



# Can artificially intelligent agents really be conscious?

SUBHASH CHANDRA PANDEY

Department of Computer Science and Engineering, Birla Institute of Technology, Mesra, Ranchi (Patna Campus), Patna, India  
e-mail: subh63@yahoo.co.in

MS received 16 February 2017; revised 4 September 2017; accepted 6 October 2017; published online 22 June 2018

**Abstract.** Mind and intelligence are closely related with the consciousness. Indeed, artificial intelligence (AI) is the most promising avenue towards artificial consciousness (AC). However, in literature, consciousness has been considered as the least amenable to being understood or replicated by AI. Further, computational theories of mind (CTMs) render the mind as a computational system and it is treated as a substantial hypothesis within the purview of AI. However, the consciousness, which is a phenomenon of mind, is partially tackled by this theory and it seems that the CTM is not corroborated considerably in this pursuit. Many valuable contributions have been incorporated by the researchers working strenuously in this domain. However, there is still scarcity of globally accepted computational models of consciousness that can be used to design conscious intelligent machines. The contributions of the researchers entail consciousness as a vague, incomplete and human-centred entity. In this paper, attempt has been made to analyse different theoretical and intricate issues pertaining to mind, intelligence and AC. Moreover, this paper discusses different computational models of the consciousness and critically analyses the possibility of generating the machine consciousness as well as identifying the characteristics of conscious machine. Further, different inquisitive questions, e.g., “Is it possible to devise, project and build a conscious machine?”, “Will artificially conscious machines be able to surpass the functioning of artificially intelligent machines?” and “Does consciousness reflect a peculiar way of information processing?” are analysed.

**Keywords.** Mind; artificial intelligence; consciousness models; machine consciousness.

## 1. Introduction

Since 1949, researchers are striving to construct the foundation of the forthcoming information age [1] and in this pursuit they envisaged various strategies to replicate the cognitive and mental capabilities of biological beings in computation, cognition, artificial intelligence (AI) and in engineering. Broadly, various approaches used can be classified as follows: (1) the symbolic and logical approach of classic AI [2, 3], (2) the sensory-motor approach [4], (3) neural-network-oriented design [5], (4) the bio-inspired strategy [6] and (5) the classic AI approach [3]. The basic characteristic of all these approaches is that they concentrate mostly on the intelligent behaviour showed by agents. Precisely, in all these approaches, researchers tried to inculcate the capability in artificial agent so that it can react to the environmental stimuli and subsequently choose appropriate course of action. In [3], it is given that one of the main objectives of AI is to design a system that can be considered as a “machine with minds” in the full and literal sense [7]. Further, it is obvious that if an entity consists of the mind in true sense then it must inevitably pose the attributes of consciousness. Indeed, the domain of AI

reflects substantial interest towards consciousness. In [8], it is quoted that:

*“Some machines are already potentially more conscious than are people, and that further enhancement would be relatively easy to make. However, this does not imply that those machines would thereby, automatically become much more intelligent. This is because it is one thing to have access to data, but another thing to know how to make good use of it.”*

The term “intelligence” is closely related to “consciousness” and in the last ten years there has been a growing interest towards the field of artificial consciousness (AC). Several researchers from traditional AI addressed the hypothesis of designing and implementing models for AC. It is sometimes referred to as machine consciousness or synthetic consciousness. Very often, where we would say that a task requires intelligence, we would also say that it requires consciousness. Perhaps, in a broader sense the term “artificial intelligence” is fabricated myopically as it does not cater to the actual need of the field AI. In fact, this term obscures the fact that despite an early emphasis on problem solving, the field has always had more than just intelligence

in its sights. Indeed, the goal of AI is to enable the artificial agent to display the characteristics of mental properties or exhibit characteristic aspects of systems that have such properties. It is obvious that intelligence is not the only characteristic of mental property. In [9], it is envisaged that it is ambiguous to answer the question “Can machines think”? Further, it is also ambiguous to answer the question “Can an artificial system be conscious”? This is probably because of the fact that the definition of term “consciousness” is blurred and used in vivid senses. Indeed, mental property also encompasses many other characteristics, e.g., action, creativity, perception, emotion and consciousness. The term “consciousness” has persistently been a matter of great interest at the philosophical level of human being but it is not formidably addressed within the purview of AI. It is the candid view of the author that AC should be treated as a sub-domain of the AI because of its very nature. Further, human beings are conscious creators because they know what they do and perceive. Human beings can feel what happens to them and it is usually defined as being conscious or having consciousness [10]. Furthermore, there seems to be some strong dependence and symbiosis between autonomy and consciousness. As a matter of fact, consciousness is an essential attribute of mentality and intelligence. This is the reason why designing a model for consciousness is not only a technical challenge but also a theoretical feat. Undoubtedly, the actual implementation of models pertaining to consciousness substantially unveils the complexities associated with consciousness [8, 11–21].

Philosophically, consciousness implicates the resilience of the subjective experience. In contrast, AC is mainly concerned with an artificial agent capable of performing cognitive tasks. Moreover, the objective of AC is to render the explanation of different phenomena such as limitations of consciousness and attention. AC is also strongly associated with the capacity for developing representational states with intentional content.

An extensive and analytical literature survey reveals the fact that consciousness is one aspect of the mentality and AI deals only with those aspects of mentality that do not include consciousness directly. In contrast, AC creates a direct thrust on those aspects of mentality that are endowed with consciousness. In fact, within the realm of AC the subject matter of consciousness is not peripheral; rather, it is a central tenet. However, there is no firm demarcation line that can distinguish the conscious aspects of mentality that are tackled by AC and remaining aspects of mentality dealt by the AI. There are several and often conflicting hypotheses. Some researchers believe that consciousness is the consequence of special kind of information processing related with information integration [22, 23]. Further, a different group of researchers hypothesize consciousness as a goal generation and development [24], or embodiment process [25]. It is also considered as a kind of information processing akin to the global workspace [26, 27], the replication of imagination and synthetic phenomenology

[28, 29], or emotions [30] and so forth. An indicative example, however, can be based on the idea of Carruthers P as given in [31]. Indeed, there is a plethora of work done by many researchers that explicitly addresses the issue of machine consciousness [21, 25, 32–39].

There are some other associated questions of paramount importance like “what are the essential attributes artificial agents must possess to display the conscious nature?” and “what type of artificial system can be declared as conscious?”. Further, it is also important to enumerate all strong arguments that are against such claims. These issues are formidable because they are closely related to the theory of computationalism. Indeed, the problem of consciousness has been treated as a hard problem [40] because of its complex nature. However, it is murkier to analyse whether the problem of consciousness is a single overarching problem or an association of number of relevant “easier” and “harder” problems. In [40], David Chalmers has given critical comments to disambiguate the senses and philosophical issues involved. In this sequel, some scientists and philosophers have even argued that it may lie beyond our cognitive grasp [41, 42]. Further, it has been observed through extensive literature survey that the topic of consciousness generates a dichotomy between the researchers. There are two domains. The researchers of first domain believe that consciousness provides a better way to cope with the environment and is a topic of substantial interest within the purview of neuroscience and AI. In contrast, the researchers of other domain believe that there is no linking of consciousness with AI, neuroscience and cognitive science and perhaps it is a topic of philosophical interest. However, there are many scientific problems in which consciousness is of the paramount importance such as cognition, intentionality, representation, freedom, temporal integration and feeling versus functioning.

Consciousness can also be visualized from evolutionary perspective and is considered as something to do with allocation and organization of scarce cognitive resources. From the evolutionary perspective, it is important because it attempts to correlate the consciousness with nature. The evolutionary perspective of consciousness also eradicates blurredness in the way of our scientific understanding of nature of consciousness. Perhaps, the issue of consciousness is still controversial and full of obstacles and we do not yet know how to tackle it or how to measure the success obtained in this pursuit. In [43], it is mentioned that it is not possible to design an entirely new kind of conscious structure through theoretical reasoning. However, undoubtedly it is of the utmost importance to devise and build a true autonomous, conscious and efficient intelligence machine, i.e., a machine that possesses mind. Attempt has been made in this direction and AI substantially contributed in this endeavour. As a consequence, some important models have emerged such as the “blackboard model (BBM)” [44], or “agenda-based model (ABM)” [45]. These proposals are measurable by means of

how effectively they work or how well they replicate the human behaviour. However, these models do not seem to display any associated philosophical issues. Indeed, the lack of formal definition of consciousness does not preclude its research progress [46] because it is a historically proven fact that significant scientific progress has commonly been achieved in the absence of formal definitions. For instance, it is evident that well before the discovery of electron in 1892, the phenomenological laws of electrical current flow were propounded by Ohm, Ampere and Volta. Therefore, researchers have constructed their temporal definition of consciousness as and when required.

Rest of the paper is organized as follows. Section 2 briefly describes the interrelationship among body, mind, intelligence and consciousness. This section also describes the weak and strong AC. Section 3 renders various computational models pertaining to consciousness. Further, section 4 describes the machine and consciousness. This section also incorporates brief description of consciousness from the perspective of Indian philosophy. Moreover, in section 4 the question “Can a machine ever be conscious?” is critically analysed. Finally, concluding remarks are given in section 5.

## 2. Body, mind, intelligence and consciousness

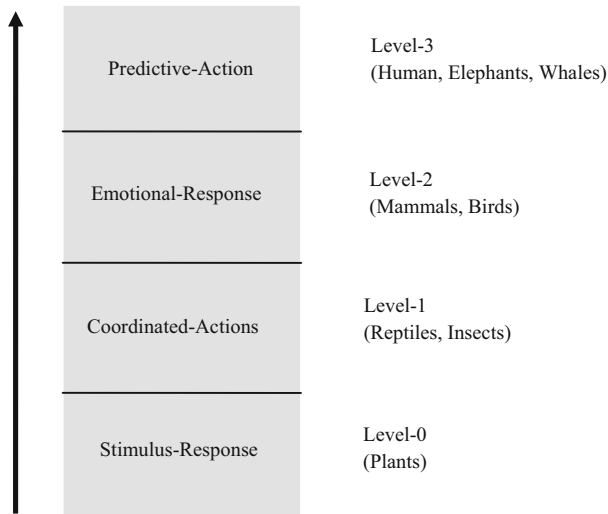
Body is the coarsest and grossest level of existence in living being and it is made up of the sense organs. Further, mind and intelligence may have some connection with the brain, as different types of living beings have different levels of intelligence. Indeed, human mind is the most fascinating thing that has mesmerized philosophers and scientists alike. However, consciousness does not seem to be merely a product of the nervous system or the brain. Furthermore, mind is subtler than the body and it is distinct from the consciousness. From the computationalism point of view, the human brain is essentially a computer. However, mind is presumably not a stored program as in a digital computer. In contrast, many other computationalist researchers are striving to create a dichotomy between modules that seems to be more computational such as vision system and those that seem to be less computational such as principles of brain organization. Perhaps, the brain organization is responsible for creativity, emotions, etc. It is obvious that nervous system or brain is not essential for the existence of consciousness, e.g., plants do not have any nervous system or brain, but they have consciousness. Further, whether a single living cell possesses mind or intelligence can be a matter of debate but it is undisputable that a single cell possesses consciousness. In this sequel, it is obvious to mention that it is not possible to describe every aspects of the brain in terms of computation. Perhaps some functioning of the brain may be similar to that of the computer whereas other functionings might show different patterns.

Perhaps, it might be quite interesting and formidable to note that if somehow we can decouple the functioning of brain for which the computational models are good from the rest of the brain’s functioning, which are good for other disciplines such as philosophy and theology, it would have certainly created significant insight in computationalism. Indeed, consciousness is the subtlest entity and is beyond the purview of body, mind and intelligence. Researchers have tried to link consciousness with the brain but everything was in vain [47].

### 2.1 Levels of subtlety of body, mind, intelligence and consciousness

There are different levels of subtlety of body, mind, intelligence and consciousness. Perhaps, this is the reason why we perform different activities to satisfy these entities. Some activities may satisfy the body while some other may satisfy the mind as the mind is just the next level above the body and is therefore in direct touch with it. Further, some activities can satisfy both body and mind partially, e.g., when we are hungry and eat some food. However, it does not satisfy the consciousness. Similarly, when we observe the scenic beauty of nature or listen to melodious music, our mind is satisfied and also our fatigued body may be relaxed and re-energized. Furthermore, there are some activities that can satisfy the intelligence, e.g., solving out a mathematical problem, composing a poem, etc. The level of intelligence is higher than that of the mind and thus satisfaction of intelligence produces the satisfaction of mind. However, hierarchically, consciousness is above the intelligence; that is why such activities may not satisfy the consciousness. Contrary to this, when we try to serve needy and deprived people our consciousness is satisfied and since the consciousness is the highest of all, obviously the intelligence and mind will also get satisfied in parallel. In [48], a “space–time” model has been proposed to quantify the consciousness. This model is based on evolutionary understanding of the brain’s component. This model does not incorporate intricate biological analysis. However, from the author’s point of view, the model proposed in [48] renders a useful framework for testing and making new forms of consciousness. In figure 1, different levels of consciousness are shown and perhaps it is pertinent to construct “smarter” artificial agents.

Philosophically, intelligence displays a dual attribute with mind. It is subtler than the mind and at the same time acts as a substrate of mind to engage it in consciousness, which in turn further strengthens the mind together with senses. Indeed, rationality and irrationality, right and wrong of actions are differentiated by means of intelligence. Intelligence is also pertinent to brain, which is the substrate of mind. The functioning of brain is similar to a feed forward hierarchical state machine and it is analogous to the hardware of computing machines. Moreover, the



**Figure 1.** Different levels of consciousness.

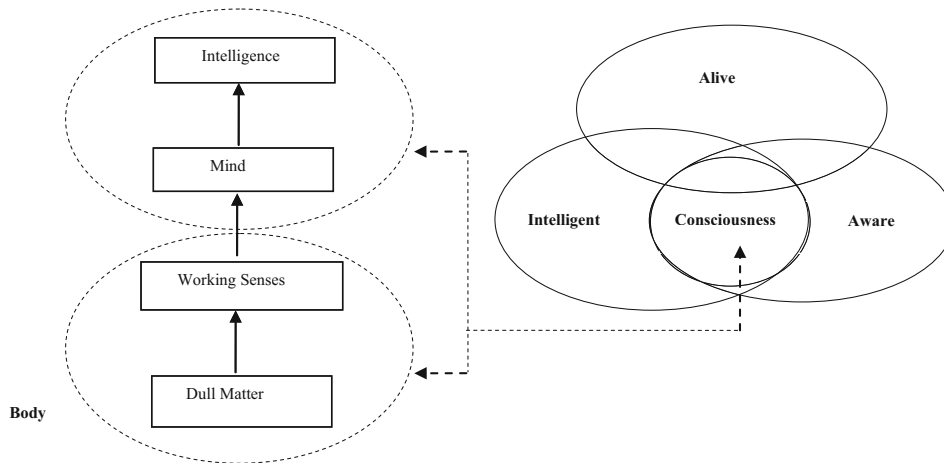
hierarchical regions of the brain possess the ability to learn as well as to predict its future input sequence. The hierarchy of intelligence, mind, working senses and dull matter is shown in figure 2.

It is indeed a matter of substantial research for a long time to inculcate the consciousness in artificially intelligent devices, and since last two decades creating conscious machine has gained initial impetus. Many researchers have contributed in this pursuit [49–63]. However, in spite of all these efforts the computational modelling of machine consciousness is still quite narrowed. This is due to the obscure nature of consciousness and related abstract notions like thought, attention, awareness, etc. Section 2.2 will provide more insight about AC.

### 2.2 AC

The main task of AI is to discover the optimum computational models to solve a given problem. The complexity of that given problem must be as high as the degree of problems solved by human being. It is remarkable that researchers from the AI domain need not be computationalists owing to the fact that computers can do things that brains do non-computationally. However, majority of AI researchers are computationalists to some extent irrespective of the fact that digital computers and brains-as-computers compute the things differently. Further, if we consider the case of phenomenal consciousness, there are very less number of researchers who believe that AI can solve it.

Consciousness is often thought to be the least amenable to being understood or replicated by AI. Perhaps, the nature of consciousness is thought to be intangible by the computations, algorithms, processing and functions of AI methods irrespective of the fact that AI is the most promising avenue towards AC. Indeed, AC is even more doomed than AI. It is because of the fact that the domain of AI is myopically understood. Moreover, the possible roles that AI might play in accounting for or reproducing the consciousness are not explicit. However, there are some foundational attributes relevant to AC, and these attributes show some common reasons given for AC sceptics. These attributes are prosthetic, discriminative and practically necessary. Generally, researchers consider three strands pertaining to AC. They are interactive empiricism, synthetic phenomenology and ontologically conservative hetero-phenomenology. At first glance it seems easy to distinguish the AI and AC. In general, AI endeavours to create an intelligent machine whereas AC attempts to create machines that are conscious. However, the subject matter of consciousness and intelligence is quite complicated and



**Figure 2.** Hierarchy of different components of intelligence.

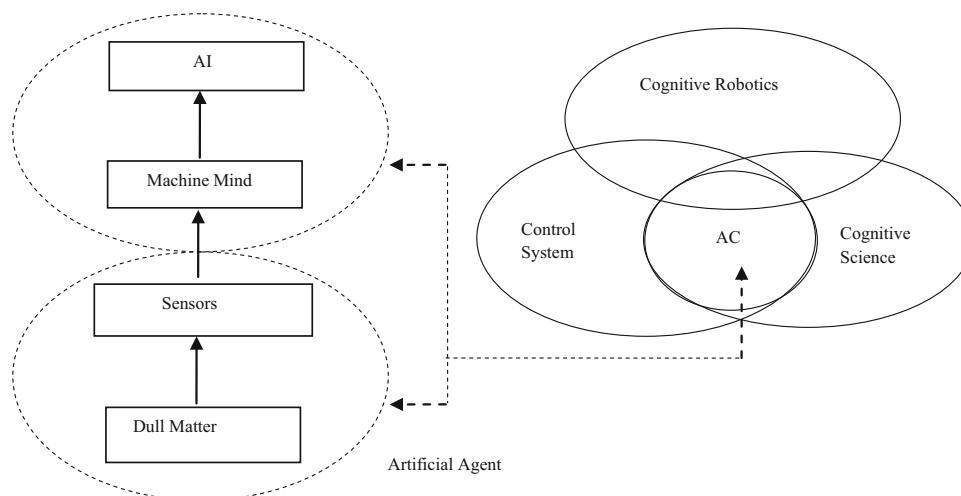
distinction between these two aspects requires philosophical foundation. Indeed, the researchers of machine consciousness domain envisage that classic functional view of the mind in terms of either functions or modules is not pragmatic to imbibe the full scope and capacity of consciousness in an artificial agent. Perhaps, this is the reason for the contradiction of traditional arguments against strong AI and it causes the myopic consideration of Searle’s Chinese Room or Block’s Chinese nation argument. Further, classic functional view of the mind renders the fact that a machine is not necessarily a Turing machine. In fact, most machines are supposed to be the replication of von Neumann’s blue print instead of a Turing machine. Moreover, different aspects such as embodiment, location, etc. flaunt the classic AI-disembodied view of a syntactical symbol crunching machine [64–67]. In general, machine consciousness exists somewhere in between the biological chauvinism and liberal functionalism. Biological chauvinism believes that only brain is conscious whereas liberal functionalism implicates that the behaviour of functional systems is equivalent to consciousness. However, the viability of only biological chauvinism is too gloomy because of the fact that existence of some kind of physical constraints is not entirely avoidable.

Indeed, the traditional field of AI emerged by the seminal field of AC and its aim is to reproduce the relevant features of consciousness using non-biological components. According to Ricardo Sanz, there are three motivations to pursue AC [68] as shown in figure 3:

1. implementing and designing machines resembling human beings (cognitive robotics),
2. understanding the nature of consciousness (cognitive science) and
3. implementing and designing more efficient control systems.

The current generation of systems related to man–machine interaction display substantial improvement in terms of mechanics and control movement. However, in spite of these improvements, the state-of-art anthropomorphic robots display very narrow capabilities of perception, reasoning and action in novel and unstructured environments. Further, keeping in view the importance of mutual interactions among the body, brain and environment it is worthy to say that the differences between the artificial and mental studies of consciousness have been endowed by the psychological and biological models of epigenetic robotics and synthetic approaches of the robotics [21, 52, 69, 70]. However, the term consciousness is more susceptible to the domain of philosophy. This might be due the fact that consciousness has no practical consequences or because it is a false problem. Perhaps, researchers are interested to focus on more well-defined aspects, e.g., vision, problem solving, knowledge representation, planning, learning and language processing. For the researcher of this category, consciousness is a by-product of biological/computational processes. In [71], it is stated that “Most roboticists are more than happy to leave these debates on consciousness to those with more philosophical leanings”. Contrary to this, many researchers give sound consideration on the possibility that human beings’ consciousness is more than the epiphenomenal by-product. These researchers have hypothesized that consciousness may be the expression of some fundamental architectural principle exploited by our brain. If this view is true, then it implicates that it is a must to tackle the consciousness in order to display the human level intelligence in artificial agents.

The researcher of pro-consciousness group believes that we have a first-person experience of being conscious and the consciousness is epiphenomenal. Further, this fact cannot be denied by theoretical reasoning. If we analyse the complex behaviour of animal kingdom, a correlation



**Figure 3.** Different motivations of artificial consciousness.

between highly adaptable cognitive system and consciousness can be observed. On the other hand, insects, worms, arthropods and similar living beings are usually considered devoid of consciousness. In fact, the recent work on AC consists of two important aspects. They are (1) the nature of phenomenal consciousness, which can be considered as a hard problem and (2) the role of consciousness in controlling and planning the behaviour of an agent. However, it is still unknown whether it is possible to solve the two aspects separately or not. Further, the interaction of a human with a robot in an unconstrained environment is the recent trend of research and it requires a better understanding of surrounding as well as the objects, agents and related events. Thus, it is obvious that the recent trend of interaction of humans and robots requires some form of AC. Many important contributions have been made by researchers pertaining to the possibilities of modelling and implementing the consciousness in machines, i.e., the conscious agent [21, 25, 32–39]. Indeed, machine consciousness is not merely a technological issue. It is also imperative to mention that it possesses many old unanswered questions pertaining to the nature of teleology, the nature of phenomenal experience, the relation between information and meaning, etc. Exhaustive literature survey revealed the fact that machine consciousness has a long past but a brief history [72]. The first use of word “machine consciousness” is observed in [73]. However, the problem has been addressed since Leibniz mill. In the present scenario, machine consciousness is considered as an excellent tool to tackle the hard problem of consciousness from a different perspective. In [74], the following is mentioned:

*“What I hope to persuade you is that the problem of the Minds of Machines will prove, at least for a while, to afford an exciting new way to approach quite traditional issues in the philosophy of mind. Whether, and under what conditions, a robot could be conscious is a question that cannot be discussed without at once impinging on the topics that have been treated under the headings Mind–Body Problem and Problem of Other Minds.”*

### 2.3 Weak and strong AC

AC is a promising field of research for at least two reasons. First, it assumes that consciousness is a real phenomenon affecting behaviour [14, 46, 75, 76]. Second, it suggests the possibility of reproducing the most intimate aspect of our mind, i.e., the conscious experience by means of machines. Further, behavioural role of consciousness is an important issue in AC and many researchers have given substantial impetus in this pursuit in an attempt to avoid the problem of phenomenal experience [11, 66, 68, 77]. Researchers classified the AC into two sub-domains. They are the “weak artificial consciousness” (WAC) and “strong artificial consciousness” (SAC). In [33, 78], it is suggested that it is possible to distinguish WAC from SAC with the help of

weak and strong AI. The WAC approach deals with agents that behave as if they were conscious, at least in some respects. However, this approach does not provide any commitment to the hard problem of consciousness. Contrary to this, SAC approach deals with the design and implementation issues of agents capable of real conscious feelings. The distinction between WAC and SAC sets a useful temporary working ground. However, it may suggest a misleading view. Indeed, setting aside the crucial feature of human mind may divert the understanding of cognitive structure of a conscious machine. Further, skipping the so-called “hard problem” could not be a viable option in the realm of making conscious machines.

The distinction between WAC and SAC is perhaps questionable due to the fact that it is not intertwined with the dichotomy between true conscious agent and “seems to be” conscious agents. Undoubtedly, human beings and most of the animals are phenomenally conscious owing to the fact that they exhibit the behavioural sign of consciousness. Further, because of the phenomenal consciousness, human beings experience pleasure, pain, sound, shapes, emotions, feelings of various sorts, etc. Moreover, human beings also feel experiences of thoughts and cognitive processes and thus it is obvious that human consciousness entails phenomenal consciousness at all levels. It is also intuitive to think that natural selection selects the consciousness without any selective advantages and thus it is obvious to be tempted to design a conscious machine without dealing squarely with the hard problem of phenomenal consciousness. If natural selection selects the consciousness without any selective advantages, there is no reason for the researchers of the machine intelligence domain to avoid this. This reveals the fact that the dichotomy between phenomenal and access consciousness as well as the differentiation between WAC and SAC is eventually factious. However, some researchers have adopted an open approach that does not rule out the possibility of actual phenomenal states in current or future artificial agents [21, 36]. Furthermore, some other researchers believe that a conscious machine is necessarily a phenomenally conscious machine [70, 79]. In [80], it is given that the possession of phenomenal experiences or P-consciousness is necessary for the conscious agent. Further, in [81] it is suggested that a human level of cognitive autonomy is also associated with consciousness.

Majority of the researchers from the machine intelligence domain corroborate the former definition; however, it is not always easy to have an explicit boundary between the two. For instance, it would be hard to deny that the machine is conscious if it exhibits all behaviours normally associated with a conscious being. Similarly, a machine might not be considered as conscious in spite of the fact that the machine exhibits all attributes of consciousness. Indeed, the domain of machine consciousness incorporates a heterogeneous network of researchers and is not unified with a set of clearly defined goals. It is also a challenging task to find out

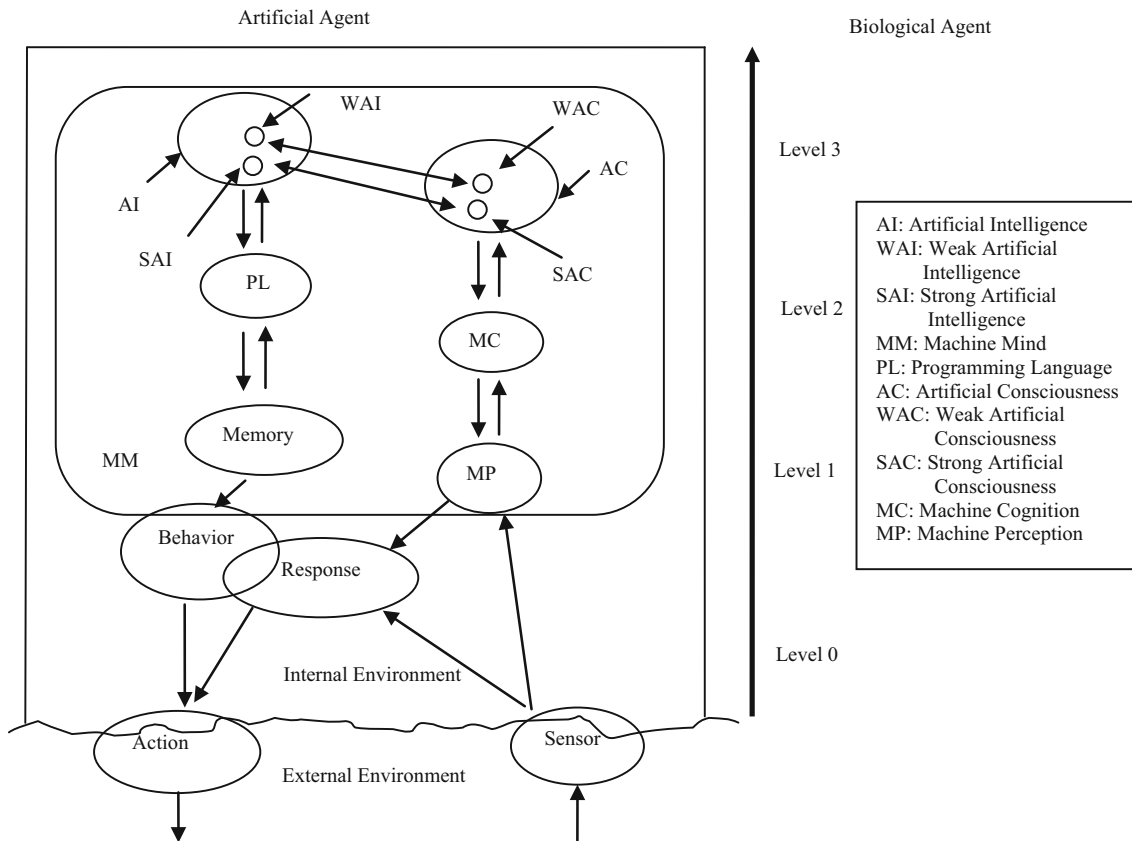


Figure 4. Interrelations among different features of an artificial agent.

how different goals of machine consciousness fit together. A modified version of inter-relation among different features of an artificial agent including WAC, SAC, machine perception, machine cognition, etc. [82] is shown in figure 4. However, some researchers of this domain also argued against the possibility of creating a conscious machine. Their argument is mainly based on a priori assumptions that “no machine will ever be like a man”. However, it is also convincing that any argument that seems to deny the possibility of machine consciousness would also recklessly defy the very possibility of human consciousness and thus such arguments seem to be faulty. For instance, a naïve adversary of machine consciousness may argue that CPUs and computer memory do not seem to be the right kind of stuff to harbour phenomenal experience and thus a computer could never be conscious. The author would like to quote the words of Lycan [83]:

*“[if] pejorative intuition were sound, an exactly similar intuition would impugn brain matter in just the same way [...]: ‘A neuron is just a simple little piece of insensate stuff that does nothing but let electrical current pass through it from one point in space to another; by merely stuffing an empty brainpan with neurons, you couldn’t produce qualia-immediate phenomenal feels!’ – But I*

*could and would produce feels, if I knew how to string the neurons together in the right way; the intuition expressed here, despite its evoking a perfectly appropriate sense of the eeriness of the mental, is just wrong.”*

### 3. Computational models of consciousness

Some researchers believe that it is not possible to replicate the consciousness by computations, algorithms, processing and functions of AI method in artificial agents. However, in spite of all drawbacks, some important computational models of consciousness emerged forth. These models are:

- a) Moore/Turing inevitability model,
- b) Hofstadter–Minsky–McCarthy model,
- c) Daniel Dennett model,
- d) Perlis–Sloman model and
- e) Brian Cantwell Smith model.

#### 3.1 Moore/Turing inevitability model

This class of proposal relies on Moore’s law [84], which states that exponential progress in the power of computers

synergized by how well future programs will do on the “Turing Test” [9] can rate how intelligent a computer is. Further, in [47, 85], it is assumed that substantial development of faster and more capable computers will inculcate intelligence in machines equivalent to human intelligence and it might subsequently surpass humans in intelligence, which in turn produces the machine consciousness of inevitable consequence.

### 3.2 Hofstadter–Minsky–McCarthy model

Many pertinent problems of consciousness have been visualized by Richard Hofstadter [86]. The Richard Hofstadter’s work mainly seems to be intended to stimulate or tantalize one’s thinking about the problem e.g., in [87] there is a chapter in which characters talk to an anthill. The anthill is considered analogous to brain and ants are considered as neurons. Hofstadter hypothesized that the anthill can carry on a conversation because the ants that compose the anthill play roughly the role neurons play in a brain. Further, in this book there is a brief evocative conversation between an anthill and anteater during which the anthill offers anteater some of its ants. This provides a playful digression on the subject, i.e., it makes the vivid possibility that “neurons” could implement a negotiation that may end in their own demise. This story reveals that Hofstadter believes anthill to be conscious and if anthill can be considered conscious why one could not use integrated circuits rather than ants to achieve the same end. The conceptual framework of Hofstadter offers a new way of thinking about consciousness by showing that introspection is mediated by models. However, in this conceptualization, most of the intricate details are left out. In fact, Hofstadter invented a novel, innovative and playful version of argumentation and it seems that this new version of argumentation is striving to break up the concepts into different granules and subsequently toss them together into so many configurations. However, in this entire transformation the original questions one might have asked get shunted aside.

Minsky wrote a paper in 1968 [88], which introduced the concept of self-model, which came forth as the central tenet to the computational theory of consciousness. In this approach consciousness is not considered very seriously as a problem. As per this model, answering of various pertinent questions about consciousness requires more concentration on the models used rather than on the questions themselves. Further, many interesting ideas have been proposed in Minsky’s famous book “The society of mind” [89]. However, many ideas given in [89] related to consciousness lack serious and plausible explanations; for instance how Freudian phenomena might arise out of the “society” of sub-personal modules that he considers within the domain of human mind needs a worthy and plausible proposal to argue for or against.

Further, John McCarthy also contributed in the domain of consciousness and he usually termed it as “self-awareness” [90]. McCarthy’s work is mainly concentrated on the problem solving capacities of the robots. Moreover, he has also focused on how robots can augment their problem solving capacity by introspection. Perhaps, self-awareness is an ability of paramount importance in robots and it can inculcate inside robots the ability to infer that they do not know something. In McCarthy’s work the word “awareness” is used in a sense that is quite different from the notion of phenomenal consciousness. In [91], McCarthy addressed the issue of “zombies”. It is a philosophical term to symbolize the hypothetical beings who behave exactly as we do but do not experience anything. The issue of “zombies” is also discussed by Todd Moody [92]. In his work, Moody listed some introspective capacities that can be imparted to a robot. However, McCarthy has given critical comments on Moody’s work:

*“Moody is not consistent in his description of zombies. On page 1 they behave like human. On page 3 they express puzzlement about human consciousness. Would not a real Moody zombie as though it understood as much about consciousness as Moody does?”*

From the author’s perspective, the Moody’s idea of a zombie also does not seem to be convincing. We can consider the example of dreaming phenomenon. Human beings can experience it but it is senseless for zombies. In fact, given any functional description of cognition as detailed and complete as one can imagine, it will still lack many information for its description. Indeed, Moody visualizes that an intricate combination of hardware and software (i.e., a robot) can emanate consciousness just like the computer-controlled heating system radiates heat energy. We can illustrate the ability of zombie by comparing it to the case of printing any sentence on the paper. What sentence we are printing on the paper does not depend on the colour of the paper. Similarly, the zombie can perform the same things as human beings can do but without experiencing it. However, Chalmers [40] infers that zombies make sense.

### 3.3 Daniel Dennett model

Daniel Dennett has contributed substantially in the domain of philosophy of mind as well as in cognitive science. Dennett’s PhD dissertation proposed a model for a conscious system [93]. This model of consciousness introduced a novel block diagram that became a standard attribute for the many theories of psychologists. Further, in his model, more impact is given on introspection rather than on the problem solving ability. Dennett’s proposed model of mind can be considered as a precursor for the computational model of mind and it constitutes one of the most ambitious and detailed proposals for how AI might account for



consciousness. Moreover, Dennett used the phrase “Cartesian theatre” to describe a hypothetical place in brain where the self becomes aware of the things. He observed that the belief of Cartesian theatre is deep rooted and persistent in philosophical and psychological writings as well as in the domain of common sense.

It is a well-known fact that very often things remain in pre-conscious or sub-conscious state inside the brain. Therefore, it is intuitive to think what happens when a series of events becomes conscious? Dennett conceptualized a ridiculous idea that to make an event conscious requires feeling it on the screen of the Cartesian theatre. Most of the traditional models of consciousness have not been accepted by Dennett. He wanted to construct a new model of consciousness and he believes that language is a key component for the human consciousness. It is difficult to think about human consciousness without having the ability of a normal human adult to express what he is thinking. Indeed, Dennett’s consciousness model assumes that consciousness is based on language and not vice versa. Further, this model is based on the fact that consciousness requires substantial ingredients of matter, perception, intellectual skills, etc. Furthermore, there must be some linguistic abilities that depend on consciousness and at the same time this basic ability must exist “before” and “independent” of consciousness.

### 3.4 Perlis–Sloman model

The development of computational model of consciousness has substantially been enriched by the researchers who were actively engaged in core problems of AI as well as pertinent philosophical issues. In [94], it is proposed that consciousness is ultimately based on self-consciousness. According to Perlis [94]:

*“Consciousness is the function or process that allows a system to distinguish it from the rest of the world...To feel pain or have a vivid experience requires a self”.*

To some extent Perlis argument seems to be non-convincing because his argument often depends on thought experiments. Therefore, there can be a possibility that one is conscious but not of anything, or of as little as possible. Indeed, the introspective thought experiments are just not a very accurate tool and it appears that Perlis’s model is not compatible with computationalist approach. In [94], he described the “amazing structure...” but it is not clear how “amazing” structures will be and thus render it non-computational. He quoted:

*“I conjecture that we may find in the brain special amazing structures that facilitate true self-referential process, and constitute a primitive, bare or unawareness, an ‘I’. I will call this the “amazing-structures-and-process paradigm”.*

Aaron Sloman has written vividly about philosophy and computation. In fact, his inclination was more towards the conscious control by incorporating the philosophical aspects and strategy of organizing complex software in comparison with phenomenal consciousness. His book “The computer revolution in philosophy” [95] has almost nothing to say about the subject. Further, he tried to assert that the concept of consciousness includes many different processes and these processes should be sorted out before hard questions can be answered.

### 3.5 Brian Cantwell Smith model

Brian Cantwell Smith model of consciousness is computational and is also an anti-reductionist. He considers consciousness as a crucial topic and his early work [96] was on “reflection” in programming languages. In [96], he proposed how and why a program written in a language could have access to information about its own subroutines and data structures. It is obvious that this work can play a vital role in maintaining a self-model and thus the consciousness in a system. Smith tried to infer that computers are always connected to the world and therefore their semantic rules must be determined by keeping in view the characteristics of these connections rather than by what a formal theory might say. However, one might put forth the example of transducers that connect them to the world and the transducers are non-computational [97]. Indeed, there is no principled way to draw a boundary between the two parts because ultimately a computer is physical parts associated with other physical parts. In [98], it is stated:

*“...The view of computers as somehow essentially a form of Turing machine...is simply mistaken. [The] mathematical notion of computation...is not the primary motivation for the construction or use of computers, nor is it particularly helpful in understanding how computers work or how to use them”.*

There is one similarity in the arguments of Smith [96] and Sloman [98]. Both argued that it is misleading to say that the Turing machine is ideal means for computationalism. In [99], it is given:

*“...computers (as distinct from robots) produce at best only linguistic and exclusively “cognitive”-programmable-behaviour’: the cognitive ‘inner’ rather than on action, emotion, motivation, and sensory experience”.*

## 4. Machine and consciousness

There are so many intuitive questions within the purview of machine intelligence, e.g., could a mechanical device ever replicate human consciousness? Could a machine really be intelligent like a human being? Is it true that dismantling of

a machine resembles surgical operation of a sentient being? If we ask the first question to an AI researcher the reply will be either (1) life-like robots are just around the corner or (2) it will never happen. Undoubtedly, computers are consistently becoming more efficient and powerful. Today's computers can replicate human expertise in dozens of areas. It is true that AI has extended its viability from library to everyday life. However, it is also true that the modern computers cannot replicate different abilities of even a four-year-old child. Indeed, the information processing capacity of even the most powerful supercomputer is equal to that of the nervous system of a snail. Further, in some peculiar kinds of information processing, computers display substantial superiority over the human being, e.g., it is easy for the machine to remember a long digit number but it is difficult for the machine to extract the gist of a statement. In contrast, for human being it is hard to remember the number but easy to extract the gist of a statement. Perhaps, it is due to the fact that machines possess a limited number of processors working at amazingly faster speed whereas human brain's processors are slower but they are in huge quantity and thus let the human brain to recognize the intricate and complicated patterns instantly. On the other hand, computers have to proceed in a logical manner, i.e., they process the things algorithmically one step at a time. Human brains have imbibed terabytes of information over the years as the humans interact with other humans and with the environment unlike machines. Moreover, brains have achieved the benefit of billion years of evolutionary gain. Therefore, it is obvious to think how a machine can replicate the human mind. Indeed, technological progress is unpredictable. However, there is no doubt that the future computers will probably perform considerably better than today's computer in certain pursuits but future machines will probably not do as well as a human being for other kinds. It is also intuitive to think that we can declare a machine to be conscious when it behaviourally renders the human resemblance. In [100], it is given that

*"...Special consideration is given to a study of consciousness as opposed to unconsciousness, with many helpful hints on how to remain conscious".*

Perhaps, the discourse of consciousness cannot be completed if we do not analyse the consciousness from the Indian philosophical perspective. Indeed, Indian philosophy on consciousness is vast and pervasive and many important discussions about consciousness have been made in Indian philosophical literature starting from the Vedas and Upanishads up to the contemporary Indian thinkers. It would be irrelevant to present an exhaustive discussion of Indian philosophy on the consciousness. However, keeping in view the subject matter of this paper, author will try to incorporate few points pertaining to the mind and consciousness from various Indian philosophical systems in sub-section 4.1 and further, in sub-section 4.2, the author will search the possibilities of consciousness in machines.

#### 4.1 Consciousness in Indian philosophy

A rigorous review of Indian philosophical literature revealed the fact that it entails a deep rooted philosophy of consciousness. Upanishads can be considered as one of the earliest works of paramount importance in this pursuit. In Upanishads [101–104], different levels of consciousness have been unveiled. The German philosopher, Arthur Schopenhauer (1788–1860), was greatly influenced by them. Alcorn [105] believes that Paul Carus [106] derived his theory of the unconscious from the Upanishads. The Upanishads, particularly Māndūkya Upanishad [107], set forth quite clearly the four states of self. They are:

- The waking state: equivalent to the "Conscious".
- The dreaming state: the "subconscious". In the subconscious state the self loses contact with reality, and the soul fashions its own world in the imagery of its dreams.
- The deep-sleep state: a deeper level of the subconscious approaching complete unconsciousness. It is the state of bliss in which there is no contact with reality, no desire, no dreams.
- The super-conscious state: according to Vedanta, it is the state in which seers get flashes of great truths in the form of vague apprehensions. These vague apprehensions are elaborated afterwards in the waking consciousness.

Indian philosophy also establishes the hierarchy of dull matters, sensors, mind, intelligence and consciousness. In [108], it is mentioned that the working senses are superior to dull matter; mind is higher than the senses; intelligence is higher than the mind and the soul is higher than the intelligence.]

Further, in pre-upanishadic period it was hypothesized that mind did generally implicate consciousness and is used to form ideas and decisions [109]. Indeed, the subject matter of consciousness is central to Vedanta. The fact that human intelligence cannot be created by any mechanistic process is presented in Vedanta. Indeed, it emanates directly from the conscious living force within the body—the soul. According to the paradigm of Vedanta, even the brain is treated as non-intelligent. Further, Vedanta states that brain uses consciousness as its computing instrument just like we use sometimes paper and calculators for computation. In [110], the interrelation and intertwining among the consciousness, intelligence, mind and senses is visualized as follows:

*"The individual is the passenger in the chariot of the material body, and the intelligence is the driver. Furthermore, mind is the driving instrument and the sensory organs can be considered as the horses of chariot. Moreover, the self is enjoyer or sufferer in the association of mind and senses."*

Further, the "Sāmkhya philosophy" interprets consciousness in a specific direction. From the point of view

of “Sāmkhya philosophy”, consciousness is treated as the essence of the self and it cannot be produced by the element. It is because of the fact that consciousness is not present in the element separately, and hence cannot be present in them all together [111]. Moreover, within the purview of Sāmkhya philosophy, consciousness is considered as the essence of the self but not the essence of mind. Furthermore, consciousness is self’s very essence and not a mere quality of it. According to Sāmkhya philosophy, the self is not the brain, nor mind nor ego nor intellect and even more it is not the aggregate of conscious states. Perhaps, it is a conscious spirit. In addition, the self is not a substance whose attribute is consciousness, but it is the pure consciousness as such. The Sāmkhya regards the self to be the subject of consciousness. The self is a steady constant consciousness in which there is neither change nor activity. In fact, all change and activity, all pleasure and pains belong to matter and its products like the body, mind and intellect. Further, in [112], it has been proposed that consciousness is an independent entity and it emerges out owing to the interaction between the self and matter. However, from the perspective of Sāmkhya philosophy, mind is a product of matter and it is not the consciousness. Moreover, the author has observed the existence of various states of consciousness in Sāmkhya philosophy as given in [107]. Perhaps, the Sāmkhya philosophers seem to be aware of the conscious, subconscious, preconscious and unconscious states of modern psychology. Moreover, the Indian philosopher “Cārvāka” did believe in the materialistic view of consciousness but it was severely criticized by Indian philosophers.

Further, in [113], it is given that consciousness is “awareness” found in all living beings and it has no independent existence of its own. It exists essentially in something else. Non-living atoms have no such property, e.g., a jar has no consciousness. Indeed, consciousness is not a quality of mind or body instead it is the property of the soul. Furthermore, it is asserted in [109] that mind is not the thinker rather the soul is thinker of ideas and therefore, the soul is the substratum of consciousness. The fact that soul is the substratum of consciousness is also given in “Mīmāṃsā”. In [114], consciousness is regarded as the permanent entity of the self and it is admitted that consciousness is self-luminous and it can be described as the knower and sometimes as cognition. Indeed, consciousness is a dynamic mode of the self and its result is seen only in the object, which becomes illuminated by it. Further, in [115], it is given that according to Ramanuja:

*“Consciousness is not a substance and is independent of sense object contact. Consciousness is continuous and eternal. It is present in present-perception, reasoning and memory. It is inseparable from the subject in all his cognitive activities.”*

#### 4.2 Could machine ever be conscious?

Indeed, there are dual aspects of word “consciousness”. Very often a conscious process turns to unconscious and vice versa is also a true phenomenon, e.g., when we learn how to drive the car, every action has to be thought out, but with practice the skill becomes automatic, i.e., conscious process becomes unconscious. Similarly, unconscious processes can also become conscious, e.g., by using the bio-feedback mechanism and intense concentration we can focus on a hidden sensation like our heartbeat. Further, cognitive activity of a living being is mostly unconscious while sleeping. However, we may wake up from this state spontaneously and this awakening can be because of many reasons, e.g., due to some internal “clock” or other internal event, or due to some external events like ringing of alarm clock, smell, pain other significant signals, etc. These signals can break the slumber while intensity of such signals may be below the threshold level and thus insignificant for sensory input. Therefore, it is explicit that there is a certain level of “awareness” even in the so-called unconscious state. Furthermore, even in wake state, most of the cognitive activities may be unconscious. In fact, both cognition and consciousness seem to be linked with complexity of response and behaviour performed by the agent. Indeed, the cognitive and conscious capacity seems to increase with an increasing level of complexity, or interconnectedness [116]. It is ridiculous to believe that a machine can ever become conscious from unconsciousness and vice versa in this sense by accessing a subset of information from the whole system. Perhaps, we cannot infer that the operating system of a computer works in an oblivious manner. Indeed, it is designed in such a manner that certain kinds of information are available to the programmer or user. It is pertinent to mention that information processing must be accomplished in real time whether it is an artificial agent or biological agent, i.e., computer or brain. Further, it is obvious that if in any device (artificial or biological) every bit of information is easily available at all the times to every process then the device would be perpetually lost in thought. Indeed, at a given point in time only some kinds of information are relevant for the system and only such information should be routed in to the system’s main processors. This fact is true even for robots of the future generation. Certainly, the machines of coming generation will need some kind of a control system that limits what goes into and out of the individual processors; otherwise the entire machine will display an abrupt and fluctuating behaviour because of the processors fight for control. In this perspective it is genuine to say that the present and future computers must be constructed on the basis of dichotomy between “conscious” and “unconscious” processing. However, it is intuitive to think that we can ever know that the computer is conscious that someone is inside the room by observing through camera-eyes and capturing the signals from sonar sensors. Indeed, it is important to ponder that ‘can machine be

conscious?'. But it is more important to analyse that 'could sensor data create consciousness in true sense?' There will always be some sort of temptation to say that it is simply a stimulus–response machine programmed to act as if it is sentient but it is not possible to come up with an experimental test that can prove it strenuously.

Philosophically, it is obvious to think that if one cannot determine that computer is sentient how one can infer that human being the creature made out of meat rather than metal is sentient. Further, is it logical to state with certainty that consciousness emanates from the brain tissue rather than from mind? Perhaps, it seems ridiculous that researchers are toiling hard to conclude whether a computer can be conscious or not while our own consciousness is beyond our conceptual grasp. In fact, the human mind, which is merely an outcome of evolution process, is biologically incapable of understanding the intricacies of consciousness with certainty. Further, the idea that brain is some kind of a computer is not greeted with much enthusiasm in research communities except some researchers of the expert systems [117, 118]. It is like an attempt to measure the dimension of any object precisely and accurately with a non-calibrated device. Therefore, it is hard to believe that consciousness is similar to a computer program. From the author's perspective the dilemma about whether a computer would be sentient cannot be solved anytime soon; hence, we do not expect that someone will tell us with certainty pertaining to this issue. In fact, the computational framework of brain renders it mostly for memories. The recent trend of information-based computational theory can be considered as the most comprehensive means for understanding the functioning of brain. However, the idea of computation being central to the brain seems to be a metaphor. Further, the author would like to quote Maudlin's argument that claims that a Turing machine can be conscious. The "necessity condition" of the argument of Maudlin [119] claims that

*"Computationalism asserts that any conscious entity must be describable as a nontrivial Turing machine running a nontrivial program."*

However, Maudlin has not given any defence for his argument. Presumably the idea is that TMs are universal, and that universal machines can compute whatever function the user wants. Therefore, Turing machine can be considered as a good candidate for conscious experience. Indeed, Maudlin's argument is based on a substantial assumption and thus Sprevak [120] refuted it:

*"...it is not true that a universal computer can run any program. The programs that a computer (universal or otherwise) can run depend on that machine's architecture. Certain architectures can run some programs and not others. Programs are at the algorithmic level, and that level is tied to the implementation on particular machines."*

The author also candidly agrees with this statement. Further, architecture is a set of primitive operations and basic resources available for building computations [121] and computationalism could be formulated in terms of input–output functions. The primitive characteristics of TM are (i) changing (or preserving) a single square of the tape, (ii) moving the head along the tape one square left or right and (iii) changing the state to one of a finite number of other states. The specification of computational architecture requires a high level of abstraction, e.g., it is not desirable to note what a TM tape is made of, or how long it takes the head to move to a new position. Indeed, computer program is a set of instructions required for building a machine that, when combined with appropriate context, generates a computational process. Further, computational process can be treated as a series of primitive operations. The relationships between contexts and outputs define the mathematical function that a machine computes. It is an experimental fact that the neural circuitry of human brain is considerably slower than silicon. It implicates that basic behavioural responses must be pre-computed and then looked up by fast indexing technique. One indication for the acceptance of the computational theory of brain is given in [122]. In [122], it is proposed that complex phenomena have simple explanations once the framework of computation is accepted. Furthermore, from the machine point of view, it is obvious that different machines can compute the same mathematical function in different ways and this can be the consequence of TM universality. It implicates that consciousness cannot be generated from mathematical functions because the same mathematical function may create different sorts of AC in different machine architectures. This is due to the fact that different architectures will have different primitives for generating the computational processes. This is the reason why Sprevak [121] stated that we need to make clear whether computationalism is a crux for mathematical functions and in this case we do not have to care about the architecture of our machine or whether it is better expressed in terms of processes, architectures and programs. In the latter case we have to care about the architecture of our machine.

It is of paramount importance to consider the formation of memory circuits. From computational perspective, the peripheral visual circuitry constructs code for natural images and subsequently this code represents the image. Moreover, the computational model introduces a peculiar dimension of neural substrate of the code. This neural substrate of the code should use an economical signalling strategy. In fact, neurons display the signals in discrete spikes and half of the metabolic cost is in spike generation. Therefore, it is obvious that if the code uses less spikes it is preferable. It is astonishing that these codes when created in computational simulations appear very similar to the experimental observations from neural recordings [123, 124]. This method implicates that there must be a trade-off between simplicity and accuracy of the code

Level	Computational Model
Operating System	Schedule all behaviors and establish trade-off among debugging and runtime behaviors.
Debugging	Analyze unusual events using off-line mode.
Run Time	Pick a suite of behaviors to handle the current situation.
Behavior	Sensory-motor coordinated actions.
Routines	Task-specific tests of the environment.
Calibration	Encode environmental statistics in specific circuits.
Neural	Models of specific neurons in circuits including excitation and inhibition
Synapses	Models of neuron's components and roles of different neural transmitter e.g. neural spike signaling, basic unit of memory

**Figure 5.** A proposal for the brain's use of computational hierarchies.

against the complexity of its description. Furthermore, the visual codes incorporate the description of visual data but these codes do not include a prescription for what to do with it. Therefore, it is intuitive to consider them as non-conscious. In this sequel, the author would like to give the examples of primates. In primates the eyes have pronounced fovea (the point on the retina where visual activity is the highest) so that the resolution at the point of gaze is greatly increased over the periphery. The consequence of this is that the primate eyes use rapid ballistic movements almost all the time to orient the high-resolution gaze point over interesting visual targets. It is interesting to think that visual coding can generate such a kind of flip flop in AC.

Indeed, consciousness cannot depend on what function a machine computes. Several good arguments have been given in [119]. Let  $f$  be a computable function that can be performed by TM after taking many steps. It is explicit that a given computable function can be primitive, i.e., the function can be performed in a single operation for some architecture. However, some other architecture may take several steps to perform the same function, e.g., a primitive function may take several steps on a TM. Further, let the computation of mathematical function  $f$  create a short session of phenomenal experience. Now, we can imagine a computational architecture for which  $f$  can be treated as a computational primitive. It is obvious that executing a single primitive in a single computational step is not enough to create conscious experience. Indeed, it seems hard to see how different elements of conscious episode might change. This is due to the fact that there is no

supervenience in the computational framework that could be changed for conscious experience. Moreover, when we study memory formation, the use of memory is postponed. Similarly, when we study the use of memory, the existence of memory is assumed. In fact, they are the two abstractions from many that must be assumed to handle the richness of human brain computation. Figure 5 shows a more complete view for brain's use of computational hierarchy. In figure 5, different sub-tasks are described at different levels of abstraction. Perhaps, the modularity in brain's use of computational hierarchy will preclude the generation of AC in machine because the consciousness of human being is a continuous function instead of being discrete. Further, in the case of phenomenal consciousness the situation is entirely different and is very difficult to explain it by the theory of computationalism. This is because of the fact that computers do not have experiences. It is possible to construct a digital climate control system but that will never let the system to feel heat or cold. Actually every step in physical mechanism proceeds in a programmed manner. There is no role of consciousness and no one is tempted to invoke conscious sensation in physical systems. Perhaps, the ability to produce consciousness or to experience the things is the key factor that exempts the brain from the hypothesis of computationalism. Over and above, the meaning of consciousness is multifaceted; however, it is a "hard problem" [40]. It is really hard to explain how a physical system can have vivid experiences with seemingly intrinsic qualities such as the spectrum of rainbow or redness of a tomato. It is easy to talk about the feeling and

sensations but all these things are arduously indefinable. What we can utmost do is that we can attach a label to certain tastes, but we really do not have any idea whether other people are feeling the same sensation of a given taste or different. Furthermore, the theory of computation possesses many associated limitations and perhaps all these limitations are concerned with mathematical infinities, which in turn produce limitations for the AC.

Indeed, different interrelated aspects of AC are essential for exploring the internal and external environment of the artificial agent. In [125], it is mentioned that consciousness could be viewed as a constantly changing process, interacting with a constantly changing environment in which it is embedded. Further, some researchers visualized consciousness as a “force field”, similar to an electromagnetic field [126, 127]. These views reflect the thought that it is not possible to display consciousness through neural or electrochemical processes. Indeed, it would have little meaning to us, even if we claimed we had succeeded. It is a scientific fact that concepts at one level are very often not transferable in a meaningful way to another level. It is not possible to refute that the qualities and properties that were irrelevant at lower levels emerge in a bizarre manner at each new level in the hierarchical organization of matter. This view is also reflected in [128] while mind is compared with quantum reality:

*“The mind is not a part of the man–machine but an aspect of its entirety extending through space and time, just as, from the point of view of quantum mechanics, the motion of the electron is an aspect of its entirety that cannot be unambiguously dissected into the complementary properties of position and momentum.”*

It seems logical that mind may need some quantum mechanical description. In [129, 130], it is argued that there are some non-computable aspects of mental phenomena or mind, e.g., understanding and insight. It is not possible to simulate such aspects on a machine. Such phenomena may even require a new physics, new laws and principles that are not mechanistically derivable from lower levels. With this perspective, consciousness seems to fundamentally transcend physics, chemistry and any mechanistic principle of biology [82].

## 5. Conclusion

Body, mind, intelligence and consciousness are mutually interrelated entities. However, consciousness is subtler than intelligence, mind, senses and body. AC is mainly concerned with the consciousness possessed by an artificial agent and thus makes it capable to perform the cognitive tasks. AC attempts to explain different phenomena pertaining to it, including limitations of consciousness. There are two sub-domains of AC. They are the “weak AC” and

“strong AC”. It is difficult to categorize these two sub-domains due to the fact that they are not related with the dichotomy of true conscious agent and “seems to be” conscious agents. Further, researchers have given few computational models of consciousness. However, it is not possible to replicate the consciousness by computations, algorithms, processing and functions of AI method. In fact, however vehemently we say that the computer is conscious, it is ridiculous to imbibe that sensor data can create consciousness in a true sense. Indeed, consciousness is not a substance and is independent of sense object contact and it cannot be produced by the element. Further, consciousness is “awareness” found in all living beings; it is the essence of the self and non-living atoms have no such property. Furthermore, consciousness cannot depend on what function a machine computes. Over and above, exploration, cognition, perception and action are fundamental to all levels of existence and are related to consciousness. In this sense, consciousness is inseparable from life, expanding and evolving together by means of inter-relation and inter-action. Hence, consciousness can be regarded as a driving force in nature. However, it does not seem that artificial consciousness can augment the intensity of artificial intelligence successively in a periodic manner and thus render an ultra-intelligent machine. The conclusion of this paper is accomplished by the statement of I. J. Good:

*“Let an ultra-intelligent machine be defined as a machine that can far surpass all the intellectual activities of any man however clever. Since the design of machines is one of these intellectual activities, an ultra-intelligent machine could design even better machines; there would then unquestionably be an ‘intelligence explosion,’ and the intelligence of man would be left far behind. Thus the first ultra-intelligent machine is the last invention that man need ever make [131].”*

## References

- [1] Shannon C E and Weaver W 1949 *The mathematical theory of communication*. Urbana: University of Illinois Press
- [2] Haugeland J 1985 Artificial intelligence: the very idea. In: *Mind design II*. Cambridge, MA: MIT Press
- [3] Russell S and Norvig P 2003 *Artificial intelligence: a modern approach*. New York: Prentice-Hall
- [4] Pfeifer R 1999 *Understanding intelligence*. Cambridge, MA: MIT Press
- [5] Sporns O 2011 *Networks of the brain*. Cambridge, MA: MIT Press
- [6] Pfeifer R, Lungarella M *et al* 2007 Self-organization, embodiment, and biologically inspired robotics. *Science* 5853: 1088–1093
- [7] Haugeland J 1985 *Semantic engines: an introduction to mind design*. Cambridge, MA: MIT Press

- [8] Minsky M 1991 Conscious machines. In: *Machinery of consciousness*, National Research Council of Canada
- [9] Turing A 1950 Computing machinery and intelligence. *Mind* 49: 433–460
- [10] Nagel T 1974 What is it like to be a bat? *The Philosophical Review* 4: 435–450
- [11] Baars BJ 1988 *A cognitive theory of consciousness*. Cambridge: Cambridge University Press
- [12] McCarthy J 1995 Making robots conscious of their mental state. In: Muggleton S (Ed.) *Machine intelligence*. Oxford: Oxford University Press
- [13] Edelman G M and Tonoi G 2000 *A universe of consciousness: how matter becomes imagination*. London: Allen Lane
- [14] Jennings C 2000 In search of consciousness. *Nature Neuroscience* 3(8): 1
- [15] Aleksander I 2001 The self ‘out there’. *Nature* 413: 23
- [16] Baars B J 2002 The conscious access hypothesis: origins and recent evidence. *Trends in Cognitive Science* 6(1): 47–52
- [17] Franklin S 2003 IDA: a conscious artifact? In: Holland O (Ed.) *Machine consciousness*. Exeter, UK: Imprint Academic
- [18] Kuipers B 2005 Consciousness: drinking from the fire hose of experience. In: *Proceedings of the National Conference on Artificial Intelligence*, AAAI-05
- [19] Adamin C 2006 What do robots dreams of? *Science* 314(5802): 1093–1094
- [20] Minsky M 2006 *The emotion machine: commonsense thinking, artificial intelligence, and the future of the human mind*. New York: Simon & Schuster
- [21] Chella A and Manzotti R 2007 *Artificial consciousness*. Exeter, UK: Imprint Academic
- [22] Tononi G 2004 An information integration theory of consciousness. *BMC Neuroscience* 5: 1–22.
- [23] Tononi G 2008 Consciousness as integrated information: a provisional manifesto. *Biological Bulletin* 215: 216–242
- [24] Manzotti R and Tagliasco V 2005 From behaviour-based robots to motivation-based robots. *Robotics and Autonomous Systems* 51: 175–190
- [25] Holland O 2004 The future of embodied artificial intelligence: machine consciousness? In: Iida F (Ed.) *Embodied artificial intelligence*. Berlin: Springer, pp. 37–53
- [26] Shanahan M 2005 Global access, embodiment, and the conscious subject. *Journal of Consciousness Studies* 12: 46–66
- [27] Shanahan M 2010 *Embodiment and the inner life: cognition and consciousness in the space of possible minds*. Oxford: Oxford University Press
- [28] Aleksander I, Awret U *et al* 2008 Assessing artificial consciousness. *Journal of Consciousness Studies* 15: 95–110
- [29] Chrisley R 2009 Synthetic phenomenology. *International Journal of Machine Consciousness* 1: 53–70
- [30] Ziemke T 2008 On the role of emotion in biological and robotic autonomy. *BioSystems* 91: 401–408
- [31] Carruthers P 2001 *Précis of phenomenal consciousness*, [http://lgxserver.uniba.it/lei/mind/forums/002\\_0002.htm](http://lgxserver.uniba.it/lei/mind/forums/002_0002.htm) [accessed 1.5.08]
- [32] Buttazzo G 2001 Artificial consciousness: utopia or real possibility? *Spectrum IEEE Computer* 34(7): 24–30
- [33] Holland O (Ed.) 2003 *Machine consciousness*. New York: Imprint Academic
- [34] Adami C 2006 What do robots dream of? *Science* 314(58): 1093–1094
- [35] Aleksander I 2008 Machine consciousness. *Scholarpedia* 3(2): 4162
- [36] Aleksander I, Awret U *et al* 2008 Assessing artificial consciousness. *Journal of Consciousness Studies* 15(7): 95–110
- [37] Buttazzo G 2008 Artificial consciousness: hazardous questions. *Journal of Artificial Intelligence and Medicine* (special issue on Artificial consciousness) 44(2): 139–146
- [38] Chrisley R 2008 The philosophical foundations of artificial consciousness. *Journal of Artificial Intelligence and Medicine* (special issue on Artificial consciousness) 44(2): 119–137
- [39] Manzotti R and Tagliasco V 2008 Artificial consciousness: a discipline between technological and theoretical obstacles. *Journal of Artificial Intelligence and Medicine* (special issue on Artificial consciousness) 44(2): 105–117
- [40] Chalmers D J 1996 *The conscious mind: in search of a fundamental theory*. New York: Oxford University Press
- [41] McGinn C 1989 Can we solve the mind–body problem? *Mind* 98: 349–366
- [42] Harnad S 2003 Can a machine be conscious? How? *Journal of Consciousness Studies*
- [43] Kim J 1998 *Mind in a physical world*. Cambridge, MA: MIT Press
- [44] Hayes-Roth B 1985 A black board architecture for control. *Artificial Intelligence* 26(3): 251–321
- [45] Currie K and Tate A 1991 O-plan: the open planning architecture. *Artificial Intelligence* 52(1): 49–86
- [46] Koch C 2004 *The quest for consciousness: a neurobiological approach*. Englewood, CO: Roberts & Company Publishers
- [47] Moravec H 1988 *Mind children: the future of robot and human intelligence*. Cambridge, MA: Harvard University Press
- [48] Kaku M 2014 *The future of the mind: the scientific quest to understand, enhance, and empower the mind*. New York, United States: Doubleday
- [49] Anderson M L and Oates T 2007 A review of recent research in meta-reasoning and meta-learning. *AI Magazine* 28(1): 7–16
- [50] Baars B J 1998 *A Cognitive theory of consciousness*. New York, United States: Cambridge University Press
- [51] Blackmore S 2002 There is no stream of consciousness. *Journal of Consciousness Studies* 9(5–6): 17–28
- [52] Chrisley R 2003 Embodied artificial intelligence [review]. *Artificial Intelligence* 149(1): 131–150
- [53] Clowes R, Torrance S and Chrisley R 2007 Machine consciousness—embodiment and imagination [editorial material]. *Journal of Consciousness Studies* 14(7): 7–14
- [54] Densmore S and Dennett D 1999 The virtues of virtual machines. *Philosophy and Phenomenological Research* 59(3): 747–761
- [55] Gamez D 2008 Progress in machine consciousness. *Consciousness and Cognition* 17(3): 887–910
- [56] Haikonen P O A 2007 Essential issues of conscious machines. *Journal of Consciousness Studies* 14(7): 72–84
- [57] Rolls E T 2007 A computational neuroscience approach to consciousness [review]. *Neural Networks* 20(9): 962–982
- [58] Rosenthal D M 1991 *The nature of mind*. Canada: Oxford University Press

- [59] Sloman A and Chrisley R 2003 Virtual machines and consciousness. *Journal of Consciousness Studies* 10(4–5): 133–172
- [60] Sun R 1997 Learning, action and consciousness: a hybrid approach toward modelling consciousness. *Neural Networks* 10(7): 1317–1331
- [61] Taylor J G 2007 CODAM: a neural network model of consciousness. *Neural Networks* 20(9): 983–992
- [62] Velmans M 2002 Making sense of causal interactions between consciousness and brain. *Journal of Consciousness Studies* 9(11): 69–95
- [63] Velmans M 2009 How to define consciousness: and how not to define consciousness. *Journal of Consciousness Studies* 16(5): 139–156
- [64] Chrisley R 1995 Non-conceptual content and robotics: taking embodiment seriously. In: Ford K, Glymour C and Hayes P (Eds.) *Android epistemology*. Cambridge: AAAI/MIT Press, pp. 141–166
- [65] Hirose N 2002 An ecological approach to embodiment and cognition. *Cognitive Systems Research* 3: 289–299
- [66] Shanahan M P 2005 Global access, embodiment, and the conscious subject. *Journal of Consciousness Studies* 12(12): 46–66
- [67] Pfeifer R, Lungarella M *et al* 2007 Self-organization, embodiment, and biologically inspired robotics. *Science* 5853(318): 1088–1093
- [68] Sanz R 2005 Design and implementation of an artificial conscious machine. In: *Proceedings of IWAC2005*, Agrigento
- [69] Rockwell T 2005 *Neither ghost nor brain*. Cambridge, MA: MIT Press
- [70] Manzotti R 2007 From artificial intelligence to artificial consciousness. In: Chella A and Manzotti R (Eds.) *Artificial consciousness*. London: Imprint Academic
- [71] Arkin R C 1998 *Behavior-based robotics*. Cambridge, MA: MIT Press
- [72] Ebbinghaus H 1908 *Abriss der psychologie*. Leipzig: Von Veit
- [73] Nemes T 1962 *Kibernetik gépek*. Budapest: Akadémiai Kiadó
- [74] Putnam H 1964 Robots: machines or artificially created life? *The Journal of Philosophy* 61(21): 668–691
- [75] Miller G 2005 What is the biological basis of consciousness? *Science* 309: 79
- [76] Seth A, Dienes Z *et al* 2008 Measuring consciousness: relating behavioural and neurophysiological approaches. *Trends in Cognitive Sciences* 12(8): 314–321
- [77] Aleksander I and Dunmall B 2003 Axioms and tests for the presence of minimal consciousness in agents. *Journal of Consciousness Studies* 10: 7–18
- [78] Searle J 1980 Minds, brains, and programs. *Behavioral and Brain Sciences* 3: 417–457
- [79] Koch C and Tononi G 2008 Can machines be conscious? *IEEE Spectrum* (June): 47–51
- [80] Block N 1995 On a confusion about a function of consciousness. *Behavioural and Brain Sciences* 18: 227–287
- [81] Tononi G 2004 An information integration theory of consciousness. *BMC Neuroscience* 5(42): 1–22
- [82] Liljenström H 2011 Intention and attention in consciousness dynamics and evolution. *Journal of Cosmology* 14
- [83] Lycan W G 1981 Form, function, and feel. *The Journal of Philosophy* 78(1): 24–50
- [84] Moore G E 1965 Cramping more components onto integrated circuits. *Electronics* 38(8): 114–117
- [85] Moravec H 1999 *Robot: mere machine to transcendent mind*. New York: Oxford University Press
- [86] Hofstadter D R and Dennett D C 1981 *The mind's I: fantasies and reflections on self and soul*. New York: Basic books
- [87] Hofstadter D R 1979 *Godel, Escher, Bach: an eternal golden braid*. New York: Basic Books
- [88] Minsky M 1968 Matter, mind, and models. In: Minsky M (Ed.) *Semantic information processing*. Cambridge, United States: MIT Press, pp. 425–432
- [89] Minsky M 1986 *The society of mind*. New York: Simon & Schuster
- [90] McCarthy J 2005 Making robots conscious of their mental states. In: *Proceedings of the Machine Intelligence Workshop*, retrieved December 31, 2005 from <http://www.formal.stanford.edu/jmc/chinese.html>
- [91] McCarthy J 1995 Todd Moody's zombies. *Journal of Consciousness Studies*, retrieved from <http://www-formal.stanford.edu/jmc/zombie/zombie.html>
- [92] Moody T C 1994 Conversations with zombies. *Journal of Consciousness Studies* (12): 196–200
- [93] Dennett D C 1969 Content and consciousness. In: *International Library of Philosophy and Scientific Method Series*. London: Routledge and Kegan Paul
- [94] Perlis D 1997 Consciousness as self-function. *Journal of Consciousness Studies* 4(5–6): 509–625
- [95] Sloman A 1978 *The computer revolution in philosophy*. Hassocks, Sussex: The Harvester Press
- [96] Smith B C 1984 Reflection and semantics in LISP. In: *Proceedings of POPL'84, 11th ACM SIGACT-SIGPLAN symposium on Principles of Programming Languages*, Salt Lake City, Utah, USA, pp. 23–35
- [97] Harnad S 1990 The symbol grounding problem. *Physica D* 42: 335–346
- [98] Sloman A 2002 The irrelevance of Turing machines to artificial intelligence. In: Scheutz M (Ed.) *Computationalism: new directions*, pp. 87–127
- [99] Said K A M, Newton-Smith W H, Viale R and Wilkes K 1990 *Modelling the mind*. Oxford: Oxford University Press, pp. 63–82
- [100] Allen W 1971 *Getting even*. United States: Random House Publisher, ISBN: 978-0394473482
- [101] Dasgupta S 1949 *A history of Indian philosophy*, 4 vols. Cambridge: University Press
- [102] Deussen P 1912 (a) *The philosophy of the Upanishads*. Translated by Geden A S and Clark T. Edinburgh: T. & T. Clark (b) *The system of the Vedānta*. Translated by Johnston C. Chicago: Open Court
- [103] Garbe R 1926 Vedānta. In: Hastings J (Ed.) *Encyclopedia of religion and ethics*. Edinburgh: T. & T. Clark, 13 vols., vol. 2, p. 597
- [104] Geden A S 1921 Upanishads. Hastings J (Ed.) *Encyclopaedia of religion and ethics*. Edinburgh, London: T. & T. Clark, vol. 12, pp. 540–548
- [105] Kenny D T 2016 A brief history of psychoanalysis: from Freud to fantasy to folly. *Psychotherapy and Counselling Journal of Australia* 4(1)



- [106] Carus P 1891 *The soul of man: an investigation of the facts of physiological and experimental psychology*. Chicago: Open Court Publishing Co.
- [107] Hume R E 1931 *The thirteen principal Upanishads*, 2nd ed. rev. by Haas G C O. London: Oxford University Press
- [108] Verse 42, Chapter 3, The Bhagwad-Gita.
- [109] Safaya R 1975 *Indian psychology: a critical and historical analysis of the psychological speculations in Indian philosophical literature*. New Delhi: Munshiram Manoharlal Publishers Pvt. Ltd.
- [110] The Katha-Upanisad, 1.3.3-4
- [111] Samkhya Pravāsana Sūtra, V. 129; iii. 2021
- [112] Radhakrishnan S 2008 *Indian philosophy*, vol. 2 with an introduction by Mohanty J N, p. 281
- [113] Misra S Upaskara Commentary of Vaiśeṣika Sūtra -III, 1,4
- [114] Prakāraṇa pañcika, P-63
- [115] Radhakrishnan S 2008 *Indian philosophy*, 2nd ed., vol. 2, p. 680 with an introduction by Mohanty J N
- [116] Århem P and Liljenström H 2007 Beyond cognition – on consciousness transitions. In: Liljenström H and Århem P (Eds.) *Consciousness transitions – phylogenetic, ontogenetic and physiological aspects*. Amsterdam: Elsevier, pp. 1–25
- [117] Anderson J 1983 *The architecture of cognition*. Cambridge, MA: Harvard University Press
- [118] Newell A 1990 *Unified theories of cognition*. Cambridge, USA: Harvard University Press
- [119] Maudlin T 1989 Computation and consciousness. *The Journal of Philosophy* 86(8): 407
- [120] Sprevak M 2007 Chinese rooms and program portability. *The British Journal for the Philosophy of Science* 58(4): 755
- [121] Pylyshyn Z W 1984 *Computation and cognition*. Cambridge: Cambridge University Press
- [122] Hinton G E 2005 Computation by neural networks. *Nature Neuroscience* 3: 1170
- [123] Olshausen B A and Field D J 1997 Sparse coding with an over complete basis set: a strategy employed by v1?. *Vision Research, Elsevier* 37(23): 3311–3325
- [124] Rao R P N and Ballard D H 1999 Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive field effects. *Nature Neuroscience* 2: 79–87
- [125] James W 1890 *The principles of psychology*. Cambridge, MA: Harvard University Press
- [126] Popper K, Lindahl B I B and Århem P 1993 A discussion of the mind–brain problem. *Theoretical Medicine* 14: 167–180
- [127] Lindahl B I B and Århem P 1994 Mind as a force field: comments on a new interactionistic hypothesis. *Journal of Theoretical Biology* 171: 111–122
- [128] Delbrück M 1986 *Mind from matter?* Palo Alto, CA: Blackwell Scientific Publishers
- [129] Penrose R 1989 *The emperor's new mind*. GB: Oxford University Press
- [130] Penrose R 1994 *Shadows of the mind*. GB: Oxford University Press
- [131] Good I J 1965 Speculations concerning the first ultra-intelligent machine. *Advances in Computers* 6: 31–88