms of our thinking, forethought tells us that the individual masters the world by means of two systems, one of which functions only in terms of time and the other only in terms of space" (Freud, 1915<sup>[89]</sup>). He believed that temporal dimension was accessible to us only as a function of acts of consciousness. As these acts in turn depended on rapid, periodic and discontinuous impulses from the unconscious – preconscious system, Freud believed that perception of time itself was discontinuous. He wrote,

"I further had a suspicion that this discontinuous method of functioning of the system lies at the bottom of the origin of the concept of time" (Freud, 1925<sup>[90]</sup>). For a more detailed review on the interface of Freudian theory and consciousness, the readers are referred to other sources (Holt, 1989<sup>[121]</sup>; De Sousa, 2011<sup>[67]</sup>).

Perception of time differs across cultures. In the Judeo-Christian culture, time is perceived as having a 'linear' form (i.e., past-present-future). We believe that the past is 'behind us', the future is 'in front us', and the present time is 'where we are now'. This concept of time is based on the notion that time is linear and unidirectional. As pointed out, our awareness of ourselves and others as growing, developing and ageing beings across the life span is a major source of our perception of time as linear in nature. Other cultures (e.g. Mayan) do not perceive time as a linear and uniform phenomenon and their time calendars consist of multiple and simultaneously existing time categories. These categories may include 'practical time', 'social time', 'religious time' and so on. Many indigenous cultures do not perceive time as linear and describe it as having a 'circular' or 'cyclic' form. Time is perceived as 'static' and the individual person is believed to be 'in the centre of time' (i.e., surrounded by concentric 'time circles'). Life events are placed in time along and across the 'time circles' according to their relative importance to the individual and his or her respective community. For example, more important events are placed closer to the individual and are perceived as being closer in time; unimportant or irrelevant are placed in peripheral time circles, although they may have happened more recently according to the linear concept of time (Penrose, 2005<sup>[181]</sup>).

Consciousness, like time, is difficult to define. What St. Augustine remarked about time can be equally true of consciousness, which is, when no one asked him, he knew what time was; however when someone asked him, he did not. One of the key features of consciousness is what seems to be temporal synchrony – in contrast to the idea that our conscious perceptions are non-synchronised (Dennett, 1991<sup>[75]</sup>). In fact at any given time, the nervous system is bombarded with a wide variety of visual, auditory and tactile inputs. What we perceive as the external reality is in fact the organisation and interpretation of this sensory data and is one of the fundamental aspects of consciousness. As Julian Barbour has argued, time may be a collage of haphazardly arranged moments whose continuity is an illusion of memory. Thus, it seems that time is a creation of consciousness (Frank, 2011<sup>[86]</sup>).

Time has been attributed to the innermost dimension of consciousness. There have been theories about the possibility of large extra dimensions to develop a theory of consciousness: According to this view, consciousness has a special extra dimension or 'brane' in the super-string theory, thus the ordinary space time becomes a part of the 'hyperspace' organised by consciousness (Smythies, 2003<sup>[222]</sup>). Similar ideas are expounded by Penrose and Hameroff. In their Orchestrated Objective Reduction (Orch-OR) model (Hameroff, Kazniak and

Scott, 1996<sup>[112]</sup>), they conceptualise consciousness as the successive quantum superposition of the tubulin protein conformations in the brain. They propose that with each conscious moment, 'a new organization of Planck scale geometry is selected irreversibly' (Penrose, 2010<sup>[182]</sup>). This leads to an apparent illusion of time. Thus without consciousness, there would be no time (Davies, 1995<sup>[64]</sup>; Penrose, 2010<sup>[182]</sup>).

### 4.3 Electromagnetic theories of consciousness

Two previously identified difficulties with the electromagnetic field theory of consciousness still hold true. The first difficulty is that, although spatiotemporal electromagnetic patterns co-varying with conscious experiences have been identified in rabbits and cats, no analogous patterns have yet been found in humans. Evidence is cited that this is very likely because the relevant patterns are inaccessible from the scalp (McFadden, 2002<sup>[161]</sup>). Recording from the surface of the human brain will be necessary. Such electrocorticography (ECoG) recordings are feasible in the context of the localising epileptogenic foci, but logistical difficulties have so far prevented their being done, with a view to identifying spatial patterns co-varying with conscious sensations. The second difficulty is that, although electromagnetic fields can certainly cause neural firing, the same mathematical calculations that show the need for ECoG reveal that the spatial patterns proposed as being conscious become unidentifiable such a short distance away from their source that they are ill-suited to causing behaviour by activating neurons in other areas of the brain. This difficulty is rendered unimportant by an accumulation of empirical evidence that consciousness is actually not causal for behaviour (Pockett, 2002<sup>[184]</sup>).

The essence of the hypothesis is that conscious experience (a.k.a. sensation) will prove to be identical with certain spatiotemporal patterns in the electromagnetic field. These patterns are at present generated only by living brains, but in principle they can be generated by hardware instead of software. The characteristics of the patterns have been left largely unspecified, except that they will probably be transiently occurring, brain-sized, spatial patterns of electromagnetic intensity or amplitude (i.e., voltage). One of the points which I then thought to be in favour of the theory was that such localised electromagnetic fields are known to be capable of causing neurons to fire, which in principle offers a mechanism by which consciousness could cause behaviour (Silverman, 2008<sup>[212]</sup>).

So what, after all that, is the present status of the electromagnetic field theory of consciousness? The theory has, I think, survived the criticism that electromagnetic fields are unsuited for a directly causal role in voluntary movement (a.k.a. behaviour). The answer here is that it is becoming increasingly likely that voluntary movements are *not* directly caused by consciousness. And if behaviour is not directly caused by consciousness, there is no requirement for putatively conscious electromagnetic fields to directly cause behaviour.



However, on the down side of the ledger for the theory, we have still not managed to describe a single empirical example of a spatial electromagnetic pattern that co-varies with a particular kind of *human* conscious experience. The only overt advance in that direction over the past seven years has been to determine the methodology that will very likely be necessary for making such measurements (Pockett, 2007<sup>[185]</sup>; Randall, 2011<sup>[190]</sup>).

#### 4.4 Dynamic geometry, brain function modelling and consciousness

A geometric interpretation towards understanding brain function has been proposed since many years. This interpretation assumes that the relation between the brain and the external world is determined by the ability of the central nervous system (CNS) to construct an internal model of the external world using an interactive geometrical relationship between sensory and motor expression (Roy and Llinas, 2008<sup>[202]</sup>). This approach has opened new vistas not only in brain research, but also in understanding the foundations of geometry itself. The approach named the tensor network theory is sufficiently rich to allow specific computational modelling and addresses the issue of prediction, based on Taylor series expansion properties of the system, at the neuronal level, as a basic property of brain function. It has actually proposed that the evolutionary realm is the backbone for the development of an internal functional space that, while being purely representational, can interact successfully with the totally different world of the so-called external reality (Pellionisz and Llinas, 1985<sup>[178]</sup>; Baianu *et al.*, 2011<sup>[15]</sup>).

Now if the internal space or functional space is endowed with stochastic metric tensor properties, then there will be a dynamic correspondence between events in the external world and their specification in the internal space. We shall call this dynamic geometry since the minimal time resolution of the brain (10–15 ms), associated with 40 Hz oscillations of neurons, and their network dynamics is considered to be responsible for recognising the external events and generating the concept of simultaneity. The stochastic metric tensor in dynamic geometry can be written as five-dimensional space-time where the fifth dimension is a probability space as well as a metric space. This extra dimension is considered to be an imbedded degree of freedom. It is worth noticing that the above-mentioned 40 Hz oscillation is present both in the awake and dream states, where the central difference is the inability of phase resetting in the latter (Leznik, Makarenko and Llinas, 2002<sup>[151]</sup>; Penrose, 2010<sup>[182]</sup>).

This framework of dynamic geometry makes it possible to distinguish one individual from another. Dynamic geometry plays a pivotal role in understanding the external world through the CNS. This internal geometry is sense-dependent in contrast to the deductive geometry used in modern physics. The weak chaotic nature of the oscillations of single neurons makes the metric of functional geometry a probabilistic one. The probabilistic nature of the geometry makes



it possible to construct a well-defined mathematical transformation between the outside world and the internal world using the tensor network theory. The concept of dynamic geometry will shed new light on the issue of consciousness and its neuronal correlates (Penrose, 2005<sup>[181]</sup>).

#### 4.5 The role of gravity in a theory of consciousness

Even as the role of gravity in consciousness is at present still at the level of experimental studies, researchers have offered a view on the origin, implications and potential applications of gravity (Hu and Wu, 2006<sup>[129]</sup>). The connection between quantum entanglement and Newton's instantaneous universal gravity and Mach's Principle is natural. To a certain degree, this view is a reductionist expression of this connection with important consequences. Readers are advised that these propositions are outside mainstream physics and other authors may hold dissimilar views on some of the points made.

Microscopically, gravity is assumed to be feeble and negligible and macroscopically it is ubiquitous and pervasive. It seems to penetrate everything and cannot be shielded. However, there is no consensus as to its cause, despite the efforts of many people. Presumably, this state of affairs is due to the lack of any experimental guidance (Hu and Wu, 2004<sup>[128]</sup>). There are many general and technical articles written on the subject. The propositions are explained in the following text.

Gravity originates from the primordial spin processes in non-spatial and non-temporal pre-space time and is the macroscopic manifestation of quantum entanglement. Thus, gravity is non-local and instantaneous, as Newton reluctantly assumed and Mach suggested. It implies that all matters in the universe are instantaneously interconnected and many anomalous effects in astronomy such as red shift, dark energy, dark mass and the Pioneer effect may be resolved from this perspective. Potentially, gravity can be harnessed, tamed and developed into revolutionary technologies to serve mankind in many areas, such as, instantaneous communication, space time engineering and space travel (Scharf, 2012<sup>[206]</sup>).

There are several existing approaches that provide some hints as to the said mathematical forms. These approaches are all based on non-local hidden variables, that is, the principle of non-local action. They include Bohmian mechanics (Bohm and Hiley, 1993<sup>[27]</sup>), Adler's trace dynamics, Smolin's stochastic approach (Smolin, 2006<sup>[221]</sup>) and Cahill's process physics (Siddharth, 2001<sup>[211]</sup>). In addition, other existing alternative approaches on gravity may also provide some hints. For example, Sakharov's induced gravity is a well-known alternative theory of quantum gravity in which gravity emerges as a property of matter fields (Tegmark, 2000<sup>[229]</sup>). Researchers, however, regard gravity as a product of quantum entanglement (Hu and Wu, 2006<sup>[129]</sup>).

#### 4.6 Three worlds and consciousness

It is only in the last 50 years that physics has spread its realms to explain phenomena like life, death, immortality, the concept of God and consciousness. It is important for readers to note that there have always existed three kinds of worlds, that is, the mathematical world, the physical world, and the mental world. These three worlds are interconnected, although only small parts of each world are of relevance to each other. There is a small part of mathematics that plays an important role in the understanding of physics. This includes the role of mathematics in quantum physics, the constants of nature and mathematical formulae that help in the elucidation of physical theories (Penrose, 2005<sup>[181]</sup>). Certain aspects of the physical world and certain physical structures like the brain are concerned with the development and maintenance of the mental world and mental concepts. In turn the mental world is also relevant and is needed, to understand and operate the abstract concepts of mathematics and physics (Penrose, 2005<sup>[181]</sup>).

Thus, we have three systems, three worlds that are all interlinked to each other and needed for each others' existence. There is thus no doubt that quantum physics and mathematics will have their share when it comes to explaining certain facets and theories of consciousness. These theories may run parallel to those of neuroscience and cognitive science, or in turn may even help fill the gaps posited by those theories (Penrose, 2005<sup>[181]</sup>; 2010<sup>[182]</sup>).

#### 4.7 The anthropic principle and consciousness

How important is consciousness for the universe as a whole? Can a universe exist without any conscious inhabitants whatsoever? Are the laws of physics designed in order to allow the existence of conscious life? Is there something special about our location in the universe, either in space or time? These are the kinds of questions that are addressed by what has come to be known as the anthropic principle. The principle has many forms, but the most clearly acceptable form addresses the spatio-temporal location of conscious or intelligent life in the universe. This is the *weak* anthropic principle. The argument can be used to explain why conditions happen to be just right for life on earth at the present time (Stein, 2011<sup>[224]</sup>). This has been known as the weak anthropic principle and has been used by physicists to explain the relations between various constants of nature like the gravitational constant, mass of a proton, age of the universe, and so on. This in turn also forms the basis for various theories put forth by quantum physics. By the use of this principle we show that consciousness is inevitable by virtue of the fact that sentient beings, that is 'we', have to be around to observe the world and thus sentience does not have any selective advantage (Stein, 2011<sup>[224]</sup>). Some physicists also believe that this principle is the reason for the evolution of consciousness and that consciousness is here without it being favoured by natural selection.

## 4.8 Quantum physics, computers and neuroscience

There are some points of difference between brain action and computer action that seem to be of greater importance than the ones so far mentioned, having to do with a phenomenon called brain plasticity. It is wrong to regard the human brain as simply a fixed collection of wired up neurons. The interconnections between neurons are not in fact fixed, but are changing all the time. I do not mean that the location of axons or dendrites in the neuron will change, but refer to the synaptic junctions where communications between different neurons actually takes place. Often these occur at the dendritic spines, which are tiny protuberances where contact with the synaptic knobs can be made. Contact here means just leaving of a small gap called the synaptic cleft. These spines under certain conditions can break away or make new contacts. Thus, if we consider the brain and its neuronal connections to be a computer, then it is a computer that is capable of changing its circuitry independently all the time (Aleksander, 2000<sup>[3]</sup>).

Many scientists appear to be of the opinion that the development of parallel computers holds the key to building machines with capabilities like the human brain, and that this, in turn, may help us create consciousness in the laboratory. The motivation for this type of computer comes from the study of neural architecture and an attempt to imitate the operation of the nervous system, as different parts of the brain indeed seem to carry out separate and independent calculational functions (e.g. with the processing of visual information in the visual cortex). The oneness of conscious perception seems to be at quite a variance with the picture of a parallel computer. This analogy is, however, more appropriate when we look at the unconscious functioning of the human brain. Various independent movements like walking, fastening, breathing and even talking, can all be carried out simultaneously and more or less autonomously without one being even consciously aware of them. There also seems to be some relation between this oneness of consciousness and quantum parallelism. In the quantum theory, different alternatives at the quantum level are allowed to co-exist in linear superposition. Thus, a single quantum state could in principle consist of a large number of different activities, all occurring simultaneously. This is what is called quantum parallelism (Nayak *et al.*, 2011<sup>[170]</sup>).

## 4.9 Concluding remarks of this section

There is no doubt that extraordinary advances have been made in our understanding of the physical world in which we reside and consciousness, and this has come as a result of careful physical observation and rigorous experimentation. Physical reasoning of great depth along with mathematical arguments ranging from the complicated, but routine, to inspirational leaps of the highest order have added in this endeavor. Thus we have learnt from physics the geometry of space through to Newtonian mechanics, to the magnificent structures of classical mechanics, Maxwell's electromagnetic field theory and

thermodynamics. More recently we have come across Einstein's general and special theory of relativity and the deeply mysterious yet profoundly accurate and broad ranging quantum mechanics the quantum field theory.

The absence of strong experimental data relating to the normal quantum proposals of consciousness has led to a curious situation in theoretical fundamental physics research. A general consensus exists for real progress to be made, so we have to move beyond standard particle models of physics and cosmology to understand consciousness, it is necessary to have a quantum theory in nature that encompasses gravity in addition to strong and weak electromagnetic forces. However, as experiments in this area are absent, the efforts of theoreticians have been directed very much into the internal world of mathematical desiderata. Despite all this there is lot to be learnt from the nature of quantum physics and how it defines a newer understanding of consciousness.

## **5. Contributions of Philosophy to Consciousness**

### **5.1 General issues in philosophy in relation to consciousness**

Human beings have been interested in thinking since time immemorial. Thinking in philosophy of various issues may be an end in itself. In issues that involve consciousness, existence and being, it is well known that abstraction is inevitable. Consciousness has been intellectually perplexing and has thus vexed philosophers through the ages. It is worthy to note in philosophy that the science of consciousness cannot avoid being pluralistic in nature. There is at first a neuroscience and structural or materialist approach to consciousness, where various different areas of neuroscience converge when one speaks or thinks of consciousness. The second approach is the dualism approach where one thinks of the mind and body as separate, although both may contribute to an understanding of consciousness. The brain itself can never be thought of as isolated phenomena and has to be related to both the body and the mind. One way to establish the relation between brain, body and mind is to get out of the mess we have landed ourselves in, and by reifying mind: By accepting that the body includes the brain, and mind is just a collection of brain functions (Singh and Singh, 2011<sup>[217]</sup>).

Even if that were so, consciousness is an integral part of philosophy; but the very essence of thought being pluralistic makes consciousness bend in the same direction (DiFrancesco, 2008<sup>[77]</sup>).

### **5.2 Some philosophical approaches to consciousness**

First we look at an intentionalist approach that holds the view that conscious states are nothing more than intentional states, that is, states exhibiting

intentionality or the capacity to represent something beyond themselves. The difficulty with this approach is that qualia seem devoid of intentionality. Qualia thus seem to be an extra element, an aspect of consciousness over and above the intentional content. The overall experience of a toothache may include the thought that one is in pain — a thought that representing as it does one's current situation exhibits intentionality, but the pain itself is a further non-intentional component (Humphrey, 2011<sup>[124]</sup>).

Some philosophers deny that there are any qualia to account for in the first place. This is what is called an eliminativist position. This deals with a philosophically problematic phenomenon by suggesting that its problematic nature gives us reason to doubt its existence — to eliminate it entirely from our picture of the world, rather than trying to explain it. These philosophers do not deny that we have conscious experiences, but deny that these experiences feature properties of the sort that qualia are taken to be. Thus, there are no properties that are not essentially intrinsic that are un-analysable in terms of their relations or subjective, that is, analysable only from the first person point of view (Smith and Jovic, 2003<sup>[220]</sup>).

If qualia cannot be dismissed as unreal then how does an intentionalist theory of consciousness deal with them? The answer is a philosophy called representationalism, which is the view that qualia are nothing more than representational properties of conscious experiences. This is where we have the higher order theory of consciousness. This is the idea that what makes any particular mental state a conscious state is that it is the object of a higher order mental state that represents it. Some may take these states to be thoughts, some take them to be more akin to perceptions, or a thought about a thought or rather the inner perception of the perception itself. There are also philosophical views that all consciousness and its correlates are ultimately material (materialism) and that the material and mental are equally ultimate (dualism). These are paid the most attention to by contemporary philosophers of the mind (Crane, 2001<sup>[56]</sup>).

A third view is known as idealism and holds that everything is ultimately mental. It proposes that objects like tables and chairs really exist only in so far as a mind perceives them to exist. This is, however, not regarded as a serious option by most contemporary philosophers. The first alternative to the three views mentioned above is that neither mind nor matter is metaphysically ultimate and that what is ultimate is rather a single kind of stuff that is neutral, between and more fundamental than either of them. This in a nutshell is the metaphysical theory known as neutral monism. For instance, what exactly are the colourless, odourless, tasteless particles of which physics speaks — molecules, atoms, quarks, gluons and so forth? We know from science only that the material world is a collection of fundamental entities having a certain causal structure described in mathematically precise details by physics. However, what is it that fleshes

out this causal structure, the entities that bear these causal relations and form components of one another in the vast causal network described by physics, is something we do not know. This view regarding consciousness is called structural realism. Structural, because all we know of the world is its structure rather than its intrinsic nature. It is called realist, because there is really a physical world existing external to our minds (Papineau, 2002<sup>[177]</sup>).

Despite advances made in our understanding of consciousness we seem left metaphysically in the position where subjectivity lies at the core of the mental and is the main obstacle in the way of the material account of conscious experience. There is a sense in which qualitative conscious states may be identified with states of the brain. Perception of a brain state and introspection of a mental state may be seen as two different ways of representing the same thing. The dualist view may settle this issue partially, but a question often asked is how do discrete brain processes add up to a meaningful unified experience. This is known by neuroscientists, cognitive scientists and mind philosophers as the binding problem and is reflected merely by a small gap in our scientific knowledge (Searle, 1992<sup>[208]</sup>). Singh and Singh (2011<sup>[217]</sup>) have recently attempted to explain how brain processes add to a meaningful unified experience in their formulation called 'the lattice of mental operations'.

### 5.3 Consciousness, mind and matter in Indian philosophy

Consciousness and its relation to the physical body had been thoroughly analysed in the Indian philosophy of ancient times. There are many concepts in ancient Indian literature that could lead to scientific answers to some of the questions posited by brain scientists and modern consciousness researchers. In Indian philosophical literature, thought is often described as being fast and one that never comes to a stop (Abhedananda, 2008<sup>[1]</sup>). Interestingly, according to modern physics, a faster than light object called the tachyon cannot be brought to rest. If the 'mind' indeed contains superluminal objects, it will be possible to describe the properties and processes of the human 'mind' in terms of mathematics and physics along with quantum mechanics (Hari, 2008<sup>[13]</sup>). Indian philosophy has always been considered complex and mysterious. It was written a long time ago and in Sanskrit (a language not spoken here today). Consciousness was often described there in the context of spiritual progress.

Indian philosophy looks at the human body as a piece of hardware made up of matter. Every living being has an accumulation of experiences or information, in other words a memory (called *manas* in this literature). Consciousness is regarded as that part of the mind that 'knows'. It is like the computer operator, and one that knows everything that is a part of the living beings' activity. Indian philosophy emphasises that consciousness is the same as free will, different from and independent of any living being's memory and its contents and mechanism. The philosophy makes a distinction between information and consciousness. The former

produces experiences in response to external inputs, just like a computer software, while consciousness is the ability to really know and choose (Swami, 1999<sup>[228]</sup>).

Modern Indian philosophy divides consciousness into two components, namely, free will and the mind. Free will is independent of all causes. It is the ability to decide consciously and independent of any reason from the past or present and without expecting anything in the future. Manifestation of free will is not an unconscious, non-deterministic, random occurrence. It is independent of space and time, does not depend on any memory and is not bound by any rules or logic. It is said to be *nishkarana*, meaning that it is not the effect of any cause. Neither can its existence be described nor can its occurrence be predicted. *Manas* keeps accumulating a lot more as life goes on. It is a sense like sight, touch, hearing, smell and taste. It is said to be *sukshma* (meaning soft and subtle) as compared to the body which is *sthula* (meaning gross and hard). The mind has been described as being faster than matter and the mind never comes to rest (Mukherjee, 2002<sup>[168]</sup>). *Manas* is different from the body and neither of the two can be transformed into one another.

Indian philosophy is dualistic in the sense that it asserts that just like in the computer, the living brain's software, the mind is also real information and cannot be created from ordinary matter all by itself. Life and consciousness is thus a process of the interaction of both mind and body. Life begins when the mind starts interacting with a body and lasts as long as this interaction continues. The reincarnation principle of eastern religions, Hinduism and Buddhism, states that the living mind or soul never ceases to exist when a person dies and simply survives and starts interaction with another body when such a suitable body is found. This principle has, however, not yet been proved by modern science. Indian philosophy is mostly known as monism as it explains that consciousness alone appears in various forms in the universe. It is like gold and all objects in the universe are ornaments made from this gold. This fact can be realised only by spiritual means and thus the monistic part of Indian philosophy does not conflict with the dualistic part described above (Abhedananda, 2009<sup>[2]</sup>).

Indian philosophy insists that each individual is born with their very own *karma* (subconscious memory of past actions whose consequences will take place in the future), *vasanas* and *samskaras* (subconsciously remembered skills, inclinations, likes and dislikes and so on.). Consciousness is thus more complicated than subjective knowledge and inference. The subjective experience arises because of the ever-present consciousness observing the mind's contents and thoughts (Hari, 2008<sup>[113]</sup>; 2010<sup>[114]</sup>).

## 5.4 The philosophy needed for a scientific theory of consciousness

Given the understanding of the form of theories in the physical sciences and what such theories actually deliver, and the arguments in favour of modular



hierarchies in the human mind/brain, the general form of a scientific theory of consciousness may be sketched.

First, it would involve causal descriptions, at a psychological level, of some phenomena generally regarded as being conscious. Second, a theory would involve identification of the different types of processing performed by different physiological structures such as the cortex, thalamus, basal nuclei and so on, and the differences between the processing performed by substructures of these structures. The expectation would be that there would be more similarity between the types of processing performed by different cortex modules than between a cortex module and a thalamus module, for example. The causal relationships at the psychological level would then be mapped into sequences of combinations of such processes. The descriptions at the highest level would be approximate, but a comprehensible description of an end-to-end conscious process would be possible. Third, an important issue is whether it is possible to create the needed hierarchy of intermediate-level descriptions. The earlier discussion indicates that the combination of resource constraint, modifiability, reparability and constructability requirements tend to result in a modular hierarchy with the right properties to support such a description hierarchy (Coward and Sun, 2004<sup>[53]</sup>).

A description of a conscious behaviour in terms of high-level modules would be comprehensible but approximate for a macro-behaviour. However, a path would exist for describing various small components of the behaviour to any desired degree of detail, with the use of the lower-level modules, at various intermediate levels of the description hierarchy. The critical point is that some degree of inaccuracy will be inherent in the higher levels of descriptions of consciousness, but this is not necessarily a failure of the science. For one thing, it is also present in the physical sciences. Approaches to understanding consciousness based on delineation and separation of subsystems that perform different types of information processes (thus forming intermediate level descriptions) are in actuality possible (Coward, 2005<sup>[54]</sup>).

A scientific theory of consciousness must be based on an understanding of what a scientific theory should be like and what it actually delivers. A scientific theory of consciousness will have some critical qualitative characteristics, based on the properties of theories in the physical sciences (and the manner in which understanding is possible in complex computational systems). First, it will be made up of a hierarchy of causal descriptions. At a high (e.g. psychological) level there will be many different types of descriptive entities, a relatively low information density in the descriptions and a relatively high degree of approximation. At a detailed (e.g. physiological) level there will be few different types of descriptive entities, a higher level of information density and a much lower degree of approximation. There will be a number of intermediate levels of description with intermediate numbers of the types of entities, information densities and degrees of approximation. This hierarchy of description is



necessary, if understanding of a very complex system is to be possible within the human mental capabilities. Each level can describe a cognitive phenomenon in its own terms, but the differences in information density mean that although a description of a complete psychological phenomenon at the highest level would be comprehensible, only descriptions of small segments of that phenomenon will be comprehensible at more detailed levels. It must be possible to map descriptions between levels, and there must be a clear understanding of when translation to a deeper level is necessary, to achieve a required degree of accuracy (Coward and Sun, 2004<sup>[53]</sup>).

Second, a number of practical needs/constraints tend to result in the resources of the brain being organised into a modular hierarchy, and modules at different levels in this hierarchy have the properties of entities at different levels in a hierarchy of descriptions. In complex computational systems, it is this modular hierarchy that makes human understanding possible, and an equivalent hierarchy in the brain would enable some scientific understanding of consciousness. Resource constraints mean that modules will not in general correspond with the features, as perceived by an outside observer. The question of the possibility of such a theory centres round the issue of whether an appropriate hierarchy of descriptions can be constructed within the limits of human intellectual capabilities (Sun, 2002<sup>[226]</sup>).

From this perspective, there are some common problems with some existing theories. One is that the approximate nature of higher-level descriptions is not recognised, and another is that attempts are made to match modules with superficial cognitive features rather than deeper (resource-driven) types of information processes. A few, very brief comments about some existing approaches would be as follows, in relation to the above (Sun, Coward and Zenzen, 2005<sup>[227]</sup>)

- The approximation inherent in higher-level descriptions means that the wholesale exclusion of approaches like folk psychology is not necessarily appropriate, because such approximate descriptions might sometimes be, in some ways, analogous with the higher levels of description in the physical sciences (Arico *et al.*, 2012<sup>[11]</sup>).
- The issue with computational/mathematical modelling without reference to lower-level (e.g. physiological) structures at all, is that such modelling is analogous to trying to directly implement the user manual for a complex computational system. Such an implementation would be possible in principle, given unlimited information handling resources, but the result would sometimes bear little resemblance to the actual system architecture, and any system features/functionalities not explicitly addressed in the user manual ('accidental capabilities') would in general not be present in the implemented user manual (Coward, 2001<sup>[52]</sup>).
- The limitation to searching for some physiological activity that correlates with consciousness is that it can be anticipated that any one physiological

structure will participate in a wide range of functionalities (as discussed earlier). It is probably not the activity of one structure that will correlate with consciousness, but rather a specific combination and sequencing of modular activities that will help understand consciousness. Such an activity combination and/or sequencing would define the presence of conscious behaviour, but unlike the proposal of Lamme (2006)<sup>[144]</sup>, there might be consistency between the psychological and physiological measures.

- The invocation of quantum mechanics is almost equivalent to an argument that no hierarchy of descriptions exists that can bridge the gap. Given that biochemistry may be understood at a level of description higher than quantum mechanics, there seems to be no clear reason why understanding of conscious phenomena would require descriptions exclusively at such a low level, except in the same sense that all sciences will require quantum mechanics, as the degree of required precision increases (Sun, 1999<sup>[225]</sup>).
- How does this relate to the so-called difficult problem? Why is it that when our cognitive systems engage in visual and auditory information processing, we have a visual or auditory experience? How can we explain why there is something like a mind to entertain a mental image, or to experience an emotion (Chalmers, 1995<sup>[40]</sup>)? This is not exactly a scientific question within the current scope of science, although a scientific understanding will probably reduce the sense of mystery around the questions by giving a better perspective on the issue.

### 5.5 The function of a consciousness: Neurophilosophical insights

It is widely held that we understand something only when we can explain it, and explaining a natural phenomenon typically, if not always, means locating it in its distinctive causal nexus. Moreover, when those phenomena are biological, locating a phenomenon often means specifying what function it has for the relevant organism, and that is deciphering something about the phenomenon that tends to benefit organisms or confer on them some adaptive advantage. This is especially so with mental phenomena and their associated brain processes. We expect that a satisfactory explanation of perceiving, thinking and other mental occurrences will involve coming to know how those processes and the brain events that subserve them function to benefit the organism (Koch, 2012<sup>[138]</sup>).

It is widely acknowledged that thoughts, desires and emotions sometimes occur without being conscious. As all these states occur in both conscious and non-conscious forms, an additional question about their function arises: What, if any, function do conscious versions of these states have that non-conscious versions lack? It is plain that we will not fully understand perceiving, desiring, and thinking without knowing what functions they serve. So it may also seem obvious that we also cannot fully understand the consciousness of these states without understanding the function of such consciousness. Understanding

consciousness, it seems, requires knowing how the consciousness of psychological states contributes to the well-being of the organism. This concerns the consciousness of psychological states. Not only do we distinguish between psychological states being conscious or not, we also distinguish the conscious from the unconscious condition of individual organisms themselves. A creature is conscious if it is awake and responsive to sensory stimuli as against being asleep, anaesthetised or unconscious. There is no doubt that an organism's being conscious enables it to interact with its environment in ways that greatly enhance its well-being and survival (Winkielman and Schooler, 2011<sup>[243]</sup>).

However, even when an organism is fully conscious, many of its psychological states may fail to be conscious states. Fully conscious humans often have many thoughts and desires that are not conscious states, and have subliminal perceptions, which are also not conscious. So, we cannot infer from the function of an organism's being conscious to a function of the consciousness of its psychological states (Humphrey, 2011<sup>[124]</sup>).

The difference between these two functions is sometimes overlooked (Merker, 2005<sup>[164]</sup>; Morsella, 2005<sup>[166]</sup>), perhaps because it is assumed that the psychological states of an awake organism are invariably conscious. There is a second issue that is sometimes confused with this, about the function that accrues specifically to psychological states being conscious. Thinking, perceiving and emotions have significant functions independently of whether they are conscious. Although, some have assumed that psychological states never occur without being conscious or at least that qualitative psychological states never do (Block, 2001<sup>[25]</sup>). If one assumes that, one will not distinguish between the function these states have independently of being conscious and the function that is added by their being conscious. One would see the function of consciousness as simply the function that conscious states have, ignoring what function might be added specifically by those states being conscious.

However, not all psychological states are conscious, and there is little reason to think that only conscious psychological states tend to benefit the organism in some significant way. Therefore, we must distinguish the function that is due specifically to a psychological state's being conscious from the function that that state has independently of its being conscious. Even for states that are conscious, we must distinguish the function that is specifically due to its being conscious from the function that results from others of its psychological properties. Restricting attention to the cognitive and desiderative states, a number of suggestions are current as to how the consciousness of these states is useful. It has been held that consciousness enhances processes of rational thought and planning, intentional action, executive function and the correction of complex reasoning. There is also an undermining reliance on evolutionary selection procedures in explaining why such states so often occur consciously in humans (Rosenthal, 1986<sup>[199]</sup>; 2005<sup>[200]</sup>; 2008<sup>[201]</sup>).

## 5.6 Is consciousness embodied in the true sense: Critical issues

It seems that more pages are published on consciousness these days than on any other subject in the philosophy of mind. Embodiment and situated cognition are also trendy. They mark a significant departure from orthodox theories, and are thus appealing to radicals and renegades. It is hardly surprising, then, that consciousness, embodiment and situated cognition have coalesced (Varela, 1996<sup>[238]</sup>). Both embodiment and situated cognition are exciting, and being exciting is an additive property. But excitement is not always correlated with truth, and the embodied and situated approach to consciousness may be easier to sell than to prove.

The term 'embodied' is most generically used to mean involving the body (Gallagher, 2005<sup>[92]</sup>). To say that a mental capacity is embodied can mean one of the two things. It can mean that the capacity depends on the possession and use of a body, not just a brain. I exclude the brain, because the brain is part of the body, and all materialists (and some dualists) believe that mental capacities involve the brain. Some embodiment theorists think other parts of the body are important as well. Other embodiment theorists do not go quite this far. Instead, they say that embodied mental capacities are the ones that depend on mental representations or processes that relate to the body (Dennett, 1991<sup>[75]</sup>). Such representations and processes come in two forms: There are representations and processes that represent or respond to the body, such as a perception of bodily movement, and there are representations and processes that affect the body, such as motor commands. We can call the first class 'somatic' and the second class 'enactive'. On this use of the term 'embodiment,' everyone agrees that, say, proprioceptive states of the central nervous system are embodied. The controversy concerns whether other forms of perception and cognition are embodied. For example, only an embodiment theorist would say that vision is embodied in any of the ways described here. To say that consciousness is embodied is to say that consciousness depends either on the existence of processes in the body outside the head or on the somatic or enactive processes that may be inside the head (Noe, 2005<sup>[173]</sup>).

Situated and embodied approaches have a tendency to drift towards excessive radicalism. Practitioners argue that orthodox conceptions of the mind will be completely undermined once we recognise a place for the body and world in mental life. I think we should resist such extremes. In issuing that warning, the bulk of this discussion has been critical, but that was not my ultimate purpose. Recent studies on self-consciousness have focussed on awareness of the acting body, and work on the unity and function of consciousness may move in the same direction. If these forecasts are right, a complete theory of consciousness will be an embodied theory, in a moderate sense of the term. A complete theory will implicate systems that are involved in representing and controlling the body. The contributions of these systems

are highly significant. They give us a sense of agency, ownership and unity. These are pervasive aspects of conscious experience. Moreover, the mechanisms that give rise to consciousness may have evolved in the service of action. If so, consciousness is not about sensing; we can do that without consciousness. Nor is consciousness about making life more pleasant or more miserable; these are just side effects. Rather consciousness is about acting — it emerges through processes that make the world available to those systems that allow us to select behavioural means to our ends. In resisting radically situated and embodied theories, we must not lose sight of this fundamental fact (Prinz, 2002<sup>[188]</sup>; Davidson and Novell, 2010<sup>[63]</sup>).

## 6. The Dark Side of Consciousness

The evidence that there is a right hemisphere bias in terms of self-awareness is overwhelming. Yet, to explain the existence of self-awareness, one might turn to an evolutionary analysis. The cost of developing and maintaining brain tissue is inherently expensive, and the addition of any brain function must result in benefit. Although this is an extreme over-simplification of the evolutionarily approach, the idea is intuitively practical for examining brain/behaviour relationships. It can be assumed that the processes of self-awareness and theory of mind require a great deal of energy to maintain. It is assumed that these processes are expensive, and an evolutionary approach would examine what benefit these abilities would provide to 'offset' such a cost. We have known that the right hemisphere is dominant for quite a number of processes, such as spatial awareness and many emotional processes (Joseph, 1988<sup>[134]</sup>). Certainly self-awareness is not the only function of the right hemisphere. However, it is quite definite that the amount of brain tissue and function dedicated to self-awareness is great. To outweigh the cost, self-awareness must provide substantial benefits.

A larger advantage would be derived from what has been termed 'cognitive goldilocks' (Malcolm and Keenan, 2003<sup>[156]</sup>). Just like the tale, cognitive goldilocks allows one to test out a variety of scenarios. Having a concrete sense of self allows an individual to cast himself or herself into the past and future. We are able to imagine ourselves in a variety of situations. This casting would have a certain benefit, as one would be able to judge beforehand the possible advantages and disadvantages of different actions in different situations. Certainly this would provide a tremendous evolutionary advantage. However, most researchers turn to the Theory of Mind to explain the benefits of self-awareness. The theory of Mind, as indicated, is the ability to model the mental state of another individual. It has been found that there is a relationship between self-awareness and Theory of Mind in both humans (Gallup, 1998<sup>[97]</sup>) and non-human primates, such that, self-awareness appears to be a necessary condition for the Theory of Mind. Furthermore, there is good evidence that tasks requiring the Theory of Mind engage the right hemisphere.

Deception, or the dark side of consciousness, may provide one of the greatest advantages when considering the costs and benefits of self-awareness and the Theory of Mind. Simply, a tremendous advantage would be bestowed upon the individual that could deceive as compared to the individual that could not deceive. Such an advantage would justify the 'cost' associated with maintaining extra brain matter. Therefore, it is postulated that one of the advantages of higher-order consciousness is the ability to deceive, as well as detect, deception. Children deceive at an extremely high rate, with deception beginning around the age that self-awareness appears (Ritblatt, 2000<sup>[191]</sup>). These studies have revealed that deception correlates with abilities on other Theory of Mind tasks, such that the better a child is at modelling mental state, the more likely he or she is to deceive. Furthermore, animals that are self-aware (such as the chimpanzee) engage in deception, while monkeys are not seen as deceivers (de Waal, 1988<sup>[69]</sup>; de Waal, 2010<sup>[70]</sup>).

Both deception and deception detection are correlated with the right hemisphere, both neuroimaging and patient reports support the hypothesis that deception may be one of the adaptive roles of the right hemisphere. Recently, researchers have begun to examine the relationship between self-awareness, Theory of Mind and the right hemisphere. Specifically, it was found that as self-awareness increased, the deception detection ability also increased. It was the right hemisphere (and not the left) that had a true relationship in terms of self-awareness and deception detection. Interestingly, it appears that the left hemisphere often fills in information that it is unaware of. This has led some to the idea that the left hemisphere is the interpreter (Gazzaniga, 1998<sup>[99]</sup>; Gazzaniga, 2012<sup>[101]</sup>; Damasio, 2012<sup>[61]</sup>). However, the filling in of the left hemisphere does not require insight, self-awareness, or any higher-order state. The left hemisphere appears to do so in a rather blind manner. Thus, the idea that consciousness is a left hemisphere phenomenon (in terms of interpretation) is not supported. The right hemisphere, in fact, truly interprets the mental state not only of its own brain, but the brains (and minds) of others.

The data support the hypothesis that the right hemisphere is dominant for higher-order consciousness. Furthermore, it is possible that the right hemisphere may have sustained such higher-order consciousness by providing deception, which would gain a significant advantage in terms of primate evolution. The dark side of consciousness, namely deception, may be an effective way of understanding the sustaining of such complex and 'expensive' phenomena as self-awareness and the Theory of Mind (Malcolm and Keenan, 2003<sup>[156]</sup>).

## 7. Limitations of a Scientific Theory of Consciousness

In biological terms, human consciousness appears as a feature associated with the functioning of the human brain. The corresponding activities of the

neural network occur strictly in accordance with physical laws; however, this fact does not necessarily imply that there can be a comprehensive scientific theory of consciousness, despite all the progress in neurobiology, neuropsychology and neurocomputation. Predictions of the extent to which such a theory may become possible vary widely in the scientific community. There are basic reasons – not only practical but also epistemological – why the brain-mind relation may never be fully decodable by general finite procedures. In particular self-referential features of consciousness, such as self-representations involved in strategic thought and dispositions, may not be resolvable in all their essential aspects by brain analysis. Assuming that such limitations exist, objective analysis by the methods of natural science cannot, in principle, fully encompass subjective, mental experience (Koch, 2012<sup>[138]</sup>).

The monitoring of brain activities in relation to conscious processes represents only one approach in the investigation into the neurobiological principles behind the conscious mind. It is supplemented by other methods addressing neural activity, connectivity and function. In addition, there is an entire spectrum of research fields – such as psychophysical methods, comparative investigations on other primates, the study of models of neural networks and the corresponding computer-based theoretical research – that may contribute to the understanding of higher brain functions, such as, pattern recognition and language, learning and memory, voluntary movement, chronological organisation of actions and many other abilities. All in all, we are increasingly able to understand many relationships between processes in the brain and processes in the mind. The mind is a functional correlate and the processes involved are interrelated. Most activities of the human brain are unconscious and routine; conscious processes are heavily influenced by unconscious pre-conditions, such as past experience and emotions. Consciousness is mediated by the cerebral cortex, and it is activated particularly by situations that are novel or that involve difficult planning or decisions. Models for the underlying integrative features of large neural networks or even the entire cortex are particularly interesting in this context (Koch, 2004<sup>[137]</sup>; Koch, 2012<sup>[138]</sup>).

Physical theories do not fulfill the promise of a complete theory of consciousness either. A further line of thought in the spectrum of opinion holds that understanding the mind-body relationship is not possible in the current state of physics, but a future expanded physics could fully resolve the problem. The world of atoms was not accessible to physics at the beginning of the twentieth century, but the new discipline of quantum mechanics – and the underlying conceptual changes and expansions of the basic laws of physics – rendered atoms and molecules understandable in physical terms. Although it is widely criticised, and is very much a minority opinion, it must be taken seriously; but the likelihood of this happening is rare. Although direct conclusions from quantum computation on consciousness, or from quantum uncertainty on free will, are inadequate, it is obvious that the material and mental aspects of modern physics are rather remote



from the full-blooded mechanistic–materialistic determinism of the nineteenth century physics (Penrose, 2005<sup>[181]</sup>; Smolin, 2007<sup>[221]</sup>; Penrose, 2010<sup>[182]</sup>).

To sum up these considerations, it is not a stringent consequence of the applicability of physics to the brain and the unique correspondence of mental states to the physical states of the brain that all behavioural dispositions will be deducible from the physical state of the brain in a finite process. We have more reason to believe that there are limits to the decidability of brain states with respect to mental states. According to everything that we know, the brain follows the same physical laws as do machines; but a machine that we were capable of understanding could not do everything like a human, and a machine that could do everything like a human would be impossible for us to fully understand. If we know the mental state of a human, expressed by means of language and gestures, we may know more than would be possible to know through a purely physical analysis of her or his brain, however elaborate that analysis may be (Penrose, 1990<sup>[179]</sup>).

Finally, the problem of consciousness is tied closely to one of the most difficult questions surrounding our understanding of ourselves: The question of free will. Conscious thought is often involved when evaluation of a situation reveals different possible behavioural pathways of comparable emotional desirability. Naturally, the question of whether and to which extent we may consider the pathways actually taken, as determined, will not be resolved solely by insights into possible limits of a theory of the brain–mind relationship; but they make a small contribution. They say that there may be principal limits to grasping the consciousness of others. The will of another, despite being tied closely to processes in his or her brain, cannot be completely decoded by an outsider, and therefore, is not objectively understandable. Outsiders cannot claim to make certain statements about our own motives if we do not voluntarily share them. Luckily there are limits to intruding into the consciousness of others, and there is unfortunately often too little modesty and reservation when judging the motivations of others. In fact, complete mind reading is beyond human capabilities (Edelman, 2003<sup>[79]</sup>).

What the uncertainty of thoughts does have in common with the uncertainty of particles is that the difficulty is not just a practical one, but a systematic limitation, which cannot, even in theory, be circumvented. It is patently not resolved by the efforts of psychologists and psychoanalysts, and it will not be resolved by neurologists either, even when everything is known about the structure and workings of our brain (Gierer, 2008<sup>[103]</sup>).

Despite all the limitations and fallacies of various theories and the lack of consensus among various consciousness theorists, there is another problem. All these theorists, although from diverse disciplines, rarely communicate with one another, a problem that is endemic across all sciences. It is prudent that experts from various fields meet, encourage interdisciplinary dialogue and understand

each other's discipline with an aim to solve the ultimate mystery of consciousness and to come up with a unified theory of consciousness. For this, various inroads have already been made, but we must now join the highway, and there is no stopping now. We have to relinquish personal gains and quell primitive fears, while growing together to solve the riddle of consciousness.

### 8. Final Conclusions [See also Figure 1, Flowchart of Article]

The theories reviewed in this article have looked at certain domains of consciousness. I have looked at the contributions of self-psychology in consciousness

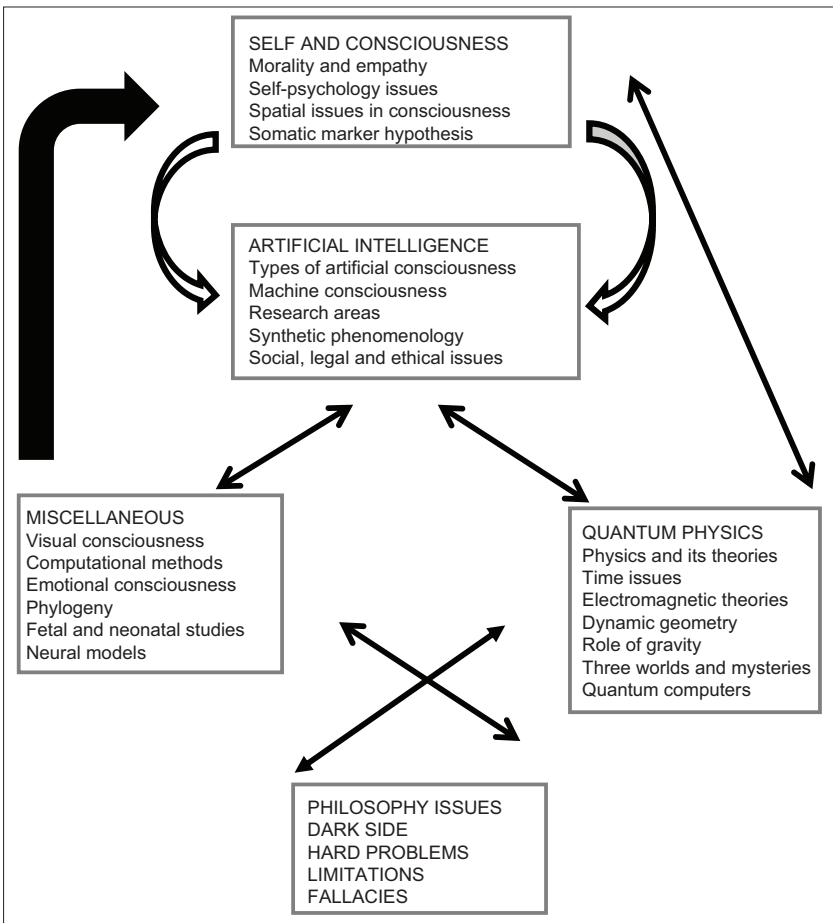


Figure 1: Flow chart of the article

and then moved to how artificial intelligence paradigms define consciousness, conscious processes and the machine correlates of consciousness. There have been advances in the role of quantum physics in consciousness, covering various physical models of consciousness. Neurophilosophy with its recent advances has added to our knowledge and cleared certain cobwebs in this area.

### **Take home message**

Various subspecialties like quantum physics, philosophy, self-psychology, artificial intelligence, visual neurobiology, emotions and their psychology and developmental psychology have contributed to the development of an integrated theory of consciousness and must be studied together with neurobiology, to gain a total understanding of consciousness.

### **Conflict of interest**

None declared.

### **Declaration**

This is my original, unpublished work, which has not been submitted for publication elsewhere.

### **Acknowledgements**

The author would like to thank Dr. Ajai Singh for his guidance and encouragement throughout the preparation of this article.

## **References**

1. Abhedananda S. True Psychology. Kolkata: Ramakrishna Vedanta Math; 2008.
2. Abhedananda S. Notes on Philosophy and Religion. Kolkata: Ramakrishna Vedanta Math; 2009.
3. Aleksander I. How to build a Mind: Machines with Imagination. London: Weidenfeld and Nicolson; 2000.
4. Aleksander I. The World in My Mind, My Mind in the World: Key mechanisms of consciousness in People, Animals and Machines. Exeter: Imprint Academic; 2005.
5. Aleksander I, Morton H. Computational studies of consciousness. Prog Brain Res 2008;168:77-93.
6. Aleksander I, Gamez D. Informational theories of consciousness: A review and extension. Adv Exp Med Biol 2011;718:139-47.
7. Altschuler EL, Ramachandran VS. A simple method to stand outside oneself. Perception 2007;36:632-4.
8. Aharanov Y, Bohm D. Significance of electromagnetic potentials in quantum theory. Phys Rev 1959;115:485-91.
9. Amodio DM, Ratner KG. A memory systems model of implicit social cognition. Curr Dir Psychol Sci 2011;20:143-8.

10. Antonov AA. From artificial intelligence to human super intelligence. *Int J Comp Info Sys* 2011;2:1-6.
11. Arico A, Fiala B, Goldberg RF, Nichols S. The folk psychology of consciousness. *Mind Lang* 2011;26:327-52.
12. Arnold DH, Clifford CW. Determinants of an asynchronous process in vision. *Proc Royal Soc London* 2002;269:579-83.
13. Baars BJ. *A Cognitive Theory of Consciousness*. Cambridge: Cambridge University Press; 1998.
14. Baars BJ. The global workspace theory of consciousness: Towards a cognitive neuroscience of human experience. *Prog Brain Res* 2005;150:45-53.
15. Baianu IC, Brown R, Glazebrook JF. A category theory and higher dimensional algebra approach to complex system biology, meta systems and an ontological theory of levels: Emergence of life, society, human consciousness and artificial intelligence. *Acta Univers Apul* 2011;(Suppl 1):176-99.
16. Barbour J. *The End of Time: The Next Revolution in Physics*. Oxford: Oxford University Press; 1999.
17. Bartels A, Zeki S. The architecture of column cells in human visual brain: New results and review. *Eur J Neurosci* 2000;12:172-93.
18. Bechtel W, Abrahamesen AA. Explanation: A mechanistic alternative. *Stud Hist Philos Biol Biomed Sci* 2005;36:421-41.
19. Bermudez J. The moral significance of birth. *Ethics* 2005;106:378-403.
20. Bernat JL. Chronic disorders of consciousness. *Lancet* 2006;367:1181-92.
21. Bird CM, Burgess N. The hippocampus and memory: Insights from spatial processing. *Nat Rev Neurosci* 2008;9:182-94.
22. Blanke O, Ortrigue S, Landis T, Seeck M. Simulating illusory own body perceptions. *Nature* 2002;419:269-70.
23. Blanke O, Landis T, Spinelli L, Seeck M. Out of body experience and the autoscapy of neurological origin. *Brain* 2004;127:243-58.
24. Blanke O, Mohr C, Michel CM, Pascal Leone A, Brugger P, Seeck M. Linking out of body experience and self processing to mental own body imagery at the temporo-parietal junction. *J Neurosci* 2005;25:550-7.
25. Block N. Paradox and cross purposes in recent work on consciousness, *Cognition* 2001;79:197-219.
26. Block N. Perceptual consciousness overflows cognitive access. *Trends Cogn Sci* 2011;15:567-75.
27. Bohm D, Hiley BJ. *The Undivided Universe*. London: Routledge; 1993.
28. Bourgeois JP. Synaptogenesis, heterochrony and epigenesis of the mammalian neocortex. *Acta Pediatr Suppl* 1997;422:27-33.
29. Burghardt GM. Conceptions of play and evolution of animal minds. *Evoln Cogn* 1999;5:114-22.
30. Buttazzo G. Artificial consciousness: Hazardous questions and answers. *Artif Intell Med* 2008;44:139-46.
31. Cabanac M. On the origin of consciousness: A postulate and its corollary. *Neurosci Biobehav Res* 1996;20:33-40.
32. Cabanac M. Emotion and phylogeny. *Jpn J Physiol* 1999;49:1-10.
33. Cabanac M, Cabanac A, Parent A. The emergence of consciousness in phylogeny. *Behav Brain Res* 2009;198:267-72.
34. Cabeza R, Jacques StP. Functional neuroimaging of autobiographical memory. *Trends Cogn Sci* 2007;11:219-27.
35. Calverley DJ. Imagining a non biological machine as a legal person. *AI and Society* 2008;22:523-37.
36. Cardin V, Friston KJ, Zeki S. Top-down Modulations in the Visual Form Pathway Revealed with Dynamic Causal Modelling. *Cerebr Corte* 2011;21:550-62.

37. Carroll S. *From Eternity to Here: A Quest for the Ultimate Theory of Time*. New York: Penguin Books; 2010.
38. Carruthers P. *Language, Thought and Consciousness*. Cambridge: Cambridge University Press; 1996.
39. Carruthers P. *Phenomenal Consciousness*. Cambridge: Cambridge University Press; 2000.
40. Chalmers DJ. Facing up to the problem of consciousness. *J Conscious Stud* 1995;2:200-19.
41. Chalmers DJ. *The Conscious Mind: In Search of A Fundamental Theory*. Oxford: Oxford University Press; 1996.
42. Chella A, Macaluso I. Sensations and perception in Cicer robot: A museum guide robot, Proceedings of BICS 2006, Lesbos, Greece; 2006.
43. Chrisley RJ, Holland A. Connectionist Synthetic Epistemology: Requirements for the development of objectivity. *COGS CSRP* 1994;353:1-21.
44. Chrisley RJ. *Artificial Intelligence: Critical Concepts*. Routledge, London; 2000.
45. Chrisley RJ, Parthermore J. Synthetic phenomenology: Exploiting embodiment to specify the non conceptual content of visual experience. *J Conscious Stud* 2007;14:44-58.
46. Chrisley R. Philosophical foundations of artificial consciousness. *Artif Intell Med* 2008;44:119-37.
47. Churchland PS. *Brain-Wise: Studies in Neurophilosophy*. Cambridge MA: MIT Press; 2002.
48. Clarks A, Chalmers D. The extended mind. *Analysis* 1998;58:7-19.
49. Cohen MA, Dennett DC. Consciousness cannot be separated from function. *Trends Cogn Sci* 2011;15:358-64.
50. Cote F, Fligny C, Bayard E, Launay JM, Gershon MD, Mallet J, *et al*. Maternal serotonin is crucial for murine embryonic development. *Proc Natl Acad Sci USA* 2007;104:329-34.
51. Cotterill R. *Enchanted Looms*. Cambridge: Cambridge University Press; 2000.
52. Coward LA. The Recommendation Architecture: Lessons from the design of large scale electronic systems for cognitive science. *J Cogn Sys Res* 2001;2:111-56.
53. Coward LA, Sun R. Some criteria for an effective scientific theory of consciousness and examples of a preliminary attempt at such a theory. *Conscious Cogn* 2004;13:268-301.
54. Coward LA. *A System Architecture approach to the Brain: From Neurons to Consciousness*. New York: Nova Science Publishers; 2005.
55. Craig AD. How do you feel now: The anterior insula and human awareness. *Nat Rev Neurosci* 2009;10:59-70.
56. Crane T. *Elements of the Mind*. Oxford: Oxford University Press; 2001.
57. Crick F. *The Astonishing Hypothesis*. London: Simon and Schuster; 1994.
58. Dalgeish T. The emotional brain. *Nat Rev Neurosci* 2004;5:583-9.
59. Damasio AR. *The Feeling of What Happens: Body, Emotion and The Making of Consciousness*. London: Vintage Books; 2000.
60. Damasio AR. Feelings of emotion and the self. *Ann NY Acad Sci* 2003;1001:253-61.
61. Damasio AR. *Self comes to Mind: Constructing the conscious brain*. UK: Vintage Books; 2012.
62. Darwin C. *The Expression of Emotions in Man and Animals*. UK: John Murray; 1889.
63. Davidson I, Nowell A. *Stone Tools and the Evolution of Human Cognition*. University Press of Colorado; 2010.
64. Davies P. *About Time: Einstein's Unfinished Revolution*. Vintage Books; 1995.
65. Dawkins MS. *Through Our Eyes Only*. Oxford: W. H. Freeman and Co; 1993.
66. De Ridder D, Van Laere K, Dupont P, Menovsky T, Van de Henyning P. Visualizing out of body experience in the brain. *N Engl J Med* 2007;357:1829-33.
67. De Sousa A. Freudian theory and consciousness: A conceptual analysis. *Mens Sana Monogr* 2011;9:210-7.
68. De Sousa A. Towards an integrative theory of consciousness Part 1: (Neurobiological and cognitive models). *Mens Sana Monogr* 2013;11:100-50
69. De Waal F. *Good Natured: The Origins of Right and Wrong in Humans and Other Animals*. Cambridge, MA: Harvard University Press; 1996.

70. De Waal F. The age of empathy: Nature's lessons for a kinder world. US: Three Rivers Press; 2010.
71. Decety J, Sommerville JA. Shared representations between the self and others: a social cognitive neuroscience view. *Trends Cogn Sci* 2003;7: 527-33.
72. Decety J, Grezes J. The power of simulation: Imagining one's own and others behaviour. *Brain Res* 2006;1079:4-14.
73. Dennett D. Conditions of Personhood. In Rorty A., editor, *The Identities of Persons*. Berkeley: University of California Press; 1976.
74. Dennett DC. *The Intentional Stance*. Cambridge: MIT Press; 1987.
75. Dennett DC. *Consciousness Explained*. UK: Boston, Little, Brown and Company; 1991.
76. Denton D. *The Pinnacle of Life*. New York: Harper Collins; 1994.
77. Di Francesco M. Consciousness and the self. *Func Neurol* 2008;23:179-87.
78. Duch W. Brain inspired conscious computing architecture. *J Mind Behav* 2005;26:1-22.
79. Edelman GM. Naturalizing consciousness: A theoretical framework. *Proc Natl Acad Sci USA* 2003;100:5520-24.
80. Ehrsson HH. The experimental induction of out of body experiences. *Science* 2007;317:1048.
81. Eimer M, Holmes A. Event related brain potential correlates of emotional face processing. *Neuropsychologia* 2007;45:15-31.
82. Ekman P. *Emotions Revealed: Recognizing faces and feelings to improve communication and emotional life*. New York: Henry Holt; 2003.
83. Engel AK, Fries P, Konig P, Brecht M, Singer W. Temporal binding, binocular rivalry and consciousness. *Conscious Cogn* 1999;8:128-51.
84. Farhenfort JJ, Lamme VA. A true science of consciousness explains phenomenology: Comment on Cohen and Dennett. *Trends Cogn Sci* 2012;16:138-9.
85. Ffytche DH, Zeki S. The primary visual cortex, and feedback to it, are not necessary for conscious vision. *Brain* 2011;134:247-57.
86. Frank A. *About Time*. UK: One World Books; 2011.
87. Franklin S, Baars BJ, Ramamurthy U, Ventura M. The role of consciousness in memory. *Brain Minds Media* 2005;1:1-38.
88. Freud S. *The Psychopathology of Everyday Life*. New York: Norton; 1907.
89. Freud S. *The Unconscious*. Standard Edition. London: Hogarth Press; 1915.
90. Freud S. *The Interpretation of Dreams*, Standard Edition. London: Hogarth; 1925.
91. Gallagher S. The moral significance of primitive self consciousness: A response to Bermudez. *Ethics* 1996;107:129-40.
92. Gallagher S. *How The Body Shapes The Mind*. Oxford: Oxford University Press; 2005.
93. Gallagher S. The narrative alternative to the theory of mind. *Conscious Emotion* 2006;7:223-9.
94. Gallagher S, Marcel AJ. The self in contextualized action, *J Conscious Stud* 1999;6:4-30.
95. Gallagher HL, Frith CD. Functional imaging of theory of mind. *Trends Cogn Sci* 2003;7:77-83.
96. Gallese V. The shared manifold hypothesis: From mirror neurons to empathy. *J Conscious Stud* 2001;8:33-50.
97. Gallup GG. Self awareness and the evolution of social intelligence. *Behav Process* 1998;42:239-47.
98. Gamez D. Progress in machine consciousness. *Conscious Cogn* 2008;17: 887-910.
99. Gazzaniga M. Brain and conscious experience. *Adv Neurol* 1998;77:181-92.
100. Gazzaniga M. *The Ethical Brain*. Washington: Dana Press; 2006.
101. Gazzaniga M, Larkin P. *Who's in Charge: Free Will and the Science of the Brain*. USA: Tantor Press; 2012.
102. Gennaro RJ. *Higher Order Theories of Consciousness*. Amsterdam: John Benjamins; 2004.
103. Gierer A. Brain, mind and limitations of a scientific theory of consciousness. *Bio Essays* 2008;30:499-505
104. Goldman AI. *Simulating Minds*. Oxford: Oxford University Press; 2006.
105. Goertzel B, Pennachin C. *Artificial General Intelligence*. Berlin: Springer; 2007.

106. Gould JL, Grant-Gould C. *The Honey Bee*. New York: Scientific American Library; 1995.
107. Gray JA. The contents of consciousness: A neurophysiological conjecture. *Behav Brain Sci* 1995;150:11-23.
108. Griffin DR. *Animal Minds*. Chicago: University of Chicago Press; 1992.
109. Grush R, Churchland PC. Gaps in Penrose's toiling. *J Conscious Stud* 1995;2:10-29.
110. Haikonen PO. *The Cognitive Approach to Conscious Machines*. Exeter: Imprint Academic; 2003.
111. Haikonen PO. Towards times of miracles and wonder: a model for the conscious machine. *Proceedings of BICS 2006*, Lesbos: Greece; 2006.
112. Hameroff SR, Kazniak A, Scott AC. *Towards a Science of Consciousness: The First Tuscon Discussions and Debates*. Cambridge: MIT Press; 1996.
113. Hari S. Psychons could be zero energy tachyons. *Neuro Quantology* 2008;6:152-60.
114. Hari S. Consciousness, mind and matter in Indian Philosophy. *J Conscious Explor Res* 2010;1:640-50.
115. Hauser M. *Moral Minds*. New York: Springer Press; 2006.
116. Hesslow G. Conscious thought as simulation of behaviour and perception. *Trends Cogn Sci* 2002;6:242-7.
117. Hesslow G, Jirenhed DA. The inner world of a simple robot. *J Conscious Stud* 2007;14:85-96.
118. Heyes CM. Theory of mind in non human primates. *Anim Behav* 1994;47: 909-19.
119. Holland O, Goodman R. Robots with internal models: A route to machine consciousness. *J Conscious Stud* 2003;10:77-109.
120. Holland O. *Machine Consciousness*. Exeter: Imprint Academic; 2003.
121. Holt RR. *Freud Reappraised: A fresh look at psychoanalytical theory*. New York: Guilford Press; 1989.
122. Hubel DH, Weisel TN. The Ferrier lecture: Functional architecture of the macaque monkey visual cortex. *Proc Royal Soc London B* 1977;198:1-59.
123. Humphrey N. *Seeing Red: A Study In Consciousness*. Cambridge MA: Harvard University Press; 2006.
124. Humphrey N. *Soul Dust: The Magic of Consciousness*. UK: Quercus Press; 2011.
125. Hurley S. *Consciousness In Action*. Cambridge: Harvard University Press; 1998.
126. Husserl E. *Cartesian Meditations*. The Hague: Nijhoff; 1960.
127. Husserl E. *The Phenomenology of Internal Time Consciousness*. The Hague: Nijhoff; 1964.
128. Hu H, Wu M. Spin mediated consciousness theory. *Med Hypotheses* 2004;63:633-46.
129. Hu H, Wu M. Thinking outside the box: The essence and implications of quantum entanglement. *Neuro Quantology* 2006;4:5-16.
130. Ito M, Miyashita Y, Rolls ET. *Cognition, Computation and Consciousness*. Oxford: Oxford University Press; 1997.
131. James W. *The Principles of Psychology*. London: Macmillan; 1890.
132. Johnson MH. Subcortical face processing. *Nat Rev Neurosci* 2005;6:766-74.
133. Jordan JS. Synthetic phenomenology? Perhaps but not via information processing, Talk given at the Max Planck Institute for Psychological Research, Munich, Germany; 1998.
134. Joseph R. The right cerebral hemisphere: Emotion, music, visuospatial skills, body image, dreams and awareness. *J Clin Psychol* 1988;44:630-73.
135. Kamitani Y, Tong F. Decoding the visual and subjective contents of the human brain. *Nat Neurosci* 2005;8:679-85.
136. Kawabata H, Zeki S. The neural correlates of beauty. *J Neurophysiol* 2004;91:1699-705.
137. Koch C. *The Quest for Consciousness: A Neurobiological Approach*. Roberts and Publishers; 2004.
138. Koch C. *Consciousness: Confessions of a Romantic Reductionist*. Cambridge: MIT Press; 2012.
139. Kostovic I, Milosevic N. The development of cerebral connections in the first 20-45 weeks of gestation. *Semin Fetal Neonatal Med* 2006;11:415-22.



140. Kurzweil R. *The Age of Spiritual Machines*. London: Penguin Putnam; 2000.
141. Lagercrantz H, Hanson M, Evrard P, Rodeck C. *The Newborn Brain*. Cambridge: Cambridge University Press; 2002.
142. Lagercrantz H. The emergence of the mind: A borderline of human viability. *Acta Paediatr* 2007;96:327-8.
143. Lagercrantz H, Changeux JP. The emergence of human consciousness: From fetal to neonatal life. *Pediatr Res* 2009;65:255-60.
144. Lamme VA. Towards a neural stance on consciousness. *Trends Cogn Sci* 2006;10:494-501.
145. LaRock E. Disambiguation, binding and the unity of visual consciousness. *Theory Psychol* 2007;17:747-77.
146. Le Doux J. The human amygdala. *Curr Biol* 2007;17:868-74.
147. Lenggenhager B, Tadi T, Metzinger T, Blanke O. Video ergo sum: Manipulating bodily self consciousness. *Science* 2007;317:1096-9.
148. Letnic K, Zoncu R, Rakic P. Origin of GABAergic neurons in the human neocortex. *Nature* 2002;417:645-9.
149. Levy DN. *Robots unlimited: life in a virtual age*. UK: Vintage Books; 2006.
150. Lewis M, Haviland-Jones JM, Barrett LF. *Handbook of Emotions*. New York: Guilford Press; 2008.
151. Leznik E, Makarenko V, Llinas R. Electronically mediated oscillatory patterns in neuronal ensembles: An *in vitro* voltage dependent dye imaging study in the inferior olive. *J Neurosci* 2002;22:2804-15.
152. Light S, Zahn-Waxler C. *Empathy*. Cambridge: MIT Press; 2011
153. Logothetis NK, Pauls J, Poggio T. Shape representation in the inferior temporal cortex of the monkey. *Curr Biol* 1995;5:552-63.
154. Logothetis NK. Single units and conscious vision. *Philos Trans R Soc Lond B Biol Sci* 1998;353:1801-18.
155. Mach E. *The Science of Mechanics*. German: German Press; 1960.
156. Malcolm S, Keenan JP. Deception, detection and hemispheric differences in self awareness. *Soc Behav Personality* 2003;31:767-72.
157. Malloy I. *Essays in Cognitive Science: Collected articles on morality and consciousness*. USA: Dorrance Publishers; 2011.
158. Marian DE, Shimamura AP. Emotions in context: Pictorial influences on affective attributions. *Emotion* 2012;12:371-5.
159. Marsh L, Robbins P. Consciousness and the social mind. *Cogn Sys Res* 2008;9:15-23.
160. Matzke DJ. Consciousness: A computational paradigm update. *Acta Nerv Sup* 2010;52:134-40.
161. McFadden J. Synchronous firing and its influence on the earth's magnetic field: Evidence for an electromagnetic field theory of consciousness. *J Conscious Stud* 2002;9:23-50.
162. McFarland D. Defining motivation and cognition in animals. *Int Stud Philos Sci* 1991;5:153-70.
163. McIntosh DN, Reichmann-Decker A, Winkelman P, Willbarger J. When the social mirror breaks: deficits in automatic, not voluntary, mimicry of emotional facial expressions in autism. *Dev Sci* 2006;9:295-302.
164. Merker B. The liabilities of mobility: A selection pressure for transition to consciousness in animal evolution. *Conscious Cogn* 2005;14:89-114.
165. Metzinger T. *Being No One*. Cambridge: MIT Press; 2003.
166. Morsella E. The function of phenomenal states: Supramodular interaction theory. *Psychol Rev* 2005;112:1000-21.
167. Moutoussis K, Zeki S. Relation between cortical activation and perceptual investigation with invisible stimuli. *Proc Natl Acad Sc USA* 2002;99: 9527-32.
168. Mukherjee BD. *The Essence of the Bhagavad Gita*. Kolkata: Academic Publishers; 2002.
169. Mulhauser G. *Mind out of Matter*. Dordrecht, Kluwer Academic Publishers; 1998.
170. Nayak S, Nayak S, Singh JP. An introduction to Quantum Neural Computing. *J Glob Res Comp Sci* 2011;2:50-5.

171. Nehaniv C. *Computation for Metaphors, Analogy and Agents: Vol 1562 of Springer Lecture Notes on Artificial Intelligence*. Berlin: Springer-Verlag; 1998.
172. Nijhuis JG. Fetal behaviour: The brain and behaviour in different stages of human life. *Neurobiol Aging* 2003;24:541-6.
173. Noe A. *Action in Perception*. Cambridge MA: MIT Press; 2005.
174. Nowakowski R. Stable neuron numbers from cradle to grave. *Proc Natl Acad Sci USA* 2006;103:12219-20.
175. Panksepp J. *Affective Neuroscience*. Oxford: Oxford University Press; 1998.
176. Panksepp J. *Affective consciousness: Core emotional feelings in humans and animals*. *Conscious Cogn* 2005;14:30-80.
177. Papineau D. *Thinking About Consciousness*. Oxford: Oxford University Press; 2002.
178. Pellionisz A, Llinas R. Tensor network theory of the metaorganization of functional geometries in the CNS. *Neurosci* 1985;16:245-73.
179. Penrose R. *The Emperor's New Mind*. Vintage: London; 1990.
180. Penrose R. *Shadows of The Mind: A Search For The Missing Science Of Consciousness*. Oxford: Oxford University Press; 1994.
181. Penrose R. *The Road to Reality: A Complete Guide to the Laws of the Universe*. New York: Vintage Books; 2005.
182. Penrose R. *Cycles of Time: An Extraordinary New View of the Universe*. Word Press; 2010.
183. Phillips P. *An introduction to artificial intelligence: Have the thinking machines arrived. Perspectives In Computer Science: Karlstad University Monographs*; 2011.
184. Pockett S. Difficulties with an electromagnetic field theory of consciousness. *J Conscious Stud* 2002;9:51-6.
185. Pockett S. Difficulties with an electromagnetic field theory of consciousness: An update. *Neuro Quantology* 2007;5:271-5.
186. Prechtl HF. Ultrasound studies of human fetal behaviour. *Early Hum Dev* 1985;12:91-8.
187. Prinz J. A neurofunctional theory of visual consciousness. *Conscious Cogn* 2000;9:243-59.
188. Prinz, J. *Furnishing the Mind: Concepts and Their Perceptual Basis*. Cambridge MA: MIT Press; 2002.
189. Prinz J. *Gut Reactions*. Oxford: Oxford University Press; 2004.
190. Randall L. *Knocking on Heavens Door: How Physics and Scientific Thinking Illuminate the Universe and Modern World*. UK: Ecco Books; 2011.
191. Ritblatt SN. Children's level of participation in a false belief task, age and theory of mind. *J Genet Psychol* 2000;161:53-64.
192. Robbins P. Jack AI. The phenomenal stance. *Philos Stud* 2006;127:59-85.
193. Rolls ET. *Precis of the brain and emotion*. *Behav Brain Sci* 2000;23:177-233.
194. Rolls ET. The functions of the orbitofrontal cortex. *Brain Cogn* 2004;55:11-29.
195. Rolls ET. *Emotion Explained*. Oxford: Oxford University Press; 2005.
196. Rolls ET. A computational neuroscience approach to consciousness. *Neural Netw* 2007;20:962-82.
197. Rolls ET. *Memory, Attention and Decision Making: A Unifying Computational Neuroscience Approach*. Oxford: Oxford University Press; 2008.
198. Rolls ET, Deco G. *Computational Neuroscience Of Vision*. Oxford: Oxford University Press; 2002.
199. Rosenthal DM. Two concepts of consciousness. *Philos Stud* 1986;49:329-59.
200. Rosenthal DM. *Consciousness And Mind*. Oxford: Oxford University Press; 2005.
201. Rosenthal DM. Consciousness and its function. *Neuropsychologia* 2008;46:829-40.
202. Roy S, Llinas R. Dynamic geometry, brain function modelling and consciousness. In: Bannerjee R, Chakrabarti S, editors, *Progress in Brain Research*. Vol. 168. Netherlands: Elsevier; 2008.
203. Rushworth MF, Behrens TE. Choice uncertainty and value in the prefrontal and cingulate cortex. *Nat Neurosci* 2008;11:389-97.

204. Russell JA. Culture and the categorization of emotions. *Psychol Rev* 1991;110:426-50.
205. Russell JA. Core affect and the psychological construction of emotion. *Psychol Rev* 2003;110:145-72.
206. Scharf C. *Gravity's Engines: Bubble-Blowing Black Holes Rule Galaxies, Stars, and Life in the Cosmos*. USA: Scientific Press; 2012.
207. Searle JR. Minds, brains and programs. *Behav Brain Sci* 1980;3:417-57.
208. Searle JR. *The Rediscovery of The Mind*. Cambridge: MIT Press; 1992.
209. Searle JR. Consciousness. *Annu Rev Neurosci* 2000;23:557-78.
210. Shahnahan M. A cognitive architecture that combines internal simulation with a global workspace. *Conscious Cogn* 2006;15:433-49.
211. Siddharth BG. *The Chaotic Universe*. New York: Nova Science; 2001.
212. Silverman MP. *Quantum Superposition: Consequences of Coherence and Entanglement*. Berlin: Springer; 2008.
213. Simpson J, Weiner E. *Oxford English Dictionary*. UK: Oxford University Press; 2009.
214. Singer W. Neuronal synchrony: A versatile code for the definition of relations. *Neuron* 1999;24:49-65.
215. Singer T. The neuronal basis and the ontogeny of empathy and mind reading: Review of literature and implications for further research. *Neurosci Biobehav Rev* 2006;20:855-63.
216. Singh AR, Singh SA. Notes on a few issues in the philosophy of psychiatry. *Mens Sana Monogr* 2009;7:128-83.
217. Singh AR, Singh SA. Brain-Mind dyad, human experience, the consciousness tetrad and lattice of mental operations: And further, the need to integrate knowledge from diverse disciplines. *Mens Sana Mongr* 2011;9:6-41.
218. Singh AR, Goodwin R, Mograbi GJ, Balasubramanian R, Garyali V. A Discussion in the mind brain consciousness group 2010-2011: Let's study the structure that is the very raison de etre of our existence. *Mens Sana Monogr* 2011;9:284-93.
219. Sloman A, Chrisley R. Virtual machines and consciousness. *J Conscious Stud* 2003;10:113-72.
220. Smith Q, Jokic A. *Consciousness: New Philosophical Perspectives*. Oxford: Clarendon Press; 2003.
221. Smolin L. *The Trouble with Physics*. UK: Penguin Books; 2006.
222. Smythies J. Space, time and consciousness. *J Conscious Stud* 2003;10:47-56.
223. Starzyk JA, Prasad DK. A computational model of machine consciousness. *Int J Mach Conscn* 2011;1:156-66.
224. Stein JD. *Cosmic numbers: The numbers that define our universe*. UK: A1 Books; 2011
225. Sun R. Accounting for the computational basis of consciousness: A connectionist approach. *Conscious Cogn* 1999;8:529-65.
226. Sun R. *Duality of the Mind*. Lawrence Erlbaum Associates Mahwah: NJ; 2002.
227. Sun R, Coward LA, Zenzen MJ. On levels of cognitive modelling. *Philos Psychol* 2005;18:613-37.
228. Swami R. *Wisdom of the Ancient Sages: The Mundaka Upanishad*. Himalayan International Institute of Yoga, Science and Philosophy: USA; 1990.
229. Tegmark M. The importance of quantum decoherence in brain processes. *Phys Rev* 2000;61E:4194.
230. Thagard P. *HOT thought: Mechanisms and Applications of Emotional Cognition*. Cambridge MA: MIT Press; 2006.
231. Thagard P, Aubie B. Emotional consciousness: A neural model of how cognitive appraisal and somatic perception interact to provide a qualitative experience. *Conscious Cogn* 2008;17:811-34.
232. Tootell RB, Taylor JB. Anatomical evidence for MT and additional cortical visual areas in humans. *Cereb Cortex* 1995;5:39-55.
233. Tong F, Nakayama K, Vaughan JT, Kanwisher N. Binocular rivalry and visual awareness in the human extra-striate cortex. *Neuron* 1999;21:753-9.
234. Tononi G. An information integration theory of consciousness. *BMC Neurosci* 2004;5:42-52.

235. Treisman A. The binding problem. *Curr Opin Neurobiol* 1996;6:171-8.
236. Turiel E. *The Development of Social Knowledge*. Cambridge: Cambridge University Press; 1983.
237. Vanhatalo S, Kaila K. Developmental of neonatal EEG activity: From phenomenology to physiology. *Semin Fetal Neonatal Med* 2006;11:471-8.
238. Varela F. Neurophenomenology: A methodological remedy for the hard problem. *J Conscious Stud* 1996;3:330-49.
239. Von der Hydt R. Approaches to visual cortical function. *Rev Physiol Biochem Pharm (Berl)* 1987;108:69-150.
240. Wallach W, Allen C, Franklin S. Consciousness and Ethics: Artificially conscious moral agents. *Int J Mach Conscn* 2011;3:177-92.
241. Weiskrantz L. *Consciousness Lost And Found*, Oxford, Oxford University Press; 1997.
242. Wilkes KV. *Real People: Personal Identity without thought experiments*. Oxford: Clarendon Press, New York: Oxford University Press; 1988.
243. Winkielman P, Schooler JW. Splitting consciousness: Unconscious, conscious, and metacognitive processes in social cognition. *Eur Rev Soc Psychol* 2011;22:1-35.
244. Yang MHI. Me, My "Self" and You: Neuropsychological relations between social emotion, self-awareness, and morality. *Emot Rev* 2011;3:313-6.
245. Zahavi D. *Subjectivity and Selfhood: Investigating the first person perspective*. Cambridge: MIT Press; 2005.
246. Zeki S. Uniformity and diversity of the structure and function of the rhesus monkey prestriate visual cortex. *J Physiol* 1978;277:273-90.
247. Zeki S. A century of cerebral achromatopsia. *Brain* 1990;113:1721-77.
248. Zeki S. *A Vision of the Brain*. Blackwell: Oxford; 1993.
249. Zeki S, Marini L. 3 cortical stages of color processing in the human brain. *Brain* 1998;121:1669-85.
250. Zeki S, Bartels A. Towards a theory of visual consciousness. *Conscious Cogn* 1999;8:225-59.
251. Zeki S. *Inner Vision: An Exploration of Art and The Brain*. Oxford: Oxford University Press; 1999.
252. Zeki S, Bartels A. The process of kinetic continuity in the brain. *Cereb Cortex* 2003;13:189-202.
253. Zeki, S. The disunity of consciousness. *Trends Cogn Sci* 2003;7:214-8.
254. Zeki S. The neurology of ambiguity. *Conscious Cogn* 2004;13:173-96.
255. Zeki S. *Spelndours and Miseris of the Brain: Love, Creativity and the Quest for Human Happiness*. Wiley-Blackwell; 2009.
256. Zelazo PD. The development of conscious control in childhood. *Trends Cogn Sci* 2004;8:12-7.

## Questions that this Paper Raises

1. Is an integrated theory encompassing all fields of consciousness possible?
2. Is it possible to understand the true essence of consciousness with a grand theory that unifies all the sciences and stakeholders?
3. Does the study of quantum physics, mathematics and their recent advances hold the key to the mystery of consciousness?
4. Should philosophers sit with scientists and add their views to various theories of consciousness?

5. Will science ever be able to end the debate on the mystery of consciousness and explain what it really means?

#### About the Author



*Dr. Avinash De Sousa is a consultant psychiatrist and psychotherapist with a private practice in Mumbai. He is an avid reader and has over 130 publications in national and international journals. His main areas of research interest are alcohol dependence, child and adolescent psychiatry, transcranial magnetic stimulation and consciousness. He is also the academic director of the Institute of Psychotherapy Training and Management, Mumbai. He teaches psychiatry, child psychology and psychotherapy in over 25 institutions, as a visiting faculty. He is one of the few psychiatrists, who in addition to a psychiatry degree, has an MBA in Human Resource Development, a Masters in Psychotherapy and Counselling and an MPhil in Psychology.*