



universität  
wien

# Masterarbeit

Titel der Masterarbeit

Consciousness and Cognition: A Functionalist Approach  
and Critic

Verfasser

Mehrdad Farahmand

angestrebter akademischer Grad

Master of Science (MSc.)

Wien, im Jänner, 2011  
Studienkennzahl It. Studienblatt  
Masterarbeitgebiet It. Studienblatt

Betreuer

066 013  
Middle European Interdisciplinary Master Programme in  
Cognitive Sciences  
Univ.-Prof. Dr. Olga Markic

## Curriculum Vitae

### Education:

PhD, Philosophy, University of Vienna, in Progress

Title: Consciousness and Cognition: A Processual and Dynamic Perspective (Currently being completed)

M.D, Medicine, Medical University of Vienna, in Progress (to be completed by 07. 2011)

Ms, Cognitive Sciences, University of Vienna, in Progress (Currently being completed)

Title: Cognition and Consciousness: A Functionalist Approach and Critic

M. A., Philosophy, California State University Los Angeles, 2003

Concentrations: Philosophy of Mind

Thesis: The Odyssey of Consciousness

B.S., Behavioral Neurosciences (Psychobiology), University of Southern California, 1996

Concentrations: Psychology, Special Education

Honors Thesis: The Role of Alternative Splicing of mRNA in Sexual Differentiation.

### Experience:

Lecturer

University of Vienna, 2008- present

Course: Philosophy, Cognitive Sciences

Lecturer (Part-time Faculty), 2000 - 2004

California State University Los Angeles

Course: Philosophy

Part-time Faculty 2000 - 2004

Glendale Community College

Courses: Philosophy

Molecular Biology Research Assistant and Labor Technician 1994-1999

University of Southern California

### Grants and Fellowships:

Howard Hughes Medical Institute Undergraduate Scholarship and Traineeship

### Awards and Honors:

Who is Who Among America's Teachers, Excellence in Teaching Award, 2003, 2002.

California State University Regents Alumni Association Honors, 2004, 2003, 2002.

California State University Academic Excellent Award, 2004, 2003, 2002.

Howard Hughes Medical Institute Scholarship, 1994-1996.

## Table of Contents

<b>Abstract</b>	<b>4</b>
<b>Abstrakt</b>	<b>5</b>
<b>1. Fundamentals, Background, and History</b>	<b>6</b>
<b>1.1 The problem of consciousness</b>	<b>6</b>
<b>1.2 Mental Causation</b>	<b>11</b>
<b>1.3 The Historical and Theoretical Background</b>	<b>14</b>
<b>1.3.1 Dualism</b>	<b>16</b>
<b>1.3.1.1 Cartesian Interactionist Dualism</b>	<b>17</b>
<b>1.3.1.2 Modern Dualism</b>	<b>22</b>
<b>1.3.1.3 Problems and Plausibility</b>	<b>25</b>
<b>1.3.2 Mentalism and Idealism</b>	<b>29</b>
<b>1.3.3 Material Monism</b>	<b>33</b>
<b>1.3.3.1 Reductive Materialism</b>	<b>35</b>
<b>1.3.3.2 Eliminative Materialism</b>	<b>42</b>
<b>1.3.3.3 Non-reductive Physicalism &amp; Property Dualism</b>	<b>46</b>
<b>2. Functionalism and Its Consequences</b>	<b>54</b>
<b>2.1 Methodological and Analytic Behaviorism</b>	<b>54</b>
<b>2.2 Functionalism and Cognitivism</b>	<b>58</b>
<b>2.2.1 The Representational Theory Of Mind</b>	<b>63</b>
<b>2.3 HOT theory of Consciousness</b>	<b>67</b>
<b>2.4 Global Workspace Theory</b>	<b>72</b>
<b>2.5 Problems and Plausibility</b>	<b>79</b>
<b>2.5.1 China Brain</b>	<b>80</b>
<b>2.5.2 Chinese Room argument</b>	<b>82</b>
<b>3. The Empirical Evidence</b>	<b>88</b>
<b>3.1 Consciousness Correlates</b>	<b>88</b>
<b>3.2 Perception: The Visual Information Processing Paradigm</b>	<b>93</b>
<b>3.3 Attention</b>	<b>101</b>
<b>3.4 The Memory Paradigm</b>	<b>109</b>
<b>3.5 Self</b>	<b>125</b>
<b>3.6 Concluding Remarks</b>	<b>129</b>
<b>4. Final Thoughts</b>	<b>131</b>
<b>Bibliography</b>	<b>137</b>

## **Abstract**

In this project, I examine the nature of consciousness from a functionalist perspective. Within this context I present the traditional computationalist as well as modern versions of functionalism such as High Order Thought (HOT) and Global Workspace theory (GWS).

I begin the work with a critical presentation of the fundamental questions and concerns in this discussion. Furthermore, I present the basic tenets of competing theories of mind and consciousness and put forth their supporting as well as undermining arguments. The point of this approach is twofold. Firstly, it is of paramount importance to put functionalism within its proper intellectual and historical context, since all theories are historical beings and must be understood as such. Secondly, another purpose would be to present the motivations for a functionalist perspective. I also present the current empirical evidence in neurosciences with respect to the nature of consciousness and cognition.

My contention is that functionalism provides a good theory of cognition and mind. However, it fails to account for consciousness since it cannot account for its essential properties such as intentionality and subjectivity. Furthermore, functionalism does not consider the proper mode of human existence, which is meaning and semantics driven and considers human consciousness in a syntactical and computational manner only.

### **Abstrakt**

In diesem Projekt präsentiere ich eine kritische Analyse von funktionalistischer Behandlung von Bewußtsein und Kognition. Ich fange diese Arbeit mit einer vergleichenden historischen Präsentation von konkurrierenden Theorien von Bewußtsein und Kognition an und biete Argumente warum diese Theorien versagen.

Auf diesem Weg beschreibe ich klassische funktionalistische Theorie wie moderne funktionalistische Theorien wie High Order Thought (HOT) und Global Workspace (GW) Theorien. Zusätzlich analysiere ich die Argumente gegen Funktionalismus.

Es ist meine Position, daß Funktionalismus nicht in der Lage ist zwischen Geist und Bewußtsein zu differenzieren. Ich behaupte, daß Funktionalismus eine gute Theorie des Geistes ist, aber sie kann nicht Bewußtsein erklären, weil sie vernachlässigt die essentiellen Eigenschaften von Intentionalität und Qualität.

## Chapter One

### Fundamentals, Background, and History

In this work, I engage in a historical and analytic presentation of functionalism's treatment of the question of nature of cognition and consciousness. In this path, as a representative of a modern approach to this topic, I have chosen to introduce functionalism in general and specifically the Higher Order Thought (HOT) as it is understood and explicated by Bernard Baars, who uses the metaphor of theatre as the proper way to understand the whole cognitive machinery and consciousness. Within the context of the presentation of the HOT theory, I also present the standard scientific views with respect to cognition and consciousness. However, we must begin our journey with some background information about the history of the problem of consciousness and cognition and the fundamental questions facing any theory at the present time. It is to this task that I turn now in this introductory chapter.

#### 1.1 The problem of consciousness

The query about the nature of the consciousness and its relation to the world is one of the confounding questions of the philosophical discourse. There appears that there is no objective account, which is able to encapsulate the essence of consciousness. For example, any attempt to define consciousness in terms of some psychological function, such as awareness, attention seems to leave out an essential property of consciousness, namely why conscious states *feel* a certain way. There appears something overwhelmingly deep and indescribable about the subjective nature of consciousness, which lends itself to illustration through examples, but the objective definitions seem to be not available.

The problem must be approached with two considerations in mind. One is that consciousness is that, which makes our experience of the world possible. We could receive information from the world and process that information subconsciously, or

unconsciously (whatever these terms refer to). However, an experience is by definition conscious. Hence, consciousness makes it possible for us to understand and think about the world, in which we live, but how can we understand consciousness itself. This very problem is the impetus to reduce consciousness to something more manageable such as cognition or language. However, these approaches leave out some major and fundamental properties of consciousness. Most of the time they get rid of the very phenomenon they aspire to comprehend.

The second aspect worthy of our consideration within this context is the historical context of this discussion. It is paramount for us to remind ourselves that problems and questions have a history of their own and this history informs the way events evolve and unfold. Problems are historically rooted. So, any viable treatment of this issue must address the following questions:

1. What is consciousness?
2. What is the causal relationship between consciousness and brain/body?
3. What does consciousness do? In other words, what is the function of the consciousness?
4. What are the material substrate, or correlates, of consciousness?

Any discussion of conscious mental states, such as pains, sensory experiences, or dreams, combines the subjective and objective conceptions of these states. However, one should, but often does not stop to, differentiate the subjective aspect, of what it is *like* to have the experience, from the objective features, such as the psychological function, and the physical structure of that experience. This conflation is of no importance in our daily actions, in view of the fact that the subjective feature and the corporeal features seem to go always together. However, any serious philosophical and scientific discussion must defer to that distinction.

Thomas Nagel points to the same intuition when he raises the renowned question: “What is it like to be a bat?”<sup>1</sup> Science tells us that bats negotiate their way around their environment by echolocation. They produce high-pitched sounds and by way of the echoes, they devise the location of physical objects in the immediate environment. Consequently, bats are not aware of just raw sounds, but they are aware of physical objects in the same way that vision makes us aware of objects and not light waves. In raising this question, Nagel wants us to make a distinction between the two conceptions of conscious experience, the objective and the subjective. The main point is that we have no notion of the subjective facet of bat’s experience. In our own case, we do not worry about this peculiarity, because we consider human consciousness simultaneously in subjective and objective terms. The bat case, on the other hand, compels us to make that distinction, for the reason that we have no insight of how the bat experience feels like, in spite of having many objective facts about it. In other words, science informs us an extensively about bat brain, but we have no idea about what is it like to be a bat. Accordingly, Nagel identifies a characteristic of experience, which ostensibly eludes scientific explanation. This subjective facet is the *what-is-it-likeness*, *qualia*, or the *phenomenology* of experience.

Another important mark of consciousness is *intentionality*. A state is intentional, when it is *about* something and it refers to something. The major conundrum is how a physical state can be *about* something. How can words, marks on paper or patterns of sound, stand for something? We could postulate that the marks on paper and patterns of sound are *about* something, because we comprehend what they mean. However, this response just thrust the query only one step further and it does not settle it.

Qualia and intentionality seem to be two fundamental properties of consciousness, which emerge from our everyday understanding of our experience. We have an intuitive understanding of how it is to be conscious in our waking moments and dreams as

---

<sup>1</sup> Nagel T., (1974), *What is it like to be a bat?*, Philosophical Review 83.4, pp. 435-450.

opposed to be in a state of dreamless sleep. Furthermore, we have an intuitive understanding of what does it mean to be conscious of a thought, an object, or an emotion as opposed to be not conscious of it. Hence, we have pre-analytic notion of qualia and intentionality. This allows us to give an ostensive definition of consciousness, which allows us to outline the inclusive parameters and differentiate them from the excluding factors. Hence, we can say *a priori* that a person is conscious, if he/she has an experience of something. In other words, a person is conscious in the presence of phenomenal content. This is a primary definition, which constitutes a starting point and not an end result. In order to define a phenomenon properly, we need to outline its relation to other things, to put it in its proper context. This means that we have determined its function, its causal efficacy, and its correlates. In other words, a proper definition of consciousness would tell us its function, what causes it, what it causes, and what its structural correlates are.

Chalmers differentiates between the *hard problem* and the *easy problem* of consciousness.<sup>2</sup> According to Chalmers, the *easy problem* correlates with the objective study of brain states. At the objective level, we can designate causal functions to different physical and psychological states and structures. The objective investigation of mind, and the brain, is relatively uncomplicated as far as it is possible to investigate the question through the tools of scientific investigation. One might investigate pain as a state, which is instigated by bodily damage. Moreover, it leads to a behavior evading the noxious stimulus. Furthermore, we can investigate the various manifestations of the pain function and circuitry in different species. Yet, not any of the structural and functional investigation divulges anything about the phenomenology of pain. Apparently, causal and functional explanations overlook the qualia of mental states.

---

<sup>2</sup> Chalmers D., (1986), *The Conscious Mind: In Search of a Fundamental Theory*, Oxford University Press, pp. 32-42, 103-104, 131-132.

The elucidation of phenomenal aspect, consequently, represents the *hard problem* of consciousness. The disparity between our aptitudes to resolve the *easy problem* as opposed to the *hard problem* comprises the *explanatory gap*.

So, there seems to be an intractable problem at hand. Moreover, there seems to be only a limited number of possible solutions to this problem. Simply put the question is where does consciousness fit in our understanding of the universe. The possible responses are:

1. Consciousness is an illusion. There is only matter.
2. Consciousness is reducible to matter and it is a manifestation of material complexity of a certain kind.
3. Consciousness is a fundamental feature of the universe along with matter.
4. Consciousness is the fundamental element of the universe and matter is an illusion produced by the conscious mind.
5. There is a more fundamental aspect of reality and existence, which is ineffable in the sense that is not conceivable in a conceptual way, but it lends itself to another faculty, namely intellectual intuition and certain kind of experience. This is the intuition behind some forms of double-aspect theories with an axiological and religious emphasis.

We should also take stock of the facts:

1. There is brain and nervous system.
2. There is cognition, mental functions, and structures, which are dependent on the brain, the nervous system for their action. In other words, an end to physical activity will entail the end of mental action as well.
3. There is also subjective consciousness. This is based on the personal experience of each one of us. It is not objectively verifiable. However, this subjectivity is the condition that makes all my objectively verifiable action possible. It all begins from my personal perspective including my scientific hypothesis and this is utterly subjective in nature.

An evaluation of the possible theoretical frameworks mentioned above against the facts reveals that the positions 1 and 4 are not maintainable. One thing we cannot do is to deny the reality of matter or mind. The question becomes then what nature of the relationship between the two can possibly be. Positions 2 and 3 try to ask this question. They maintain either the mind is reducible to matter, or the mind is a property of matter, or mind and matter constitute a dual reality interacting with each other or they stand in a parallel relationship to each other. These possibilities reflect the history of theories mind in the Western traditions. Position 5 reflects a religious perspective, since it postulates a reality not accessible to discursive thought but capable of being revealed to intuition and mystic experience.

## 1.2 Mental Causation

One of the most important questions that a theory of mind must explain is how can mental states be causally efficacious in a physical world. There are three causal relationships, which need to be elucidated: *physical-to-mental* causation, *mental-to-physical* causation, and *mental-to-mental* causation. The case of *physical-to-mental* causation seems to be quite straightforward. Here, we can use the case of the sensations just as physical injuries, which cause pains. As we saw before, the physical-mental causation can be satisfactorily explained in terms of stimulus (input), response (output), and interaction among relevant mental states caused by the corresponding physiological states. Hence, physical-mental causation can be adequately explained by a functional analysis. *Mental-mental* causations are equally intuitively imaginable. Here, we speak of the case where mental states such as beliefs, desires, and feelings cause other mental states: having the belief that you are about to meet a very good friend after a long time might cause you to feel happy or anxious. The fact that our beliefs, desires, and feelings constitute a complex network of causal relations does not come as surprise to most of us.

*Mental-physical* causation, however, seems to be much harder to account for than at first glance. At the core of the problem lies the indispensable principle of causal

closedness of the physical world. According to this principle, if a  $x$  is a physical phenomenon and  $x$  has a cause  $y$ , then  $y$  must be also a physical event, property, or state. To explain a physical event, we must never resort to non-physical explanations. The reason for this principle is rooted deeply in physics. Cause and effect relationships can be in physics in terms of transfer of energy between systems and work that system does on another system. To transfer energy from one system to another and consequently do work on a system requires the application of force. This means the impetus to actively transfer energy and do work is application of force. There are, however, a certain number of forces postulated in physics:

1. Gravitational force.
2. Electromagnetic force.
3. Strong nuclear force.
4. Weak nuclear force.

Consequently, any work done by, or on, any physical system must be initiated by one of these forces or a combination of them. Now, when we claim that mental events can cause physical events, we can mean only two things:

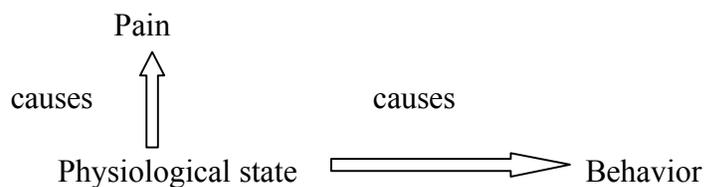
- a. Mental events cause physical events through of the above - mentioned forces.
- b. Mental events cause physical events through an extra force not included in the above list.

If we mean 'a', then we must be able to explain how a belief through gravitational force, electromagnetic force, strong nuclear force, or weak nuclear force can cause a physiological state. This seems to be an erroneous postulation based on what we know about physics. However, if we assert 'b', then we must violate the principle of thermodynamics that neither energy is created, nor it is destroyed. A supposition of a non-physical force causally efficacious in the physical world would be precisely asserting the creation of new energy.

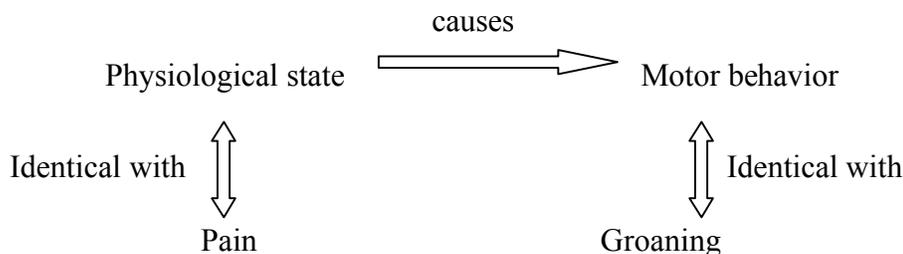
So, it seems that respecting the causal closedness principle entails rejection of the *mental-physical* causation. This is precisely what epiphenomenalism does.

Epiphenomenalism is the view that asserts that mental states are causally impotent by-products of physical causal relationships. This is a deeply unsatisfactory position, since it denies our most basic intuitions about the workings of our minds and the world.

Moreover, epiphenomenalism adds another item to our ontology of the world: an epiphenomenon. This is something in the world causes nothing, but it exists and we know of its existence. However, the question is how do we know anything? The answer is that we know x, because x is causally efficacious in the world. We know x through its causal interactions. One of these causes effects is causing us to know it. In the case of an epiphenomenon, there are no causal relations but we know of it. This is deeply contradictory. Hence, it is nonsensical. Mind seems to be the only epiphenomenon in the universe. This seems to be too arbitrary or an attempt to explain away a difficult problem and not to explain it.



Another option would be identify mental events with physical events. In other words, we reduce the mental to the physical. This is reductive materialism, which could not account the multiple realizability of mental states and led us to embrace functionalism. This point brings us to the functionalist solution to this problem.



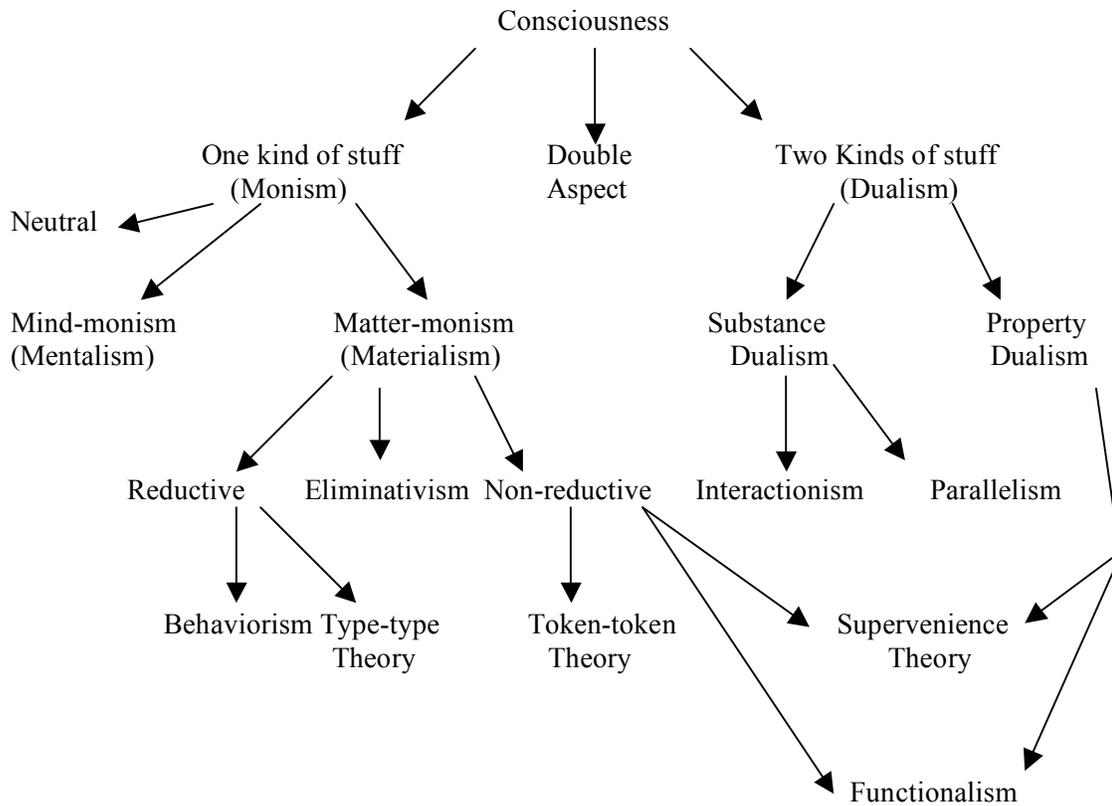
The functionalist would claim that the problem of mental causation actually disappears with functionalism. This is a consequence of the functional analysis of a mental state such as pain. Functionalism does not postulate the existence of an independent mental state, which causes a behavior. Pain is understood in terms of the totality of input-processing-output causal network. Pain signifies this whole functional state. Behavior is not caused by a mental state, but it is a component of a causal network and functional state. In other words, the complete system of interrelations between inputs, outputs, and relevant causal nodes comprise a mental state. This sounds like a very convincing position. However, there is a major problem with it. Functionalism defines pain, for instance, in terms of input-processor-output.

**Pain**= {Stimulus (input)  $\implies$  Neural Processing  $\implies$  Behavior (output)}

However, is this the complete picture of what we mean when we talk about pain. What we mean with pain is how it actually feels, its qualia. That is conspicuously absent from the functionalist account. In fact, all conscious mental states are fundamentally either qualitative, intentional, or both. Qualia and intentionality make the mental causation a seemingly intractable problem. So the problem of mental causation is intimately related to the question of the nature consciousness.

### **1.3 The Historical and Theoretical Background**

Traditionally, the landscape of the theories of consciousness is fairly varied: substance dualism, mentalism, reductive materialism, and non-reductive materialism or property dualism.



This landscape is organized around some fundamental questions in philosophy and science:

1. Monism or dualism? Is the universe composed of one kind of fundamental stuff (monism) or two kinds of fundamental stuff (dualism)?
2. Realism or idealism? Does the existence of the universe depend on subjective observation (idealism) or is it objective independent of any subject (realism)?
3. Does our knowledge of the world correspond to what is actually in the world, objective? Or our knowledge of the world is a coherent component of an intersubjective web of beliefs? Or is it simply private? We just know the content of our own minds.

The answers to these questions give rise to different theoretical frameworks, which have evolved over time and hence constitute the history of the problem at hand.

### 1.3.1 Dualism

The instinct and intuition behind a dualistic view of the world is a deep and an old one. It arises from a sense that we are more than just our bodies and our existence transcends our physical properties and final demise. This intuition is rooted in history, culture, and arts. In fact, one could argue that organized ceremonies and symbolism thereof, which constitute the beginning of arts and culture is preoccupied with death and comes after it. It seems that for mankind it was from the beginnings of the history a self-evident fact that one survives and transcends material existence in one way or another. Archeological evidence from the Paleolithic period points to elaborate burial ceremonies, which show this preoccupation of fundamental metaphysical questions such as finality and transcendence. This concern is reflected in the ancient art as well as medicine.

With the rise of ancient civilizations, the concern with afterlife and intuition that our existence transcends the physical yet it is deeply connected with the physical gave rise to organized religion and the symbolism connected to it in form of mythology. Certainly, the Egyptian religion and mythology is deeply rooted in a dualistic worldview. The kingdom of Osiris constitutes the land of the Dead, where hearts of men are judged based on their goodness or evil and are furthermore promoted to dwell in eternal peace or are condemned to eternal damnation. Here we see a close relationship between axiology and ontology. The question of what exists is closely related to what value it has and what it means. The kingdom of Osiris is given a concrete physical address in the West, where the desert begins.

In Greek philosophy and mythology, the dualist position found variety of complex treatment in symbolic and analytic forms. Moreover, this treatment is more akin to what we understand as mind, consciousness, and soul. Socrates makes a distinction between the *psyche* and the *nous*. According to the Greek mythology, the *psyche* is the non-

physical aspect of our existence, which is judged to be worthy of either eternal existence in peace and prosperity in the *Elysium*, eternal anxiety and fear in *Hades*. *Nous*, on other hand, is the intellectual intuition, seat of reasoning and intuition, source of immediate yet transcendent knowledge. The *nous* is also the source of axiology, insightful epistemology and ontological insight.

Plato elaborates the Socratic position by asserting that *nous* and *psyche* interact with the physical body. The body provides the source material for knowledge and understanding of the world through the workings of the sense data. The sense data gives us reflection of reality, *eikasias*. The highest possible of intellectual level of knowledge at this level is *pistis*, beliefs. We can only know what things seem to be, *doxa*. There are two streams of existence. One is ever changing and in constant flux. This aspect of existence is presented to us by our sense data. However, there is more fundamental dimension of existence, which is eternal and constitutes those everlasting patterns—the forms, which manifest themselves in space and time as changing appearances, which are mere reflections of *true* reality. The true reality is the Form of Good, which is the Form of Life and Being. It is the highest and deepest level of existence and can be known immediately by *noesis* (intellectual intuition) and can be arrived at through analytic mind and discursive reasoning, *dianoia*. This constitutes true knowledge, *episteme*. It unifies at once ontology, epistemology, and axiology.

### **1.3.1.1 Cartesian Interactionist Dualism**

The fundamental postulations of substance dualism are that conscious experience is fundamentally different from brain activities. However, this assumption raises further questions as to if the world contains subjective elements, then how do the subjective characteristics interrelate with the physical constituents that are essentially spatial and temporal. Additionally, which principles preside over the emergence of these subjective elements?

Descartes is the classical advocate of substance dualism. He maintains that there are two discrete but interacting realms, the mental—*res cogitans*, and the material—*res extensa*. Descartes supposes that the material sphere includes matter in motion, and all action occur by contact. According to Descartes, all physical effects are caused by bits of matter colliding with each other. Colors, sounds, smells, and all sensory impressions are not just objects in themselves, but they are impressions produced in us by the action of material particles on our sense organs. Matter in motion does not exhaust reality. Reality is composed of the mental domain as well. The realm of mind is constituted by thoughts, feelings, desires, and so on. The constituents of the mental realm have no spatial characteristics of matter, such as size, shape, and motion.

In the *Meditations*, the notion of *res cogitans* is more clearly illustrated:

“Thinking is another attribute of the soul; and here I discover what probably belongs to myself. This alone is inseparable from me. I am – I exist: this is certain; but how often? As often as I think; for perhaps it would even happen, if I should wholly cease to think, that I should at the same time altogether cease to be. I now admit nothing that is not necessarily true: I am, therefore, precisely speaking, only a thinking thing.”<sup>3</sup>

Temporality is the quality that the constituents of the mental realm share with the material domain. These two realms, however, can interact with each other regardless of their fundamental ontological disparity. Material causes can generate mental effects and mental causes can produce material effects.

Accordingly, human existence is made up of two independent yet interacting realms: *res extensa*—the material and *res cogitans*—the mental. While, the laws of mechanics govern the *res extensa*, *res cogitans* presents a different non-mechanical dimension of our existence. Descartes states in his *Discourse on Method* (Part V):

If there were machines, which have a resemblance to our body and imitated our actions as far as it was morally possible to do so, we should always have two very certain tests by

---

<sup>3</sup> Descartes, R., (1986) *Meditations*. London, Everyman, p. 88.

which to recognize that, for all that, they were not real men. The first is, that they could never use speech or other signs as we do when placing our thoughts on record for the benefit of others. For we can easily understand a machine's being constituted so that it can utter words, and even emit some responses to action on it of corporeal kind, which brings about a change in its organs; for instance, if it is touched in a particular part it may ask what we wish to say to it; if in another part it may exclaim that it is being hurt, and so on. But it never happens that it arranges its speech in various ways, in order to reply appropriately to everything that may be said in its presence, as even the lowest type of man can do. And the second difference is, that although machines can perform certain things as well as or perhaps better than any of us can do, they infallibly fall short in others, by which means we may discover that they did not act from knowledge, but only from the disposition of their organs. For while reason is a universal instrument which can serve for all contingencies, these organs have need of some special adaptation for every particular action. From this it follows that it is morally impossible that there should be sufficient diversity in any machines to allow it to act in all events of life in the same way as our reason causes us to act.<sup>4</sup>

According to Descartes, language and rational cognition grant humans certain flexibility and adaptability, which allow us to respond creatively and methodically to novel challenges in our environment. This ability to adapt through language and rationality is precisely what separates us from animals:

For it is remarkable fact that there are none so depraved, or stupid without even excepting idiots, that they cannot arrange different words together, forming of them a statement by which they make known their thoughts; while on the other hand there is no other animal, however perfect and fortunately circumstanced it may be, which can do the same. It is also a very remarkable fact that although there are many animals which show more dexterity than we do in some of their actions, we at the same time observe that they do not manifest dexterity at all in many others. Hence the fact they do better than we do,

---

<sup>4</sup> This passage from Descartes is represented in Flew, A., (1978) *Body, Mind, and Death*, New York; McMillan, p.127.

does not prove that they are endowed with mind, for in this case they would have more reason than any one of us, and would surpass us in all other things. It rather shows that they have no reason at all, and it is nature, which acts in them according to the disposition of their organs, just as a clock, which is only composed of wheels and weights, is able to tell the hours and measure the time more correctly than we do with all our wisdom.<sup>5</sup>

Consequently, a human being is best depicted as an embodied mind:

“I am not only lodged in my body as a pilot in a ship, but that I am very closely united to it, and so to speak so intermingle with it that I seem to compose with it one whole. For if that were not the case, when my body hurt, I, who am merely a thinking thing, should perceive this wound by understanding only, just as the sailor perceives by sight when something is damaged in his vessel....”<sup>6</sup>

Descartes makes furthermore the claim that our existence essentially should be understood in term of our soul and not body. This is more than an ontological claim. It is essentially an axiological statement, which asserts that the value of our existence is granted through our souls and not perishable bodies. The mind is the essence of our existence, since its activity survives all doubt and it produces doubt itself and one cannot doubt his own doubting:

“I then considered attentively what I was; and I saw that while I could feign I had no body, that there was no world, and no place existed for me to be in, I could not feign that I was not; on the contrary, from the mere fact that I thought of doubting about other truths it evidently and certainly followed that I existed . . . From this I recognized that I was a substance whose whole essence or nature is to be conscious and whose being requires no place and depends on no material thing. Thus this self, that is to say the soul

---

<sup>5</sup> *Ibid.* p.138.

<sup>6</sup> Descartes. *Works*, p. 192.

by which I am what I am, is entirely distinct from body, and is even more easily known; and even if the body were not there at all, the soul would be just what it is.”<sup>7</sup>

This can be encapsulated by the famous phrase: *cogito ergo sum*. The act of conscious cognition separates us from all other existents and it constitutes the essence of man.

As we implied earlier, Descartes maintains that mind and body interact with each other. This is a causal interaction, which obeys the laws of mechanics in general and fluid hydraulics in particular. Accordingly, external objects, which set the nerves in motion and action, stimulate the sense organs, which generate motion in the pineal gland in brain. The pineal gland then acts as causal link between the *res extensa* and the *res cogitans*, the mind. The pineal gland is the only asymmetric structure in the brain. Furthermore, it, anatomically, occupies a central place in the brain. Therefore, it was considered by Descartes to be unique and the best candidate for the locus of interaction between the mind and the body:

“It follows that the soul is really joined to the whole body, and that we cannot, properly speaking, say that it exists in any one of its parts to the exclusion of the others . . . because it is of a nature which has no relation to extension nor dimension, nor the properties of the matter of which the body is composed, but only to the conglomerate of its organs . . . It is likewise necessary to know that although the soul is joined to the whole body, there is yet in that a certain part in which it exercises its function more particularly than in all other . . . In examining the matter with care, it seems as though I had clearly ascertained that [this] part . . . is . . . a certain very small gland which is situated in the middle of [the brain] and so suspended above the duct whereby the animal spirits in its anterior cavities have communication with these in the posterior, that the slightest movements which take place in it may alter very greatly the course of these spirits; and reciprocally that the smallest changes which occur in the course of the spirits may do much to change the movements of this gland.”<sup>8</sup>

---

<sup>7</sup> Descartes, R., (1970) *Philosophical Writings*. Trans. Geach & Anscombe. New York: Nelson, p. 32.

<sup>8</sup> Descartes, *Discourse*, pp. 5, 7 - 9, 14 -17, & 22.

The choice of pineal gland is the product of state of the art knowledge in the time of Descartes. Therefore, it is antiquated. However, this is not of much bearing in this argument, since the main point is interaction and plausibility thereof and not the anatomical details that can always be updated. Descartes, however, does not explain how this interaction occurs. Realization of this problem led to the advent of *parallelism* and *occasionalism*, which postulate that the mental and the material do not interact with each other directly, but they are tuned to each other at any occasion by a third principle, namely God, as in Malebranche, or they were pre-established with each other through the principle simultaneity, as for Leibniz.

### **1.3.1.2 Modern Dualism**

The existence of consciousness seems to be an undeniable fact. Moreover, the deeply rooted intuitions and instincts that give rise to a dualistic worldview are culturally, socially, and psychologically present and effective. Hence, it comes as no surprise that there are modern theories of dualism, which need to be taken seriously and dealt with systematically. Emergence of modern technologies, on other hand, feeds the desire to re-approach the old question from new perspectives and with new methodologies. Consciousness plays such a central role to our mode of existence that it is imperative for science to address its existence and research its nature. Eccles states:

...Nowhere in the laws of physics or in the laws of derivative sciences, chemistry and biology, is there any reference to consciousness or mind. ...Regardless of the complexity of electrical, chemical or biological machinery there is no statement in the 'natural laws' that there is an emergence of this strange non-material entity, consciousness or mind. This is not to say that consciousness does not emerge in the evolutionary process but merely to state that its emergence is not reconcilable with the natural laws as presently understood.<sup>9</sup>

---

<sup>9</sup> Eccles, J.C, (1980) *The Human Psyche*, New York; Springer, p. 20.

Eccles bases his view on the premise that for any phenomenon to evolve, it must have a function and a relevant causal efficacy. It must do something, which aids the individual of a species to survive and adapt more successfully. Hence, consciousness must have a function and be causally efficacious otherwise it would not have evolved. On other hand, there are aspects of the self-conscious/conscious mind that seem to enjoy a nonmaterial existence. Eccles asserts:

Since they all...assert the causal ineffectiveness of consciousness per se, they fail completely to account for the biological evolution of consciousness, which is an undeniable fact. There is firstly, its emergence and then its progressive development with the growing complexity of the brain. In accord with evolutionary theory only those structures and processes that significantly aid in survival are developed in natural selection. If consciousness is causally impotent, its development cannot be accounted for by evolutionary theory. According to biological evolution mental states and consciousness could have evolved and developed only if they were causally effective in bringing about changes in neural happenings in the brain with consequent changes in behaviour. That can occur only if the neural machinery of the brain is open to influences from the mental events of the world of the conscious experiences, which is the basic postulate of dualist-interactionist theory.<sup>10</sup>

Eccles postulates a two-fold function of consciousness. Firstly, the function of consciousness is comprised in integrating the incoming information and sense data in the neocortex. Accordingly, consciousness creates a unified stream of conscious experience from a manifold of sense data—perceptual experience. This is the unifying function of consciousness. Secondly, consciousness stimulates neural networks and structures to give rise to a certain response and behavior in a coordinated and meaningful manner. Hence, consciousness has a function and it is causally efficacious. The net result is emergence of meaningful and coordinated experience and behavior. The body effects consciousness through sense data and information—in other words sensation—and consciousness affects the body through perception, cognition, emotion, and volition. Eccles, inspired by

---

<sup>10</sup> Eccles, (1980), p.20.

the findings of modern neurobiology, postulates the modular structure of neocortical neurons as the location of this interaction and does away with pineal gland hypothesis of Descartes.

The brain is organized threefold:

1. Neural structures with different functions such as basal ganglia, cerebellum, parietal lobe, frontal lobe, etc.
2. Neural pathways, which transmit information between variety of neural structures such as Papez circuits that transmit data between the limbic system, thalamus, hypothalamus, and frontal cortex.
3. Neural modules. The neurons are organized in columns of seven layers in cortex, three in cerebellum, and three in hippocampus. Information is received on certain layers and sends out on other layers. This modular structure of neurons is the basis of information processing capacity of brain.

Eccles states with regards this interaction:

The self-conscious mind is actively engaged in reading out from the multitude of liaison modules that are largely in the dominant hemisphere. The self-conscious mind selects from these modules according to attention and interest, and from moment to moment integrates its selection to give unity to even to the most transient experience.

Furthermore, the self-conscious mind acts upon these modules modifying their dynamic spatio-temporal patterns. Thus it is proposed that the self-conscious mind exercises a superior integrative and controlling role. A key component of this hypothesis is that the unity of conscious experience is provided by the self-conscious mind and not by the neural machinery of the liaison area of the cerebral area of the cerebral hemisphere.

Hitherto it has been impossible to develop any neurophysiological theory that explains how a diversity of brain events comes to be synthesized so that there is a unified conscious experience of a global or gestalt nature.<sup>11</sup>

---

<sup>11</sup> Eccles, (1980), p.49.

At the core, Eccles' theory of mind is no different from that from Plato and Descartes. Despite its modern formulation it still postulates a nonmaterial mind/consciousness in interaction with a material brain/body. The ontology is the same: there are two types of fundamentally distinct and separate kind of essential stuff. This introduces a schism in our ontology, epistemology, and axiology. Hence, the same objections apply to modern dualism as it does to classical dualism.

### **1.3.1.3 Problems and Plausibility**

The plausibility of dualism is based on its intuitive acceptability and its accordance to our common experience. Body and brain seem to be distinguishable, if not separable, from minds and consciousness. They seem to follow different laws of behavior. It seems like we could have a complete neurological/neurobiological profile of an individual, but still some essential psychiatric/psychological facts about the same person are outstanding. This points to the possibility that brain/body and mind/consciousness are fundamentally different. Moreover, our experience tells us that minds affect bodies and bodies affect minds. There seem to be a reciprocal causal relationship between the two spheres.

We cannot, however, leave our analysis at the level of our common sense intuitions and experiences. We need more to decide the plausibility of any theory. There we have guideline that is organized around whether the theory explains the nature of the phenomenon in question, its function, its causal efficacy, and its relationships and context.

Dualism asserts that mind/consciousness is essentially distinct from matter. However, it does not tell us what consciousness/mind is. Descartes calls consciousness/mind the substance that thinks, *res cogitans*. And he differentiates it from *res extensa*, which is spatial and temporal in nature and obey the laws of mechanics. All he says about *res cogitans* is that it does not obey the laws of mechanics and it is nor

spatial in nature and thus does not have extension. Hence, dualism does not really define consciousness in any useful manner.

A further thought along the same line is that dualism does not really distinguish between mind and consciousness. However, mind refers to mental processes that can be conscious—have phenomenal content—or they lack phenomenal content, hence unconscious. Consciousness does not seem to exhaust all mentality. Neither consciousness can be identified with any particular faculty of mind such as cognition, conation, or emotion. All these faculties can be, and are mostly, conscious, but consciousness cannot be reduced or identified with any one of them or the set of them all.

With respect to the question of function, Eccles's view seems to provide a plausible view of function of consciousness as the integrating and unifying principle of experience. It is that principle, faculty, which unites the multitude of sense data into a coherent experience. This is not, however, a view that logically follows from his dualist position. One does not need to assume dualism to assert this position. It is also historically not limited to Eccles either. Kant proposed a similar view with respect to consciousness with his theory of *transcendental unity of apperception*.

As I sit here and type these words, there is a sharp pain traveling through my head. However, there is no doubt that this pain is mine. I have no doubt that the thoughts rushing through my mind are mine. I have no doubt that the struggle to find the right words is mine. It does not feel as if there is pain in this room and I cannot locate its owner. I have no doubt that I am the owner of my experience. Where does this knowledge come from? I cannot arrive at this knowledge by inspecting the items of my experience. There is nothing in the sense data that would make this knowledge possible for me. In fact, all my experiences presuppose this ownership of my experience. Without this unity of apperception, I could not have any experiences at all. The knowledge of ownership of my experiences is not a posteriori, since it is presupposed by experience. Hence, it is a priori knowledge.

Kant recognized this crucial point and assigned the role of transcendental unity of apperception to transcendental consciousness. This unity of consciousness accounts for the ownership of experience. The ownership of experience is not subject to dispute. The empirical investigation might reveal the content of consciousness, but not its ownership. Apperception refers to all experience of which the subject is able to say 'this is mine'. Therefore, apperception is the foundation of self-consciousness and perceptive consciousness. Kant described it as the 'I think' that can be attached to all perceptual experience. It is the awareness that the perceptual experience belongs to me. Unity of apperception defines my point of view. There is never a doubt about the ownership of my experience. A doubt in the unity of apperception would mean that I stop having self-consciousness and empirical experience. For Kant, there are three elements involved in the conceptualization of experience. Firstly, there is the structuring of intuitions in time and space. Secondly, it is the unification of intuitions under one consciousness. Thirdly, it is the ability to organize all the intuitions into concepts of categories.

The prerequisite to all this is the possibility of apperceptive consciousness. The manifold representations, which are given in an intuition, would not be all my representations, if they did not belong to one self-consciousness. *Transcendental unity of apperception* is a formal unity, which all experience requires. When I stand on the beach and observe the prospect, all aspect of this experience is united. Transcendental unity of apperception is the formal unity that makes empirical consciousness and self-consciousness (my awareness of my experience) possible. Kant states: "it must be possible for the 'I think' to accompany all my representations; for otherwise something would be represented in me which could not be thought at all, and that is equivalent to saying that the representation would be impossible, or at least would be nothing to me. That representation which can be given prior to all thought is entitled intuition. All the manifold of intuition has, therefore, a necessary relation to the 'I think' in the same

subject in which this manifold is found”.<sup>12</sup> This is what we mean when we say that transcendental consciousness is foundational, and autonomous. Pure consciousness is the condition of all experience without being its object.

It must be possible for the ‘I think’ to accompany all my representations; For otherwise something would be represented in me which could not be thought at all, and that is equivalent to saying that the representation would be impossible, or at least would be nothing to me.<sup>13</sup>

Without sensibility no object would be given to us, without understanding no object would be thought. Thoughts without content are empty, intuitions without concept are blind. It is, therefore, just as necessary to make our concepts sensible, that is, to add the object to them in intuition, as to make our intuitions intelligible, that is, to bring them under concepts. These two powers or capacities cannot exchange their functions. The understanding can intuit nothing, the senses can think nothing. Only through their union can knowledge arise.<sup>14</sup>

The greatest shortcoming of dualism, however, lies in its inability to account for the interaction consciousness and brain. It cannot account for, or explain, the causal efficacy of the mental and consciousness while upholding this causal efficacy to be undeniable. The question can be put quite bluntly: how can an extended thing interact with an unextended thinking thing. For any two subjects or objects to enter a relationship with one another and interact with each other, it is essential that there is common ground between them, a form of familiarity in properties or essence that makes any interaction possible. If mind and matter are essentially different, then how can they interact? If we postulate a third entity, which makes this relationship possible, then we have complicated our ontological picture and we have *tri-ism* and not dualism. The problem is that this needs another principle and then another and so on so forth. We end up with an infinite regress problem. Moreover, we have to justify this complication of our ontological

---

<sup>12</sup> Kant Immanuel, *Critique Of Pure Reason*, translated N. Kemp-Smith, Macmillan, 1929, pp. 131-132.

<sup>13</sup> Kant, *Critique Of Pure Reason*, B 131-132.

<sup>14</sup> Kant, *Critique Of Pure Reason*, A 51/B 75.

picture by passing the test of the Ockham's razor and this does not seem to be forthcoming.

Another problem is that we violate the fundamental laws of physics by postulating a nonmaterial mind that is causally efficacious. This violates the laws of thermodynamics, which imply the causal closedness of the physical world. In physics, there is no room for non-material causal efficacy. This would contradict and contravene the principle of conservation of energy, which insists energy can neither be created nor destroyed.

### **1.3.2 Mentalism and Idealism**

Mentalism arose from the awkward interaction concept of mind and matter. The problem that perturbed Descartes' successors was that if our conscious selves reside solely in the mental domain, then how could we know the material from a mental perspective. This is an epistemic dilemma, which emerges from the ontological divide. It seems like Descartes' dualism convicts us to ignorance about the material world. Berkeley proposes a radical way out of this problem. He asserts that there is no independent substantiation for the existence of material world apart from the mental. All our experiences are just as they are, but there are no physical objects *out there* causing those experiences. Consequently, everything would continue to appear as normal, even though there would be nothing in reality except mental experiences. Mentalism, at once solves the problem of knowledge of the external world and the problem of mind-matter interaction. According to Berkeley, we know only our ideas. Hence, to be is to be perceived.

The roots of the Berkeleyian view should be looked for in the Locke's theory of perception. Although, one should reiterate that Locke was realist and empiricist. The world, according to Locke, is composed of 'insensate corpuscles' governed by the laws of mechanics. The 'insensate corpuscles' stimulate and 'move' our sense organs, which

in turn stimulate our nerves and our brains create *ideas* as a result of this interaction.

Hence, ideas are our representations of the world presented to us by the sense organs.

“All ideas come from sensation or reflection—let us then suppose the mind to be, as we say, white paper, void of all characters, without any ideas—How comes it to be furnished? ... Whence has all the materials of reason and knowledge? To this I answer, in one word, from experience. In that all our knowledge is founded; and from that it ultimately derives itself. Our observation employed either about external sensible objects, or about the internal operations of our minds perceived and reflected on by ourselves, is that which supplies our understandings with all the materials of thinking. These are the two foundations of knowledge, from whence all the ideas we have, or can naturally have, do spring.”<sup>15</sup>

These *ideas* or *ideas of sensation* differ in how accurately they represent the qualities of the substance as it is in the real world. Hence, Locke differentiates between *primary qualities* such as shape, extension, solidity, and motion, and *secondary qualities* such as sound, color, and heat. The *primary qualities* represent the actual attributes of the matter itself. While, *secondary qualities* are generated by our minds in course of their interaction with material of the sense data. Hence, they are more removed from reality and represent how things are represented in our minds and not how things in the external world. In final analysis, our *ideas of sensation* are as close we come to know the world. We know the world through our *ideas/representation* of the world.

Berkeley agrees with Locke with respect to secondary qualities. He agrees that what they represent exists only in the mind of the perceiver. Colors, sounds, tastes, etc. are representations of our experience. They do not exist independent from our minds or experience. However, Berkeley holds the same for the primary qualities. For Berkeley, Locke’s assertion that the primary qualities represent those attributes of the world that exist independent of our perception and mind is arbitrary. The shape, extension, motion, or being solid are equally are mode of experiencing these qualities the way sounds, tastes,

---

<sup>15</sup> Locke, J., (1984) *An Essay Concerning Human Understanding*, Oxford University Press, pp. 2-5.

touches, and images of the same objects are. If this is true, according to Berkeley, then what do we know about this unperceivable world? And if all we know about the world is through our perception of the world and all we know is our ideas, then what can we say about the world? How can we postulate the independent existence of this world?

“They who assert that figure motion, and the rest of the primary or original qualities do exist without the mind, in unthinking substances, do at the same time acknowledge that colours, sounds, heat, cold, and suchlike secondary qualities, do not; which they tell us are sensations, existing in the mind alone, that depend on and are occasioned by the different size by the different size, texture, and motion of the minute particle of matter.... Now, if it be certain that those original qualities are inseparably united with the other sensible qualities, and not, even in thought, capable of being abstracted from them, it plainly follows that they exist only in the mind.... For my own part, I see evidently that it is not in my power to frame an idea of a body extended and moving, but I must withal give some colour or other sensible quality, which is acknowledged to exist only in the mind. In short, extensions, figure, and motion, abstracted from all other qualities, inconceivable. Where therefore the other sensible qualities are, there must be also, to wit, in the mind and nowhere else.”<sup>16</sup>

Hence, for Berkeley, to be is to be perceived—*Esse est percipi*.

“I see this cherry, I feel it, I taste it: and I am sure nothing cannot be seen, or felt, or tasted: it is therefore real. Take away the sensations of softness, moisture, redness, tartness, and you take away the cherry. Since it is not a being distinct from sensations; a cherry, I say, is nothing but a congeries of sensible impressions, or ideas perceived by various senses: which ideas are united into one thing (or have one name given to them by the mind; because they are observed to attend one another.”<sup>17</sup>

At once, Berkeley solves the problem of mental causation and the ontological status of mind, but at what price?

---

<sup>16</sup> Berkeley, G., (1979) *A Treatise Concerning the Principle of Human Knowledge*, ed. K. Winkler, Indianapolis, Hackett, Part I, section 10.

<sup>17</sup> Berkeley, *Dialogues*, p. 81.

Berkeley's theory fails to account for another of our common sense beliefs, namely change and independent existence of changing things apart from my perception of it. As I type this paragraph, pasta is cooking in a pot on my oven in the kitchen. I know that because I prepared the ingredients, turned on the oven, and came back to type this before I eat my dinner. I fully expect the pasta to be ready in half an hour even in my absence from the kitchen. The pasta will cook regardless of my presence in the kitchen and my perception thereof. This holds everyday of our lives for a variety and multitude of things. This is precisely where Berkeley's theory fails. Having known that, he suggests that it is not my perception that ultimately matters but there is a permanent perceiver, who is the guarantor of reality, namely God. Now this is a major complication of our ontological paradigm and in no way self-evident. In fact, he has to prove the existence of God independent of his theory of mind.

Hume takes mentalism to its logical end, which appears to be *solipsism*. The worry is that if mental states are essentially private, meaning that they are available to a single individual from the first person perspective, then how can anybody ever know about the mental states of others? More importantly this question makes our interactions, communications, and discourse absurd. Wittgenstein highlights the implausibility of mentalism through the *private language argument* and metaphor of *Beetle in the box*.<sup>18</sup> Imagine that each one of us has a little box. In each box, there is something. However, only the owner of the box has access to his/her box. No one can look into another person's box and can know what in the other boxes are. Now, each person claims that he/she has a Beetle in the box. However, there is no way for anyone of us to know what others mean by claiming they have a Beetle in their box. It is logically plausible that we all have various objects we call Beetle. What does this state about the concept of Beetle at all? It implies that the concept of Beetle is inane, because we cannot know what others mean by Beetle, or what does it really refer to. Wittgenstein maintains that, in this case,

---

<sup>18</sup> Wittgenstein L., (1953) *Philosophical Investigation*, translated by G. E. M Anscombe, Oxford Blackwell, §293.

the idea of Beetle *cancels out* and it becomes meaningless. It can play no role in meaningful discourse.

Wittgenstein asserts further: “An ‘inner process’ stands in need of outward criteria” (section 580). In other words, in order to be justified in ascribing a “mental state” to some entity, there must be some true claims about the observable behavior of that entity that entail that the entity has the mental state in question. If no true claims about the observable behavior of the entity can play any role in the justification of the ascription of the mental state in question to the entity, then there are no grounds for attributing that kind of mental state to the entity.

He contends further that public verification is essential to the function of language. In other words, there are no meanings and sense in a language the claims of which can be verified by only one person. As a result, conversation about mental states cannot possibly utterly relate to private states. If mental states have any objective content, then one must deem the mental realm as inherently connected to the behavior, which makes it publicly observable and verifiable.

### **1.3.3 Material Monism**

Mentalism’s end begins with Hume’s critic and the emergence of eliminative and reductive materialism. Materialism, of any variety, rejects the distinction between subjective mind and objective brain.

The traditional materialist view in the modern times is formulated Thomas Hobbes. According to Hobbes, humans are just machines: ‘For what is a heart but a spring; and nerves but so many strings, and the joints but so many wheels, giving motion to the whole body?’<sup>19</sup> For Hobbes, sensory experience is nothing other than ‘motion in

---

<sup>19</sup> Hobbes, T., (1997 (1651)), *Leviathan*, edited by R., Tuck, Cambridge, Cambridge University Press, p. 9.

brain' generated by moving matter in the external world. Experience has no inherent characteristic, 'for motion produceth nothing but motion'<sup>20</sup>

“For seeing life is but a motion of limbs, the beginning whereof is in some principal part within; why may not say, that all *automata* (engines that move themselves by springs and wheels as doth a watch) have an artificial life? For what is the heart but a *spring*; and the *nerves* but so many *strings*, and the *joints* but so many *wheels*, giving motion to the whole body such as was intended by the artifice?”<sup>21</sup>

The intuition that there is a causal relationship between mind and brain is very deeply rooted and historically very old, as old as human civilization itself. It is based on our daily experiences and observations. Hippocrates, the father of Western medicine states:

Man ought to know that from brain and from brain only, arise our pleasures, joys, laughter and jests, as well as our sorrows, pains, grieves and fears. Though it, in particular, we think, see, hear, and distinguish the ugly from the beautiful, the bad from the good, the pleasant from the unpleasant, in some cases using custom as a test, in others perceiving them from their utility. It is the same thing which makes us mad or delirious, inspires us with dread and fear, whether by night or by day, brings sleeplessness, inopportune mistakes, aimless anxieties, absent-mindedness, and acts that are contrary to habit.<sup>22</sup>

The materialist position can be categorized in three distinct camps: the radical/eliminative, the reductive, and the emergent. The radical view maintains that consciousness does not refer to anything real at all. It is a misconception of our non-scientific and mundane theorizing about the nature of the world and ourselves. The reductive materialism does accept the reality of consciousness and mental states, it assert

---

<sup>20</sup> *Ibid.*, p.14.

<sup>21</sup> Hobbes, T., (1970) *Hobbes Selection*, editor Woolbridge F. J. E., New York, Scribner.p. 136.

<sup>22</sup> Flew, A., (1978), p.32.

that mental states are nothing above and beyond brain states. They are another way of description of brain states and functions. Emergent materialist view maintains that there is only material stuff, but they can have two types of properties: material and mental. The mental properties emerge from, supervene on, or are a function of the material substrate, but it is not reducible to it. This can be also called non-reductive materialism or property dualism.

### **1.3.3.1 Reductive Materialism**

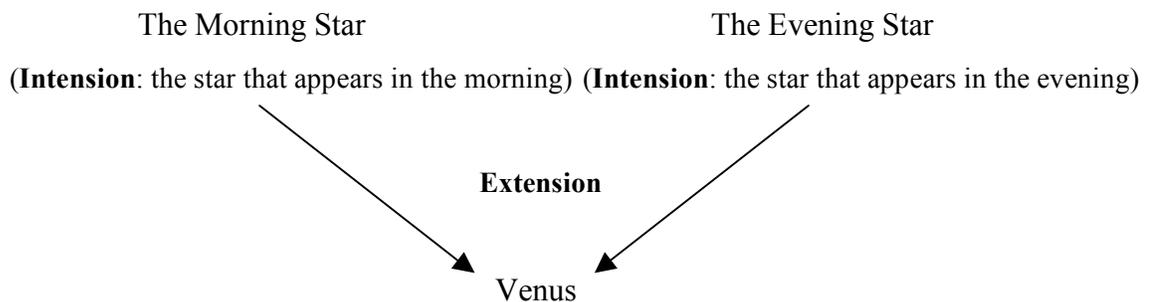
Reductive materialism does not refuse the truth of mental states, however, it states that mental states are identical to material states. Mental talk refers to the same material states, but it uses another descriptive set and language. This entails that brain states can be expressed either in physical terms of natural sciences, or in mental terms. To say that one is in pain refers to the same brain state as saying one's C-fibers are firing. Reductive materialists, such as J.J.C.Smart, accept that the apparent difference between phenomenal consciousness and brain states. Furthermore, they do admit that the description of mental states and the description of brain states do not have the same sense or meaning. However, they assert that the advances in the neurosciences will prove, and have proved, that they refer to the same thing and processes. Therefore, the mental states and events are identical to brain states and events. This is, however, a contingent identity and not a logical identity, which is an essential point of this view. Smart states:

Let us first try to state more accurately the thesis the sensations are brain-processes. It is not the thesis that for example, 'after image' or 'ache' means the same as 'brain-process of sort X' (where 'X' is replaced by a description of a certain brain process). It is that, in so far as 'after image' or 'ache' is a report of a process, it is a report of a process that happens to be a brain process. It follows that the thesis does not claim that sensation statements can be translated into statements about brain processes. Nor does it claim that the logic of a sensation statement is the same as that of a brain process statement. All it

claims is that in so far as a sensation statement is a report of something, that something is a brain process. Sensations are nothing over and above brain process.<sup>23</sup>

Here, Smart tries to point to an important distinction between the essence of phenomenon, its appearance, and its description. This is a kind of identity theory of reductive materialism, namely type-type identity theory.

At the core of identity theory is a linguistic analysis, which can be elucidated through an example. At the heart of this distinction is the difference between sense/intension and reference/extension of a concept. The sense of a concept is primarily its definition, its meaning. The reference of a concept, on the hand, is the set of all the objects in the world that can be subsumed under that definition. For instance, the concept of 'bachelor' is defined as 'an unmarried man'. This is the sense, or the intension, of the concept 'bachelor'. The reference, or the extension, of the concept is all the men in the world that fall in this set. Hence, extension looks for the correspondence of an intension in the real world. For identity theorists, this distinction lies at the heart of the mind/brain controversy.



It is possible to discuss the morning star and the evening star separately and detail facts about them independently, while failing to realize that they have the same referent,

---

<sup>23</sup> Smart, J.J.C., (1962), Sensation and brain processes, in V.C. Chappell (ed) *Philosophy of Mind*, Englewood Cliffs: Prentice Hall, p.163.

namely Venus. As long as this fact is unknown, one could spend decades, or centuries, differentiating the two, while all along we have been talking about the same phenomenon from two different perspectives. This is exactly the case with mind/body discourse as well. Talking about brain states and mental events is talking about the same thing using different vocabularies and descriptive terms, since they refer to the same thing.

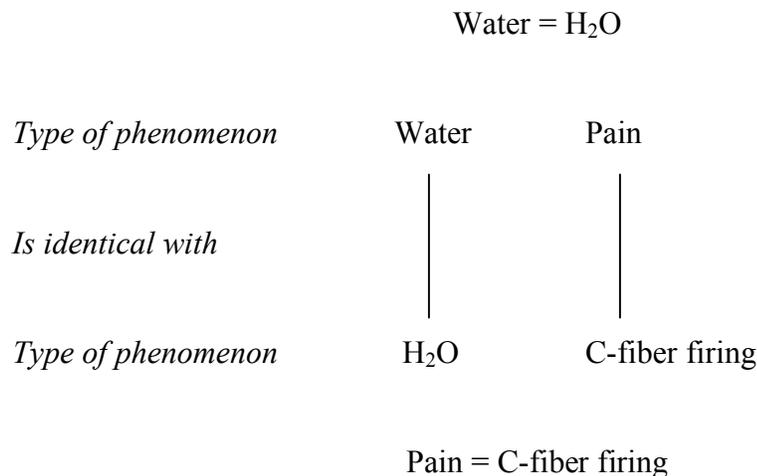
I believe that this analogy, at best, proves a double-aspect theory and never an identity theory. In the case of ‘the morning star’, ‘the evening star’ and ‘Venus’, we have three entities. There are two intensions, which refer to a single extension. The case is clear. In the case of ‘mental states’ and ‘brain states’, there are only two entities. It is clear that ‘mental states’ play the role of an intension. However, how about ‘brain states’? They seem to be simultaneously the intension and the extension to themselves. This does not make much sense. We have to choose for one option. Either ‘brain states’ are an intension, or they are the extension. If they are the intension, to what do they refer to? And if they are the extension, then what is the intension that refers to them along with ‘mental states’? Hence, the identity theory seems to claim a straightforward identity and the whole sense/reference seems to be irrelevant.

Another way to induce the same intuition, by the reductionist, is through the example of temperature, which vividly illustrates the reductive materialist point. Physicists describe temperature in terms of mean kinetic energy. This, however, does not imply the abolition of temperature vocabulary from our ontology. Neither, does it entail addition of temperature as an additional article to our ontology. The notion is that temperature is nothing above and beyond mean kinetic energy. The same holds for consciousness and brain. Conscious states do exist, but they are nothing above and beyond brain states. Mental vocabulary is just another language for description of brain states. In the same manner that has been shown that temperature is nothing other than mean kinetic energy of molecules in motion, so we shall consent to the finding that conscious states are nothing other than brain states.

Accordingly, the mind is not a being separate and distinct from the brain, but the mind is identical with the brain and mental events/states are identical with brain events/states. This is the identity theory. It exemplifies material monism, since it states that there are only material substances and their states. There are two main formulations of the reductive material monism: the type-type identity theory and the token-token identity theory.

*A. The Type-Type Identity theory:*

The type-type identity theory asserts that each type of mental state is identical with a given type of brain state. An example would illustrate this point; water is always identical with H<sub>2</sub>O regardless whether it is water in the pond, in the pool, in the river, in the ocean, in the rain, drinking water, etc.



This implies that pain is always identical with C-fiber firing. The reduction of a mental state to a brain state is also based on this identity relationship. This is, however, an *ontological reduction* and not an *analytic reduction*. The rejection of analytic reductionism implies that the intension of mental and physical are different, but their extension coincides. In formal terms:

*If a mental state M is identical with a brain state B, then M can obtain only if B obtains, and B can obtain only if M obtains.*

OR

*M obtains if, and only if, B obtains.*

This is exactly what we should expect, if we assert that M and B are identical. This is ontological reduction. Accordingly, one group of phenomena that are seemingly numerically distinct from another group of phenomena is just one set of existents and not two. The ‘if and only if’ statement is a biconditional proposition. This implies that there are *bridge laws*, which connect given types of mental states with given types of physical states. This implies, for instance, that firing of C-fibers is logically necessary and sufficient for ‘pain’ to occur. A further implication of identity of mental events with physical events is that mental properties are also identical with physical properties describable by means of two different vocabularies. Meanwhile, mental concepts and physical concepts remain nonsynonymous.

The problems with type-type theory are manifold. First, it cannot give an explanation for the *multiple realizability* of mental states. In other words, it cannot account for the various manifestations of the same mental states by different material substances. Second, it cannot explain intentionality. It cannot explain how electrochemical activity of neural networks can give rise to semantics. The fundamental question is how can a system that process information syntactically and symbolically give rise to semantics and stand for something as it is in intentionality. Third, it cannot account for phenomenology of mental states. It cannot explain how subjectivity emerges from objective states. To be fair, however, this is where most theories fail. Therefore, Chalmers considers this issue the hard problem of consciousness.

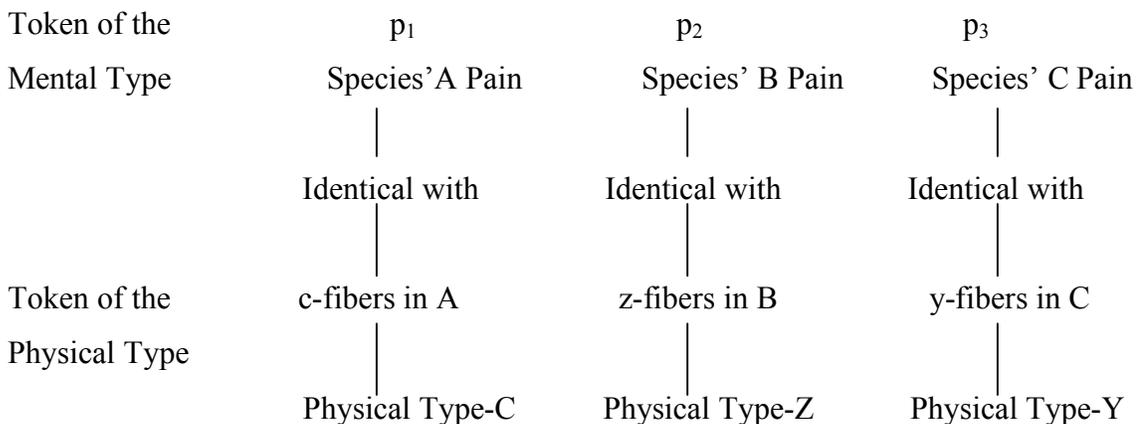
#### *B. The Token-Token Identity Theory*

As we discussed above, the type-type theory insists that each *type* of mental state must be identical with a corresponding *type* of a physical state and there are no exceptions allowed. Hence, the conditions for the satisfaction of this theory are very strict and accordingly quite inflexible. This inflexibility is precisely the problem with the type-

type theory, since we can conceive that a certain mental state in different individuals or species can be manifested by a different physical state and configurations. In other words, mental states can be *multiply realized*. Our intuition and experience tells us that mental states are embodied in all different sorts of material arrangements. The token-token theory is a response to this inflexibility of the type-type theory and it is fueled by the same intuition and informed by the same experience mentioned above.

To understand the token-token theory, we must first clarify what a token is. A token is a certain distinct instance of a type. Tokens are members of the set and the set is the type. For example, ‘chair’ is a type, but ‘the chair’ in the living room is a token of that type and the chair in the kitchen is another token etc. Hence, according to token-token theory, every token of a type of mental state could be identical with a token of a type of physical state, but it is not necessary to be the case that tokens of the same type of physical state must be involved on every case. This means that each mental state is identical with a physical state, but this physical state does not have to be the same in every case we encounter. Hence, mental state can be multiply realizable, manifested by different physical states. This theory is supported by clinical and empirical evidence of plasticity of neural structures. Different parts of the brain can assume different roles once the original structure assigned for that function has been damaged. One sees that over and over again in case of stroke patients.

Mental Type Pain (P)



Now, this does not mean that c-fiber firing is taken arbitrary. This means, however, that c-fiber firing is taken to be a contingent fact and there is no logical necessity that pain should be necessarily be limited to the activity of one type of fibers and networks. Other type of fibers can perform the same function as well.

One of the main motivations to reduce one set of phenomenon to another is to achieve, and establish, a systematic relationship between them. In this case, can the token theory establish a systematic relationship between mental states and physical states? In other words, are mental states reducible to physical states within the token-theory scheme? If this systematic relationship is unobtainable, then the coexistence of mental and physical states has to be taken to be a brute fact, which is not *a priori* acceptable fact. Assuming a phenomenon, as a brute fact should be embraced *a posteriori* after all other options have been exhausted. It should not be the first option within an explanatory scheme; otherwise all phenomenon can be explained so.

As we stated before, reduction of set of phenomena to another set of phenomena requires bridge laws that can be expressed in terms of biconditional statements. Accordingly, for a mental state M to be reducible to, and identifiable with, a physical state P, it must be the case that M occurs if, and only if, P occurs. The question is whether multiple realizability breaches this requirement. According to token theory, M can be principally manifested, and realized, in an indefinite variety of physical states. Hence, a single physical state P is not necessary for M to occur. M can equally be realized singly by P<sub>1</sub>, P<sub>2</sub>, P<sub>3</sub>, and P<sub>n</sub>. Each one these physical states are sufficient for M to occur, but none of them is necessary. This necessity is requirement of the bridge laws though. The set of them all does not constitute a necessary condition either, since this set is indefinite in number by definition. It is an open set. Only a closed set can function as a necessary condition. Hence, token theory seems to have to give up reducibility of mental facts to physical facts. This is precisely what happens with assumption of the principle of supervenience and non-reductive physicalism and property dualism, as we will see later.

The question, which still remains, is whether such theories can account for qualia and intentionality.

### 1.3.3.2 Eliminative Materialism

The eliminative camp is a diverse one. We could distinguish between three major core assertions, which shape the corresponding eliminative paradigm:

1. There is no consciousness. This is a categorical ontological statement, which denies the existence of this phenomenon altogether. The belief in the existence of consciousness, accordingly, is akin to religious faith in God in absence of any logical proofs. Rey asserts:

Why in the world should one believe in such as God? Why should one believe in such a consciousness? In both cases, of course, people have been tempted to say, 'Because, I have direct access to it.' But such a first-person breast beating begs the question...the challenge...is to come up with some non-question-begging reason to believe consciousness exists. I doubt there is any to be had.<sup>24</sup>

2. The term consciousness does not refer to anything sufficiently clear for it to be scientifically and analytically useful.

Sloman states:

People who discuss consciousness delude themselves in thinking that they know what they are talking about...it is not just one thing but many things muddled together...like multifarious our uses of energy.<sup>25</sup>

Stanovich agrees with Sloman and reasserts that 'the term consciousness fractionates into a dozen or more different usages'. Hence, it collapses into 'a botched concept'.<sup>26</sup>

---

<sup>24</sup> Rey, G., (1991), Reason for doubting the existence of even epiphenomenal consciousness, *Behavioral and Brain Science* 14(4): p.692.

<sup>25</sup> Sloman, A., (1991), Developing concepts of consciousness, *Behavioral and Brain Science* 14(4); 694-695.

<sup>26</sup> Stanovich, K.E., (1991), Damn! There goes that ghost again, *Behavioral and Brain Science* 14(4); 696-697.

3. Consciousness is an invention of *folk psychology*, which will be dispelled with scientific progress. Eventually, subjectivity and, in fact, all mental jargon will be discarded in favor of precise scientific formulations.

The view that denies the existence of consciousness altogether seems highly questionable, given that it is not capable of explaining the reason that mental states act *as if* they have subjective qualities. How can the *deception* of subjectivity emerge from physical and objective structures? Moreover, Descartes showed us that we could doubt the existence of almost everything except the doubting of a doubting mind. This doubting is also the product of our phenomenal consciousness. The denial of existence of consciousness is to deny everything we experience from doubt, to hate, to love, to sadness, to happiness, to hope, to frustration, etc. This denying all that makes up our daily existential experience. The question is how the proponents of this view would explain our experience. If they deny it, then they are contradicting their very act of questioning the nature of consciousness. And if they have another phenomenon that needs to be explained in place of consciousness, then they are just trading labels and have not achieved anything in reality.

The proponents of the view that consciousness is a vague concept and hence cannot be used in a serious scientific and philosophical discourse and investigation have right to the extent that they demand proper definitions for the sake of clarity. The question is, however, do we have clean and clear-cut definitions to work with especially with respect to topics that scientifically and philosophically matter and most and even for mundane object of everyday usage. Let us examine two cases to illustrate the point. Firstly, let us take the term 'energy' in physics and not everyday use of it. Do physicist have a clear definition of what energy *is*. The answer is: no. They tell us what energy does and how it stands in relationship to other physical parameters. Equations in physics are not definitions, but they are expressions of relationships. However, although the nature of energy is still a mystery, it has not stopped physics from properly investigating energy. The same holds for force and many other physical phenomena.

Secondly, let us take a mundane object like a 'chair'. We all know what a chair is and recognize one, when we see one. However, what is the definition of a chair. Is it something you can sit on? Here, we give a functional definition. Well, if a chair is a functional thing: we sit on it; then how about a toy chair or a museum piece that cannot be used for sitting purposes. However, if you define a chair structurally: A chair one or many legs and surface to sit on, then how about a bench carved in a rock or simply a rock. This does not mean there are no chairs or the concept of chair is nonsensical and it should be thrown out of our discourse. It means that phenomenon cannot be defined the way we want them most of the time. The desire to have precise definition for everything is more a psychological desire than a logical requirement. A definition should be stable at its core so we can differentiate a chair from a desk. However, at the edges definitions are blurry and that is good so because definitions should be relevant in a dynamic world of change. Too rigid definitions are useless. So we could never give a strict and final definition of consciousness. Reality does not allow that.

The third pathway of elimination of consciousness is to ascribe qualities of our common-sense theories and everyday explanatory schemes, namely *folk psychology*. Here, we explain our behavior in terms of our conscious thoughts, feelings, desires etc. The proponents of this branch of eliminativism claim that these theories are a direct result of our lack sufficient knowledge of the biology of the brain. These are provisional, pre-scientific, and to some degree arbitrary way of explaining phenomenon that deeply interest us, but our science has not reached the level to sufficiently explain them for us. Accordingly, there are other examples of such phenomenon in the history of science. Phlogiston was once postulated to explain the nature of combustion or *élan vital* was assumed to explain the nature of life itself. However, progress in science dispelled these notions as wrong and more importantly replaced them with proper scientific language. The same holds consciousness and mental jargon. In future, folk psychological explanations and terminology will be replaced with neuroscientific theories and language. Churchland puts this replacement in terms of a theoretical reduction. She states:

In the sense of ‘reduction’ that is relevant here, reduction is first and foremost a relation between theories. Most simply, one theory, the reduced theory  $T_R$ , stands in a certain relation to another basic theory  $T_B$ . Statements that phenomenon  $P_R$  reduces to another phenomenon  $P_B$  are derivative upon the more basic claim that the theory that characterizes the first reduces to the theory that characterizes the second.<sup>27</sup>

There are three main objections to this view.

1. Even if it so happens that folk psychology is a defective theory, it is not sufficient to presume that emotions, cognitions, and volitions are non-existent. All it proves that folk psychology cannot explain mental events, processes, and states adequately.

2. The other difficulty is the confidence on futuristic explanatory powers of neuroscience. Like any other science, Neuroscience is occupied with structures, dynamics, and functions. It is not quite apparent how the present methodology can explain intentionality, and phenomenology. The point is that maybe neuroscience will accomplish what Churchland proposes, or maybe it will not. It is not quite certain how Churchland can foresee facts about future, based on inductive argumentation.

3. Eliminativist’s argument based on the history of science is based on a false analogy. Churchland compares the case of mental entities to that of ether, phlogiston, etc. Yet, these two cases are not epistemologically similar. In the case of phlogiston, there exists an epistemically symmetrical situation. All the individuals, including scientists, have potentially equal epistemological access to investigation of these phenomena. As a result, once sufficient data has been collected, the truth about such phenomenon can be determined. However, that is not the case about consciousness. It is essentially epistemically asymmetrical. The owner of experience is the only partaker, who has direct

---

<sup>27</sup> Churchland, P., (1988) *Neurophilosophy: Toward a Unified Science of the Mind-Brain*, MIT, Cambridge Mass, p.270.

access to the phenomenological content. The third-person perspective is necessarily dependent on the first-person perspective. There seems to be a qualitative difference between the two perspectives, which seems to be intractable.

### **1.3.3.3 Non-reductive Physicalism and Property Dualism**

As results of the failures of reductive materialism, modern theories of physicalism assume a form of non-reductive materialism that either presuppose the duality of properties or they subscribe to functionalism.

Functionalists consider mental states in terms of their typical causes and effects. Mental events are causal mediators, emerging from perceptual stimuli and affecting behavior through their interactions with other mental states. For example, pain is a state that emerges from bodily injury and usually causes a want to evade the source of that harm. Any ensuing behavior depends on the interaction of this desire with other beliefs and desires. As a result, functionalism allows for the reality of mental states, even though they do not reveal themselves openly in observable action. Functionalism, however, does not commit itself on what mental states are composed of. Ontologically, it is compatible with all the theories discussed here. However, most of the functionalists are non-reductive physicalists. I will deal with functionalism extensively in the following chapter.

Property dualism does not consider conscious minds as constituted by separate stuff wholly different from the material body, but it claims that there is only one type of substance—matter—which can enjoy two distinct types of properties. The affiliation between brain and mind is framed by the *principle of supervenience*. This principle claims that mental properties *supervene* on the brain states. Supervenience can be understood in terms of entailment relationship. As a result, material properties of a certain kind and complexity entail mental properties, which *depend on* and *co-vary* with brain states, but not vice versa. However, supervenience does not imply reduction. In other words, mental states are not *reducible* to brain states. Davidson, who formulated this principle first in modern philosophy, states:

Although the position I describe denies there are psych-physical laws, it is consistent with the view that mental characteristics are in some sense dependent, or supervenient, on physical characteristics. Such supervenience might be taken to mean that there cannot be two events alike in all physical respects but differing in some mental respects, or that an object cannot alter in some mental respect without altering in some physical respects.<sup>28</sup>

Supervenience is a relationship between two sets of facts. According to this concept, one group of facts can fully determine another set of facts. The materialist position states that the higher-level facts (supervenient-facts), or mental facts, supervene on the lower-level facts (subvenient-facts), or the physical facts. Therefore, the supervenience principle declares that: higher-level facts supervene on the lower-level facts, if no two possible situations are identical with respect to their lower-level properties while differing in their higher-level properties. For example, the facts about biology supervene on the facts about physics insofar as two possible worlds that are physically indiscernible, there are also biologically indiscernible.

Supervenience can be applied *locally* and *globally*. *Local* supervenience concerns individuals and it can be defined as: the lower-property of an individual entails the higher-level properties of that individual. Local supervenience is concerned with differences *within* worlds. This is exactly the limitation of local supervenience. It does not give us much information about the truth-value of a statement. Let us remind ourselves, truth of a proposition is determined by applying the standards of truth to a statement *between* the possible worlds. Global supervenience, in contrast, is concerned with facts between the worlds. This fact makes global supervenience instrumental in determining truth-value. Global supervenience states that any two possible worlds that are physical duplicates of each other are also psychological duplicates.

---

<sup>28</sup> Davidson, D., (1994) *Mental Events*, reprinted in Davidson, *Essay on Action and Events*, Oxford, Oxford University Press, p. 214.

Another important distinction is between *logical* and *natural* (also called *nomic*) supervenience.<sup>29</sup> Higher-level facts supervene on the lower-level facts if no two logically possible situations are identical with respect to their lower-level facts but distinct with respect to their higher-level facts. The logical possibility corresponds to the conceivability of a situation. Therefore, logical supervenience needs not to be constrained by the natural laws. It is solely constrained by the criterion of rationality, the principle of non-contradiction. It is logical to conceive of a world, in which dogs can fly. However, it is logically impossible to conceive of square circles, or married bachelors. Logical supervenience is defined in terms of logically possible worlds and not deducibility in any system of formal logic.<sup>30</sup> Biological facts supervene on the physical facts. This means that once the physical facts are established in their entirety, the biological facts are also established in their entirety. Natural, or nomic, supervenience is the narrower type of supervenience. Natural supervenience establishes a structural and functional relationship between two sets of facts in the natural world. This relationship obeys the governing laws of the natural world. High-level properties naturally supervene on the physical properties if any two naturally possible situations with the same physical properties have the same high-level properties. In other words, higher facts supervene nominally on lower facts if, and only if, any world (with our laws) that is lower facts identical is also higher facts identical. Therefore, it is naturally impossible to conceive of a world, which has no gravity. However, it is logically possible to conceive of such a world. Natural possibilities are much more stringent than logical possibilities. Natural possibility has to obey natural laws.

Materialism claims either that all facts in the universe are *logically* and *globally* supervenient on the physical, or all the facts are *nomologically* and *globally* supervenient on the subvenient physical facts.

---

<sup>29</sup> Braddon-Mitchell, *Philosophy of Mind and Cognition*.

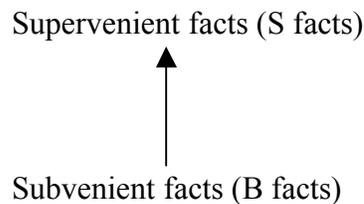
<sup>30</sup> Chalmers, D., (1996) *The Conscious Mind: In Search of a Fundamental Theory*, Oxford Univ. Press, pp. 32-42.

We can summarize relationships between the supervenient and subvenient in terms of three key properties:

- 1) *Irreducibility*: supervenient facts are neither analytically (via definition), nor ontologically reducible to subvenient facts.
- 2) *Co-variation*: this implies that there can be changes in the supervenient properties, if and only if, there are relating changes in the subvenient phenomena. However, a change in the subvenient phenomena does not necessarily entail a change in the supervenient property. This implies that phenomena cannot differ in their supervenient qualities if there is no distinction in their subvenient characteristics. Two entities that are indiscernible in their subvenient features must be indiscernible in their supervenient qualities. This accounts for *multiple realizability*.
- 3) *Dependence*: this implies that supervenient qualities are dependent for their existence on subvenient phenomena. This is, however, an asymmetric relationship; in that, subvenient phenomena exist independently of supervenient phenomena.

Hence, we can state:

S facts are supervenient on B facts, if S depends for its existence on B; S can change but if, and only if, B changes; but B can change without it necessarily being the case that S changes.



The question is whether consciousness is logically or nomologically supervenient on the brain subvenient states. In case of logical supervenience, physicalism is proved

wrong, if one can conceive of a logical situation where the brain would evolve to the same extent without consciousness. In other words, is it logically conceivable to imagine a world, which physically identical to our world but it lacks consciousness? It means there is world physical identical to this one with an individual physically identical me, but without consciousness, a zombie of me. Are zombies logically possible? Chalmers asserts that they are conceivable.<sup>31</sup> Hence, consciousness is not logically supervenient on the physical. The idea of zombie world is not logically impossible or inconceivable. In his book, *Consciousness Reconsidered*, Owen Flanagan admits that:

Consciousness did not have to evolve. It is conceivable that evolutionary processes could have worked to build creatures as efficient and intelligent as we are, even more efficient and intelligent, without those creatures being subjects of experience. Consciousness is not essential to highly evolved intelligent life. This claim is true and important.<sup>32</sup>

The question with regards to nomological supervenience can be answered either way. The question is whether the laws of nature would have required the emergence of consciousness from an evolving brain. Referring back to Flanagan, we can claim evolution could have happened without emergence of consciousness. In this case, consciousness does not nomologically supervene on the brain. This would mean the rejection of the supervenience theory of mind. However, consciousness has emerged along with an evolving brain. As Eccles argued above, this implies that consciousness had some survival value and essential function. Otherwise, it would not have been present and would have gone extinct or not emerge at all. So, for nomological supervenience to have any plausibility, the question of mental causation has to be decided. There are again two possible answers to this problem. Firstly, consciousness is causally efficacious. Secondly, consciousness is causally impotent.

Let us explore the first option: the causal efficacy of consciousness.

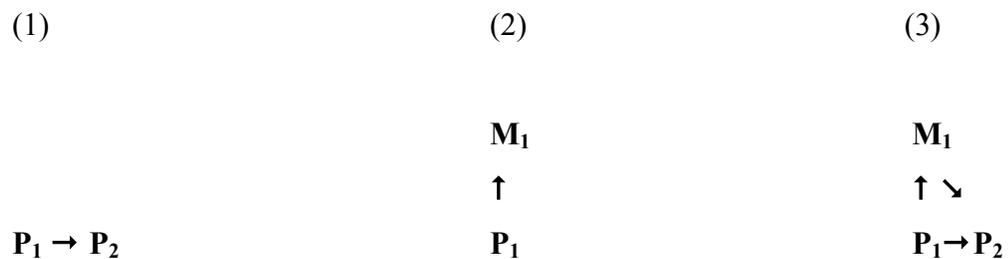
---

<sup>31</sup> Chalmers, D., (1988), *The Conscious Mind*.

<sup>32</sup> Flanagan, O., (1992) *Consciousness Reconsidered*, MIT press, Cambridge Mass, p. 60.

Now let us consider the following case:

- 1)  $P_1$  (C-fiber firing in a particular neural pathway) causes  $P_2$  (twitching of the left eye).
- 2)  $M_1$  (pain) supervenes on  $P_1$  (C-fiber firing).
- 3)  $M_1$  (pain) causes  $P_2$  (twitching of the left eye).



Now,  $P_1$  should be sufficient for the occurrence of  $P_2$ , if the theory of the causal closedness of the physical world is true. However, if one denies that  $M_1$  causes  $P_2$ , then one advocates epiphenomenalism, causal impotence of consciousness. The acceptance of (2) means that either  $P_1$  is not sufficient for occurrence of  $M_1$ , which contradicts the principle of closedness; or it means that both  $M_1$  and  $P_1$  are sufficient for occurrence of  $P_2$ , which means that physical events are *overdetermined*.

Now, the question whether overdetermination is something that physicalism can live with. Overdetermination is the notion that there can be two, or more, distinct, and individually sufficient, causes for any physical effect. The answer whether should overdetermination is not *a priori* clear, but it seems to violate the laws of conservation of energy and the principle of casual completeness of the physical world. This view has the difficulty of elucidating how mind can affect matter without violating the principles of physics, since the physical world is causally closed. Accordingly, the causes of physical events and states are always other physical events or states. Furthermore, if we delineate the causes of physical effects, we never must leave the realm of the material. This

appears to allow no room for non-physical properties, such as the conscious properties of experience. The consequence of the causal closedness of the material world is the causal powerlessness of the mental and conscious factors.

Now let us address the case of *epiphenomenalism*, causal impotence of consciousness. An epiphenomenon is a causally inefficacious by-product of some process. According to epiphenomenalism, although mental events are caused by physical events, they are only epiphenomenon. Consequently, mental events are events incapable of causing any further events. Thomas Huxley presented the classical formulation of epiphenomenalism:

All states of consciousness in us, as in brutes, are immediately caused by molecular changes of the brain-substance. It seems to me that in men, as in brutes, there is no proof that any state of consciousness is the cause of change in the motion of matter of the organism. If these positions are well based, it follows that our mental conditions are simply the symbols of consciousness of the changes which take place automatically in the organism; and that, to take an extreme illustration, the feeling we call volition is not the cause of a voluntary act, but the symbol of that state of the brain which is the immediate cause of that act. We are conscious automata...<sup>33</sup>

This entails that despite our common sense belief our hopes, desires, and mental states affect are causally impotent and their effect is an illusion. This viewpoint *upward causation* from brain to mind, but it denies the *downward causation* from mind to brain. Epiphenomenalism respects the causal closedness of the world, while allowing the reality of a causally impotent consciousness. Accordingly, the consciousness is an epiphenomenon caused by the brain, but it has no ability to affect the brain.

Epiphenomenalism is an implausible view for various reasons. Firstly, it goes against our experience that our desires, thoughts, and feelings move us to act in a certain way. The proof is not on us to justify our experience, but the proponents of this view to

---

<sup>33</sup> Huxley, T. H., (1893) *Methods and Results*, Appleton-Century-Crofts, New York, p. 244.

prove that our experience is illusion and how matter can generate the 'as if' effect. Secondly, epiphenomenalism does not fit the materialist ontology. The materialist ontology is concerned with phenomenon, their qualities, their structure, and their function. The laws of nature explain the relationships of these phenomena. However, epiphenomenalism introduces a new entity in this ontology, the epiphenomenon. Moreover, there is no other case where such postulation has been made in the physicalist ontology. This seems arbitrary.

## **Chapter Two**

### **Functionalism and Its Consequences**

Functionalism envisages mind as a function. A function of a thing is what it does. This also implies that there is an inherent dependence on the structures and arrangements, which make function possible. Functions are always embodied. Function and form are distinct, but they are inherently interlinked. This dependency of function on form should not be confused with a reduction of function to the form. Although a function always depends on a structure, or a form, to be realized; many structures and forms can simultaneously realize it. In other words, a function is inherently multiply realizable. This means that the same ‘software’ can be executed by many different ‘hardware’ arrangements.

#### **2.1 Methodological and Analytic Behaviorism**

Theories do not arise in vacuum. They have a historical root, which constantly informs them. Therefore, it is of great importance to expose the roots of theory in order to understand its progress, development, evolution, limitations, basic tenets, and future directions. Same holds for functionalism. It is rooted in philosophical positivism and psychological behaviorism and the desire to study psychology empirically and evaluation of verifiable assertions. Watson states:

Psychology as a behaviorist views it is a purely objective experimental branch of natural science. Its theoretical goal is the prediction and control of behavior. Introspection forms no essential part of its method nor is the scientific value of its data dependent upon the readiness, which they lend themselves to interpretation in terms of consciousness.<sup>34</sup>

Wundt had previously tried to marry subjectivism/introspection and objectivism/experimentation. For Wundt, consciousness and mind were identical and the

---

<sup>34</sup> Watson, J.B., (1913), Psychology as the behaviorist views it, *Psychological Reviews* 20: p.158.

main task of psychology was precisely to study this phenomenon. However, his methodology was based on presenting a stimulus to trained subjects, who carefully introspect about the stimulus and proceed to give a detailed report of their minute-to-minute experience. Hence, Wundt's methodology was based on two basic principles, namely *introspection* and *reportability*. However, the problem was that there was no way to give a clear and definite list of what constitutes contents of consciousness in order to differentiate one phenomenon from another. This problem was partly also exacerbated by the different levels and methods of training the subjects. There was no uniform methodology and this seemed to be a fundamental flaw of the methodology and the paradigm itself. Moreover, this methodology would leave out other animals. Hence, it was impossible to study mind and its evolution across many species.

According to the behaviorist perspective, it is useless to speculate about the nature of mind, consciousness, and mental states. Psychology should be about tangible and quantifiable parameters such as behaviors, the stimulus that generate them, and quantifiable, measurable, and verifiable physiological functions.

The main assumption of behaviorism is that consciousness is reducible to behavior. If that is the case, then we don't lose anything by restricting our investigation to the study of stimulus and response, since behavior can be sufficiently explained by studying the stimulus and response. Here, behaviorism makes a methodological prescription as well as an ontological assertion. The methodological recommendation is comprised in limiting our investigation to stimulus, response, and all quantifiable, measurable, and verifiable parameters. The ontological contention claims that consciousness/mind is reducible to behavior. According to ontological behaviorism, there exist no mental facts above and beyond behavioral facts. Moreover, there exist no mental states above and beyond actual or possible behavior. This claim denies any causal role to mental states. Accordingly, there are only observable behaviors or dispositions to behave in a certain way.

Methodological behaviorism asserts a weaker version of ontological behaviorism, mainly, for the sake of setting the constraint of the behaviorist project in psychology. Accordingly, there is no requirement to hypothesize private, inner, and subjective events, states, and processes, for they are fundamentally subjective. Purely subjective concepts can play no role in communication, because discourse is intersubjective by definition. Consequently, subjective concepts cannot be scientifically studied. Mental concepts can have sense, or intension; but they lack reference, or extension. For instance, the states of being in pain could mean some inner state, but it does not refer anything that actually exists.

The emphasis of methodological behaviorism is the application of behaviorist philosophy to experimental psychology. The basic tenets of methodological behaviorism are: Firstly, behavioral data comprise the only dependable and acceptable kind of information in psychological research. Secondly, psychological theories may not appeal to, or refer to, internal, subjective, and private mental states in their explanatory system.

According to this theory, the disparity in the behavior of individuals can be explained by their history of reinforcements and punishments. Here, it seems indefinite what the foundation of the postulated history can be. It sounds like some kind of complex or simple memory. But memories are mental states.

The analytic behaviorism commits to both ontological as well analytic reductionism. This is in contrast to identity theories, as we saw earlier, which upheld only ontological reductionism. According to analytic behaviorism, propositions describing mental events/states can be translated sufficiently into statements expressing physiological states/events.

The claim of ontological behaviorism is counterintuitive. Our intuition about our behavior is not based on some enigmatic and mystical premonition. It is based on the fact of our experience. Moreover, we do not come to realize our experiences through an objective third-person perspective and observation of our behavior. We have a subjective first-person experience of them. Chappell asserts:

If behavior were true, I could find out that I myself had a pain by observing my behavior, but since I do not find out that I have a pain, when I do, by observing my behavior...behaviorism is not true.<sup>35</sup>

Moreover, we cannot automatically determine, explain, and much less predict the behavior of others by observing their behavior. It becomes an impossible task when others try to conceal their true mental states by acting the opposite. Chappell states further:

If behaviorism were true I could always in principle find out when you had a pain by observing your behavior, but since I cannot always find out, even, in principle, that you have a pain when you do, whereas I can always observe your behavior it follows that behaviorism is not true.<sup>36</sup>

Chomsky puts the greatest, and most, devastating criticism of behaviorism forth in his analysis of language. He suggests that problems of explaining language in behaviorist terms are insuperable. People's verbal behavior is almost impossible to predict by what they are told, the verbal stimulus.

A typical example of 'stimulus control' for Skinner would be the response to a piece of music with the utterance Mozart or to a painting with the response Dutch. These responses are asserted to be 'under the control of extremely subtle properties' of the physical object or event. Suppose instead of saying Dutch we had said Clashes with the wallpaper, I thought you like abstract work, Never saw it before, Tilted, Hanging too low, Beautiful, Hideous, Remember our camping trip last summer?, or whatever else might come into our mind when looking at a picture (in Skinnerian translations, whatever other responses exist in sufficient strength). Skinner could only say that each of these responses is under control of some other stimulus property of the physical object. If we look at a red chair and say red, the stimulus is under the control of the stimulus 'redness'; if we say chair, it is under the control of the collection of properties (for Skinner the object)

---

<sup>35</sup> Chappell, V.C. (ed), (1962), *Philosophy of Mind*, Englewood Cliffs, Prentice Hall, p.10.

<sup>36</sup> *Ibid.* p.10.

‘chairness’, and similarly for any other response. This device is as simple as it is empty. Since properties are free for the asking (we have as many of them as we have nonsynonymous descriptive expression in our language, whatever this means exactly), we can account for a wide class of responses in terms of Skinnerian functional analysis by identifying the ‘controlling stimuli’. But the word ‘stimulus’ has lost all objectivity in this usage. Stimuli are no longer part of the physical world; they are driven back into organism. We identify the stimulus when we hear the response. It is clear from such examples, which abound that the talk of ‘stimulus control’ simply disguises a complex retreat to mentalistic psychology. We cannot predict verbal behavior in terms of stimuli in the speaker’s environment, since we do not know what the current stimuli are until he responds. Furthermore, since we cannot control the property of a physical object to which an individual will respond, except in highly artificial cases, Skinner’s claim that his system, as opposed to the traditional one, permits the practical control of verbal behaviour is quite false.<sup>37</sup>

This criticism ushered in the age of functionalism and Cognitivism.

## **2.2 Functionalism and Cognitivism**

Functionalism dismisses the idea of mind as a thing, a logical substance. As the label ‘functionalism’ implies, this view envisages mind as a function. The function of something is the task it performs. Here, we should distinguish function from structure, which is the set of arrangements that make it possible for the thing to perform its function, tasks. It is, in other words, the embodiment of the function. It is also apparent that the same function can be performed by many different structures. Hence, functions are multiply realizable. This also implies that a function cannot be identified with the physical structure, which embodies it. However, a function always needs a structure in order to perform its task. A function is necessarily embodied and embedded in a structure.

---

<sup>37</sup> Chomsky, N., (1959), Review of B.F. Skinner’s Verbal Behavior, *Language* 35(1), p.51.

At the core of the functionalist view of consciousness and mind is the notion of information processing. An information processing unit encode input data, store them, transform them, and then they generate an output. Accordingly, there is a flow of information. At the level of input, there is analysis, pattern recognition, selection, and attention at work. Information is stored in variety of memory systems. Examples of transformation of information are thinking, problem solving, planning, creativity, and language. Output is comprised in functions such as motor skills, speech, and communication.



Now, the question is whether the functional analysis is applicable to mental states. Let us take the traditional instance of pain as our case in point. Pain is usually caused by some kind of noxious stimulus to body. There is a physical reaction to the noxious stimulus. As a result, there is some behavioral response at different levels such as behavioral reaction of wincing and groaning. From a functional perspective, pain can be analyzed in terms of an input in this case the noxious stimulus, an output in this case wincing and groaning, and importantly an internal activity between different processing units within the system in this case interaction between different mental states such as desire, feelings and beliefs. According to functionalism, once we provide for a given mental event an input, an output, and specify the interrelations between various mental states; then we have exhausted the description of the given mental state. There is nothing else to be posited or explained. This explains what a mental state such as pain is according to functionalism. However, it does not tell us what the mind is. From the point of view of traditional functionalists and psychologists, the mind was understood as a kind of 'black box'. It is locus of interaction between the inputs and outputs. Theoretically, the workings of the 'black box' are compatible with non-physical entities such as souls. However, the vast majority of functionalists are physicalists and they try to explain the

activities of the mind in terms of neurophysiology and natural sciences. Hence, the investigation of the workings of the mind falls under the jurisdiction of natural science and empirical investigation. Moreover, functionalism is committed to the multiple realizability of functions by different forms. Hence, functionalists make room for the presence of mind in inorganic arrangements such as computers or silicon based minds as well. However, it must be re-emphasized that functionalism seems to be metaphysically compatible with non-physical ontologies as well. In other words, metaphysical functionalism is ontologically neutral.

One of the metaphysical consequences of functionalism is a commitment to description of events in terms of different levels. The function of a computer, for instance, can be described in terms of hardware or software. This is not just distinction for the sake of convenience, but it is also a metaphysical commitment to distinct ontological levels of existence. Thus, we can think of talk of minds in terms of higher-level events and physiology in terms of lower-level states. Here, high and low do not entail value judgments but refer to ontological distinct levels of existence. So mental explanations are abstraction of the physiological events. We should not think of these ontological distinctions in terms of distinct substances but properties. Hence, mental descriptions are abstract and higher-level accounts of the underlying physiological lower-level properties. Here, the functionalists insist that although mental properties are realized by physiological properties, this does not mean that mental properties are identical or reducible to the physiological states.

Previously, we stated that a functional analysis of a mental property or state could be exhaustively achieved by specifying in input, an output, and interaction between different mental states. However, this account does not explain how these interactions occur. In other words, what is the ontological status of these interactions? Causality seems to be the only viable candidate that is compatible with the physicalists' ontological commitments and preferences. Hence, we can describe the mind as a causal structure and we should understand mental states and properties in terms of their causal roles. Mental

states are loci in complex causal networks. The function of these causal networks is to enter into causal interactions with the outside world through sensory inputs and causally affect the outside world by generating a behavior or output. Consequently, the mental entity is more than an object. It is an agent. It belongs to a specific environmental context. This relationship must be understood in terms of bidirectional causal efficacy. What distinguishes one mental event from another is precisely this contextually specified and distinct input-output causal interrelations associated with each mental state. In other words, mental events differ from each other because they are involved in different input-output causal affiliations.

So, what are the strengths of functionalism? Well, the valuable, and indispensable, strength of functionalism is that it allows us to remain physicalists without unnecessary strictness of reductive materialism. Functionalism is physicalism but it is liberal physicalism. It allows for the emergence of mind in different species and from different material configurations. This is due to the fact that functionalism is compatible with multiple realizability.

Another advantage of functionalism is that it conceives of mind in terms of causes and effects. This makes it describable, analyzable, and to some degree predictable. Moreover, the causal view of mind also corresponds to our intuition and common sense view of how our minds and the world interact with each other.

Functionalism conceives of mental states and events in terms of functions and systems, which comprise inputs, outputs, information processing units, and the relations thereof. Hence, the question whether mind causes behavior and vice versa is superfluous. These are all constituents of the same system and function. Hence, functionalism seems to circumvent the whole mental causation problem. It is the whole system of relations between inputs, outputs, and all the relevant functional states, which constitutes being in a certain mental state and not some specific stage or aspect of the process. The point is that if a mental state is understood in terms of a function of a system, which includes the input (potentially situated in the environment), a processing unit with recurrent

information processing loops, and an output, a behavior, that potentially changes the environment of the system, then the issue becomes a matter of direction of flow and processing of information and how the system behaves in the environment and what function it fulfills and not necessarily a question what caused what. Functionalism does not eliminate the mental causation problem. It shifts the priority and importance to the analysis of functions of the system. Hence, the casual relations between the components are maintained, but they are secondary to the function of the system as a whole.

In order to dispel any epiphenomenalist attempt to render consciousness useless, the functionalist position must propose an evolutionary advantage of consciousness, an indispensable function of consciousness. Mandler (1975), for instance, suggests: Relational processes operate primarily if not exclusively on conscious content. In addition, the choice, these include evaluation, comparison, grouping, categorization and serial ordering. In short, practically all novel relational orderings require that the events to be ordered must simultaneously present in the conscious field...Other relations have been established and stored subsequent evaluations are frequently unconscious.<sup>38</sup> This sounds very much alike to what Baars has to say about the function of consciousness, as we will see later. Mandler asserts that conscious operations bestow a number of clear and requisite evolutionary advantages:

1. Consciousness gives the individual a ‘troubleshooting function’ within the context of self-consciousness for all conscious functions and certain unconscious function can ascend to conscious inspection.
2. Consciousness gives the individual the possibility and capacity to strategize about future plans by retrieving and making conscious past desires and memories in light of present context and future possibilities.
3. Consciousness expands the possible ways to interact with the present environment.

---

<sup>38</sup> Mandler, G., (1975), *Mind and Emotion*, New York, Wiley, p.54.

Mandler states further:

Many of these functions permit the individual to react reflectively instead of automatically, a distinction that has frequently been made between humans and lower animals. All of them permit more adaptive transactions between the organism and the environment. Also, in general, the functions of consciousness permit a focusing on the most important and species relevant aspects of the environment.<sup>39</sup>

For Dixon, consciousness should be identified with ‘an action system in which the final product of interactions between sensory-inflows, stored information and need states is delivered up for the elaboration of plans and responses’.<sup>40</sup> Consciousness evolved in order to:

Hallmark those features of the external scene, which were at any one time of maximum importance to survival and upon which plans of actions should be based. A second and related functions of a consciousness system would be the provision of a mean whereby organism could contemplate their own need states, to mediate between inner and outer demands, and given the limited capacity of the effector system, to establish priorities for action.<sup>41</sup>

These remarks come in direct contradiction to what we said within the context of logical and nomological supervenience. This seems to support the notion that consciousness is nomologically supervenient on the physical. However, the problem back then was the question of mental causation and overdetermination, which functionalism proposes to have circumvented and resolved.

### **2.2.1 The Representational Theory Of Mind**

One of the hallmarks of functionalism is its depiction of mind as a symbol processor. A simplified system, for instance, would have a sensory input unit, belief processor—belief box, a desire processor—desire box, and an output unit that produces

---

<sup>39</sup> Mandler, (1975), p.57.

<sup>40</sup> Dixon, N.F., (1981), *Preconscious Processing*, Chichester Wiley, p.3.

<sup>41</sup> *Ibid.*

certain behaviors and responses to specific sensory inputs. Here, we do not deal with individual beliefs or desires but with complex belief systems and desire systems. The notion of symbol processing can be understood as follows: my belief that ‘tomorrow is Monday’ is coded as symbols stating this proposition and as such this proposition is placed in my ‘belief box’. In the same token, having desires to finish some projects on Monday are deposited in my ‘desire box’. Moreover, the belief box and the desire box are interconnected to the other cognitive, emotive, and conative systems in my mind. In fact, the collection of these systems constitutes the mind.

Here, we think of mind in terms of functional systems. This also allows different entities, living or non-living, to contain such functional systems. Hence, they can have minds. What makes a mind what it is, is function and not the constituent material. The representational theory of mind necessitates the supposition of mental representations, which are made of systems of systems. In fact, these systems of symbols must function as the mental representations. Fodor calls these the ‘Language of Thought’. This is akin to the software, or the ‘machine code’, found in computers. So in our case, the proposition ‘tomorrow is Monday’ is coded in symbols of the ‘Language of Thought’. To appreciate the subtlety of this point, we must distinguish between sentences and propositions. Sentences relay meanings, which can be understood as propositions. In other words, propositions constitute the semantics and sentences comprise the syntax. The sentence ‘tomorrow is Monday’ means something, relays some core information, which can be expressed in English sentences, French sentences, German Sentences etc. The syntax of each one of these sentences from different languages is understandably different. However, they all relay a message, a meaning, which makes it possible to realize that all these sentence are saying the same thing. The Language of Thought could be understood as universal syntax processor of minds, which codes for propositions—semantics. We can think of the same idea in terms of ‘semantic engines’. A semantic engine is a machine that can manipulate sentences—syntax—without taking into account their meaning—semantics. However, this manipulation is such that this purely syntactical

processing can generate meaning. So a person confronting a semantic machine would freely interact with this machine and would never suspect that the machine does not ‘understand’ what is ‘says’. A semantic engine is a syntactical processor, which works with formal relations among sentences.

Computers are instances of semantic engines, which fundamentally work with binominal symbols 0 and 1, which can, and in fact do, produce semantics. The symbols with which computers work are meaningful, but the computer does not ‘care’ about their meanings; they work at the level of formal relations between symbols.

Another example of semantic engines is the realm of formal logic. Here, we translate each proposition into a specific symbol and then perform logical operation and then again we translate the symbols back into propositions and look at the results in a meaningful way. For instance, let us let P stand for the proposition: If tomorrow is Monday, then I must pay my rent. Let Q stand for ‘tomorrow is Monday’. This is a simple formalization of the propositions and now we can apply a formal logical operation in form of syllogism to P and Q and derive a conclusion:

If P, then Q

P

Therefore, Q

The main point is that this derivation of Q from this constellation depends solely on the form of the argument. It is absolutely independent what P and Q stand for.

According to RTM, mind is a semantic engine. In other words, propositions of language and mental representation are coded in mind in the Language of Thought. This means that minds can process mental representations without processing intelligent homuncular agents, who understand the semantics beyond the syntax and formal processing of symbols. In this sense, brains just like computers are realizations of semantic engines. To be accurate, minds are specifically organized semantic engines realized by brains.

The pressing question at this point is how do we get from syntax to semantics or what is the relationship between symbol and meaning. The proponents of RTM, and in fact most philosophers, would propose that the meaning of a symbol is in no way inherent to the symbol. The relationship between symbol and meaning is a matter of context and convention, and custom. This holds certainly for computer, where the programmer assigns meaning to the symbols. But how does that work in brains? In other words, what is the source of meaning in the Language of Thought? One option that we can right off the hand reject is the possibility of presence of inherent meaning in the symbols of the Language of Thought. So, whence comes semantics? It would be also circular to assert that the meaning arises from the interpretation of the thinkers themselves. This assumes what we want to prove, namely the origins of semantic. Hence, it is viciously circular. Now there are different possibilities for RTM to resort to:

1. Semantics arises from the causal relationship between the agents and their environment.
2. Condensation of information in the processing loops: it is of common knowledge that information processing is not just linear in nature in that information flow from A to B to C and then there is an outcome. But information flows in feedback or feed-forward loops over and over again. Each turn accordingly symbolizes a higher lever of information processing. By each turn, information is condensed and this continuous condensation eventually produces semantics.

As we might expect, neither of these options provides a satisfactory response to the posed question. In the modern discussion, this problem is referred to as the *Grounding Problem*.

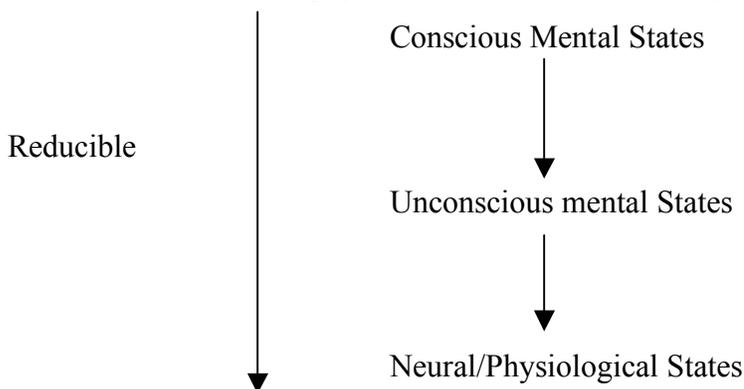
Traditional position of functionalism does not differentiate between mind and consciousness. It simply ignores the elements that make it problematic. This is also a realization of the subsequent functionalist position, to which we now turn, such as higher

order of thought theories (HOT theories) and a special case of HOT, namely the global workspace theory of Bernard Baars.

### 2.3 HOT theory of Consciousness

As we saw above, one of the most difficult challenges of any theory of mind is to account for mental causation. In other words, it is required to show how it is possible for mental properties to be causally efficacious in a physical world. Moreover, we determined that the problem of mental causation can be only satisfactorily resolved when we properly account for qualia and intentionality as fundamental properties of conscious experience. The combination of these problems has made traditionally the reductionism such an attractive option. One hopes to be able to reduce consciousness to something, which is more tractable and relatable to the physical properties we know.

High Order Thoughts (HOT) theories proceed to resolve this problem by postulating that consciousness is dependent on some sort of thoughts, which stand in intimate relation with unconscious cognition. In other words, consciousness is reducible to certain types of mental states that do not have to be necessarily conscious themselves. The advantage of this approach is that it reduces consciousness to other mental states, which can be easier related or reduced to neural states. Hence, the problem of reduction of consciousness to the physical states is transformed by introducing a hierarchy:



In other words, the direct reduction of consciousness to neural states is avoided by imposing intermediary states between the two. Consciousness, consequently, is indirectly reduced to neural states through intermediary mental states.

One of the most articulate proponents of the HOT theory of consciousness is David Rosenthal first enunciated in *Concepts of Consciousness* (1986). According to Rosenthal, mental states can be divided in two types:

1. Intentional states such as beliefs, hopes, desires, etc. These are propositional attitudes.
2. Phenomenal mental states such as feelings and sensations.

This distinction can be understood in terms of mental properties. Rosenthal states:

“All mental states, of whatever sort, exhibit properties of one two types: intentional properties and phenomenal, or sensory, properties.

...Some mental states have both intentional and phenomenal properties. But whatever else is true of mental states, it is pain that we would not count a state as a mental state at all unless it had some intentional property or some phenomenal property.”<sup>42</sup>

With respect to phenomenal states, Rosenthal further asserts:

“ Examples of sensory states that sometimes occur without consciousness are not hard to come by. When a headache lasts for several hours, one is seldom aware of it for that entire time... But we do not conclude that each headache literally ceases to exist when it temporarily stops being part of our stream consciousness, and that such a person has only a sequence of discontinuous, brief headaches.”<sup>43</sup>

According to HOT theory, a mental state becomes conscious when there is a *thought about* it. Rosenthal suggests, “a mental state’s being conscious consists in one’s

---

<sup>42</sup> Rosenthal, D., (1986), Two Concepts of Consciousness, *Philosophical Studies*, 49, p. 332.

<sup>43</sup> *Ibid.* p. 349.

having a thought that one is in that very mental state”.<sup>44</sup> Accordingly, a mental state becomes a conscious state, when one has a higher-order thought, or awareness, about that mental state. In a sense one can identify consciousness with *meta-mental* attitude. It is a mental attitude about other mental states. This assumes that consciousness presupposes conceptual thought. This also assumes that no mental state is inherently conscious, but it becomes conscious when it is object to a thought about it. In other words, consciousness is the product of a special kind of inter-relation among different mental states.

Subsequently, a mental state becomes conscious when the subject has a higher order thought about that mental state. This is a special sort of higher order thought: an *appropriately* obtained belief *about* the mental state that one is in that mental state.

Simply put, p is conscious of x, if and only if p believes x and knows that he believes p. what does it mean that the higher order thought has to be acquired *appropriately*? For HOT, this means that this belief may not have been obtained through inference or sense data. This formulation of consciousness requires that one is conscious of mental state *as* something. In other words, conscious experience is *aspectual*. This is a direct consequence of idea that all thoughts must be constructed from concepts. Furthermore, there is no consciousness without beliefs. Here, it is important to emphasize that HOT does not insist that one must be conscious of the underlying belief in each conscious experience. But the belief is the requirement for a mental state to become conscious. The underlying does not have to be conscious and in fact in most cases these beliefs are unconscious. Hence, conscious mental states presuppose unconscious beliefs. Rosenthal states:

“...it is natural to identify a mental state’s being conscious with one’s having a roughly contemporaneous thought that one is in that mental state. When a mental state is conscious, one’s awareness of it is, intuitively, immediate in some way. So we can stipulate that the contemporaneous thought one has is not mediated by any inference or

---

<sup>44</sup> Rosenthal, D., (1991), The Independence of Consciousness and Sensory Quality, *Philosophical Issues* 1, p.31.

perceptual input. We are then in a position to advance a useful, informative explanation of what makes conscious state conscious. Since a mental state is conscious it is accompanied by a suitable higher-order thought, we can explain mental state's being conscious by hypothesizing that the mental state itself causes that higher-order thought to occur.”<sup>45</sup>

There is a further categorization of consciousness, according to Rosenthal.<sup>46</sup> *Creature consciousness* is the property of an entity to have consciousness. This is a general property of creature/species. Furthermore, creature consciousness can be sub-categorized in *transitive* and *intransitive* consciousness. Transitive consciousness is property of a creature being conscious *of* something. This is the *directionality* of consciousness. *Intransitive* consciousness refers to a creature being simply conscious, a degree of vigilance that can be divided in awake, somnolence, stupor, and coma. *Transitive* consciousness, hence, implies *intransitive* consciousness. *State* consciousness is the type of consciousness, which different conscious states have. *State consciousness* can be further divided in phenomenal consciousness, which refers to qualia of mental states, and *access consciousness*,<sup>47</sup> which refers to functionally definable roles of mental state consciousness. The access consciousness in this context refers to beliefs, judgments, and in general propositional attitudes that are not necessarily phenomenological. One could think of them as being relevant in the agent's rational decision-making capability.<sup>48</sup> Or we can think of access consciousness as related higher-order thoughts, beliefs, or representations as in Rosenthal (1993, 1986). The following diagram illustrates the above-mentioned classification of consciousness according to the HOT theory.

---

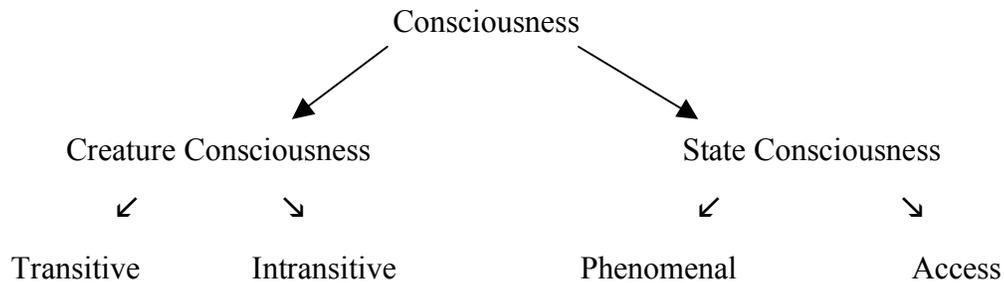
<sup>45</sup> *Ibid.*, pp. 335-36

<sup>46</sup> Rosenthal, D., (1993), State Consciousness and Transitive Consciousness, *Consciousness and Cognition*, 2, pp 269-270.

<sup>47</sup> Block, N., (1995), A Confusion about a Function of Consciousness, *Behavior and Brain Sciences*, 18.

<sup>48</sup> Dretske, F., (1995), *Naturalizing the Mind*, MIT Press.

Tye, M., (1995), *Ten Problems of Consciousness*, MIT Press.



As we can see, HOT theory of consciousness relates closely to notion of consciousness as an *inner awareness*. This is an awareness of mental states, which functions as a monitor of mental representations and states. This view also points to the functional interaction between consciousness and attention. Consciousness can be conceived as perception, or awareness, of the states of the mind.<sup>49</sup>

The question at this point is how does HOT theory account for phenomenal consciousness. The general approach of HOT theories with respect to grounds of phenomenal consciousness is to reduce it to some sort of representational content, or intentionality and causal-functional role.<sup>50</sup> There are two main trains of thought within the HOT tradition. One trend tries to reduce phenomenal consciousness to *inner awareness*.<sup>51</sup> This is like a monitoring system, which scans the perceptual data received through sense experience.

The other trend is more faithful to the main idea of HOT theories of consciousness in general. As we states above, a mental state is a conscious state, when it is the object of a higher-order representation or beliefs. It was also emphasized that the

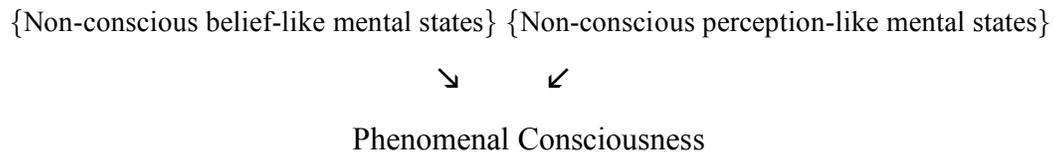
---

<sup>49</sup> Armstrong, D., (1981) *The Nature of Mind*, *The Nature of Mind*, University of Queensland Press, reprinted in *Readings in Philosophy of Psychology*, vol. 1, ed. Block, N., Harvard University Press, Cambridge, p. 199.

<sup>50</sup> Refer to Stanford online encyclopedia of philosophy.

<sup>51</sup> Tye, M., (1995), *Ten Problems of Consciousness*, MIT Press.  
Dretske, F., (1995), *Naturalizing the Mind*, MIT Press.

higher-order belief and representation does not have to be conscious itself.<sup>52</sup> Moreover, perceptual and quasi-perceptual mental states with a non-conceptual intentional/representational content are the candidates to reach phenomenal consciousness.<sup>53</sup> Hence, we can see that for the HOT theories the nature of these non-conscious mental states, giving rise to phenomenal consciousness, must be either belief-like or perception-like. Another important requirement is that the movement from the underlying non-conscious state to phenomenal consciousness must be proceeding non-inferentially.



One of the most exciting HOT theories is the global workspace (GWS) theory of Bernard Baars. Here, uses the metaphor of theatre to flesh out the intuitions of is theory. In the next section, I will proceed to provide the basic theoretical elements of GWS theory. In the next chapter, we will delve into the theater metaphor and will fill in the details of the theater metaphor with the standard neuroscientific current knowledge.

## 2.4 Global Workspace Theory

One the main problems of the HOT theories mentioned so far was that they must either postulate an inner awareness/organ to account for phenomenal consciousness. Or, they must rely on a potentially very large amount of ever presently available non-conscious beliefs to account for qualia. This latter brand of HOT theories is called *non-dispositional* HOT theory. A way out of this dilemma is to propose that there is a theatre

---

<sup>52</sup> Rosenthal (1995).

<sup>53</sup> Stanford.

of operation where the higher-order beliefs are made available to make the mental states phenomenally conscious. This theatre of operation is the working memory, which is a functionally active short-term memory system. This theory is called the *dispositional* HOT theory. In contrast to *non-dispositional* theories, which are pushed to the corner of postulating the presence of very large *actual* higher-order beliefs, the *dispositional* theories allow for the *potential* presence of such high-order beliefs in the working memory.

The global workspace theory, proposed by Bernard Baars, represents a non-dispositional exponent of the HOT theory of consciousness. According to Baars, we are only aware of a small percentage of the information being processed at any one moment.<sup>54</sup> The criterion, by which a piece of information becomes conscious, is determined by the salience and importance of that piece of information with respect to our current goals and projects. To elucidate his point, Baars utilizes the metaphor of theatre. According to Baars, the “the unconscious processors in the theatre audience receive broadcasts from a conscious “bright spot” on the stage. Control of the bright spot corresponds to selective attention”.<sup>55</sup>

This model is composed of four major components:

- 1) *The players*: these are the elements competing for access to consciousness. These elements are comprised of: A) *Outer senses* or sensory modalities, such as visual system, auditory system, somatosensory system, gustatory system, olfactory system, and the submodalities of heat, pressure, and vibration. B)

---

<sup>54</sup> Baars B.J., (1988), *A Cognitive Theory of Consciousness*, Cambridge University Press.

Baars B. J., Franklin S., (2003), How conscious experience and working memory interact, *Trends in Cognitive Sciences*, 7.

Baars B. J., Franklin S., (2007), An architectural model of conscious and unconscious brain function, Global Workspace Theory and IDA, *Neural Networks*, 20.

<sup>55</sup> Baars B.J, p. 957.

*Inner senses* such as visual imagery, inner speech, and dreams. C) *Ideas* such as imaginable ideas, verbalized ideas, and fringe conscious intuitions.

- 2) *The spotlight of attention illuminating the stage of working memory.* In this part selective attention illuminates and chooses one of the players on the stage of working memory. The working memory receives conscious input, control inner speech, uses imagery for spatial tasks. All this is done under voluntary control.
- 3) *The nonconscious audience.* The audience is comprised of: A) *Memory systems* such as lexicon, semantic networks, autobiographical and declarative memory, beliefs, knowledge of the world, and knowledge of oneself and others. B) *Motivational systems* such as desires, and goals. C) *Emotional complexes.* D) *Interpretive systems* such as recognition of faces, speech, objects, events, syntactic analysis, spatial relationships, social inferences. E) *Automatism* such as skill memory, details of language, action control.
- 4) *Context operators behind the scenes.* This consists of the director, the spotlight controller, and local context. This is the self.

According to our theatre model, working memory functions like a theatre stage or a global workplace for all mental events. Consequently, working memory presents that inner realm in which we can practice and review items such as telephone numbers, social security numbers, or we hold the narratives and the stories of our lives. Working memory is thought to include *inner speech* and *visual imagery*. Inner speech is what we hear ourselves saying quietly, when we rehearse a telephone number or when we read a passage in a text. Inner speech is composed of three components, in turn. It includes a *speaking*, a *hearing*, and a *comprehension* component. Verbal aspect of working memory seems to involve the same parts of cortex that is used in production and comprehension of language. *Broca's* area, located in the left lobe of the brain on the border between temporal lobe and the frontal lobe anterior to central fissure, seems to be responsible for

production of language. *Wernicke*' area, located on the border between parietal lobe and temporal posterior to central fissure, is responsible for the comprehension of language.

The function of visual imagery seems to help us display and solve spatial problems. Visual imagery is the mind's eye. It uses the same parts of cortex that are involved in visual processing. We will discuss visual processing in more detail in a subsequent section. For now, it suffices to say that visual imagery involves the occipital lobe, medial temporal lobe, and posterior parietal lobe of the brain.

One of the striking features of working memory is its surprising limited capacity. Studies have indicated that we can hold seven unrelated items in the verbal component and four items in the visual component. Another crucial feature of working memory is that it operates *serially*. This means that it processes items one at a time. To use our metaphor, working memory functions by presenting each individual actor addressing the audience one at a time by the spotlight being shifted from one player to the next.

As we asserted above, the players in the theatre metaphor represent the contents of the conscious experience. In other words, they are the different sources of conscious experience. One the most important hallmarks of the sources of the conscious experience, or players, is that they show the properties of *cooperation* and *competition* between themselves. An example of cooperation is trying to hold a conversation with a friend on a busy and noisy street. Here, the auditory input is complemented by the visual input through reading the friend's lips as he speaks. An example of competition is trying to study for an exam in a room where your friends are gossiping about a person of interest to you. Another important characteristic of conscious experience is its relative small capacity to process information. This can be understood and referred to the limited capacity of working memory.

Attention provides the spotlight through which different players are led to the forefront. In other words, the contents of consciousness emerge as the spotlight of attention falls on different sources of conscious experience on the stage of working memory. When the spotlight falls on an actor, then that player comes to consciousness.

This testifies to the privileged status of the actors and the regulatory character of attention function. Attention operates as the sensor that allows which actor should be allowed to speak and when. On the flipside, it seems that attention also regulates the negative function of suppression of the activity of certain actors at certain times. Players are privileged, since they are the only ones who are able to publicize information to the audience of specialists and experts. Remember, that this audience is made of unconscious sources of memory, cognitions, emotions, motivations, and automatic routines. The members of the audience not only receive information from the actors on the stage, but also they can communicate among themselves. Moreover, they can form coalitions in order to compete or cooperate with other members of the audience or the actors on the stage. However, the only way to address and relay information to the whole audience is through the actor in the spotlight on the stage.

As we have alluded above the audience plays the pivotal role in this model of consciousness. It is believed that the audience is comprised of the nonconscious cognitive, emotive, motivational, automatic system, which composes the majority of the mental life. The number and the scope of these nonconscious system are so vastly distributed that they constitute a vast society of specialized systems. Consciousness seems to play the role of mobilizing these vast assemblies of specialized societies. Consciousness, according to this model, is analogous to the public arena in a democratic society, where the concerns of the individual members and different interest groups are made public for sake of solving problems. This implies that; although, audience members are highly specialized and to a large extent autonomous, they cooperate with each other in order to overcome the challenges that the organism faces. It seems that audience members are connected to each other by parallel circuitry in order to exchange information consciousness. Evidence shows that many of these functions do not require consciousness. Nonconscious processing seems to be quite effective in many cases.

The role of the nonconscious contextual operators is to shape the events behind the scenes and set the background for the actors under the spot light to play their role.

The claim that they are invisible behind the scenes implies that they are not represented as the content of consciousness. However, they are directors that make the contents of experience possible. Consequently, they make experience possible. This is an apperceptive and transcendental role with respect to the activity of the actors, which represents perceptive consciousness. Our experience is always informed by the nonconscious expectations and contextual information.

One of the central assumptions of the GWS theory is that there is an intimate link between consciousness and attention. This link is represented and reiterated in our daily use of language. Take for example the propositions such as ‘we look in order to see’ or ‘we listen in order to hear’.<sup>56</sup> According to Baars, “the distinction is between selecting an experience and being conscious of the selected event. In everyday language, the first word of each pair (“look”, “listen”) involves attention; the second word (“see”, “hear”) involves consciousness”.<sup>57</sup>

Another major assumption of the GWS theory is that the majority of information processing in mind is done unconsciously by specific module in the mind. These modules are function specific processors spread throughout the brain. Such modules process information such as motion, color, shape, tone, etc.

A further assumption of GWS maintains that “conscious contents evoke widespread brain activation”.<sup>58</sup> The function of consciousness is the integration of information from the above-mentioned functionally specific information processors into an experiential whole. In other words, the function of consciousness is to ‘bind’ diverse unconscious data into qualitative intentional wholes. We can summarize the function of consciousness according to Baars:

---

<sup>56</sup> Baars B., 1997b.

<sup>57</sup> Baars B., 1997b.

<sup>58</sup> Baars B., Franklin, 2007, p. 956.

1. Consciousness relates input to its context. It gives a meaning and definition to input and hence it clarifies the input in our perceptual and cognitive experience of it.
2. Consciousness is specifically required for problem solving and learning of novelties in specific and learning and problem solving in general.
3. Consciousness plays a crucial role in adaptation of a mental event by raising its access priorities.
4. Consciousness allows for flexibility in response to challenges.
5. Self-consciousness, in form of inner speech and imagery, permits reflection and control.
6. Consciousness recruits a multitude of mental resources; conscious and unconscious, to muster up a response.

In sum 'consciousness appears to be the major way in which the central nervous system adapts to novel, challenging, and informative events in the world'.<sup>59</sup>

Dahaene and Nacache propose a version of GWS, which is similar to Baars' theory, but it goes further in that they propose to have identified the brain areas associated with conscious awareness.<sup>60</sup> Accordingly, conscious awareness depends on the simultaneous activation of several parts. The specific area depends on the modality becoming conscious. For example, conscious awareness for face recognition is the fusiform gyrus; while conscious awareness for motion is associated medial temporal lobe. In general, the areas involved in the GWS and conscious experience are located partly in prefrontal cortex, anterior cingulate, and further modality specific areas.

Dahaene et al suggest a three state when a visual stimulus becomes conscious.<sup>61</sup>

---

<sup>59</sup> Baars, B., Mc Govern, (1996), p. 92.

<sup>60</sup> Dehaene, S., Nacache, L., (2001) Toward a cognitive neuroscience of consciousness: Basic evidence and a workspace framework, *Cognition*, 79, pp. 1-37.

<sup>61</sup> Dehaene, S., Changeux, J.P., Nacache, L., Sakur, J., Sergent, C., (2006) Conscious, preconscious, and subliminal processing: A testable taxonomy, *Trends in Cognitive Sciences*, 10, pp- 204-11.

1. Conscious state: In this state neurons in the basic visual processing module are activated in conjunction with networks in the parietal, prefrontal, and cingulate cortex that are also involved in attention.
2. Preconscious state: Principally conscious visual experience is possible, since the relevant networks are sufficiently activated or activate-able. However, the top-down attention system is insufficiently active.
3. Subliminal state: In this state the conscious visual processing system is not sufficiently active regardless of activation level of the attention system.

This implies that conscious visual awareness requires the activation of two parallel systems:

1. The bottom-up visual processing system.
2. The top-down attention processing system.

The insufficient activation of either system leads to a lower grade of conscious experience.

## **2.5 Problems and Plausibility**

Functionalism seems to provide an excellent framework to capture the notion of mind within a theoretical context. However, functionalism seems to fail to distinguish properly between mind and consciousness in its classical formulation. The subsequent HOT and GWS theories make great strides in improving in this. However, the question remains unanswered whether they have been successful. This problem revolves around two essential aspect of consciousness: (1) the qualia of conscious states, which expresses their privacy and subjectivity; and (2) the intentionality, which expresses the ‘aboutness’ and referential nature of propositional attitudes such as beliefs, feeling, and desires. Two major arguments express this dissatisfaction with the functionalist theories. Ned Block’s ‘China Mind’ argument formulates functionalism’s failure to address the qualia issue; and John Searle’s ‘Chinese Room’ argument points to the functionalism’s shortcoming with respect to intentionality.



### 2.5.1 China Brain

The China brain, by Ned Block, provides an example of arguments, which point to the problem of qualia.<sup>62</sup> Suppose the people of China arrange themselves in a causally and functionally isomorphic mode, with every single person acting as a neuron with appropriate connections. Now, the function of the Chinese population is to simulate a program. This program imitates the workings of human brain at the neuron-to-neuron level. Given that neurons are basically input-output devices. This program allows for the simulation of an input-output system. Accordingly, the members of the population are provided with two-way communication devices.

The kind of communication devices they are equipped with tells them the number of other individuals who call in. Furthermore, the individuals are taught that certain combination of incoming calls will prompt them to take up certain actions, namely contacting other members of the population. So, when individual X receives contacts from individuals A, B, and C in a specific spatial and temporal order, then individual X will, in turn, contact individual Y. The program will provide precise and clear instruction about the workings of this system to each individual within the population. The original signal, which will start up this process, comes from outside of China. This fact allows this system to imitate the environmental inputs to the brain. The output will also culminate in some observable and verifiable action of Chinese nation. This is akin to the behavior of

---

<sup>62</sup> Block, N., (1978) *Troubles With Functionalism*, *Minnesota Studies in Philosophy of Science* 9, pp. 261-325, 278-280.

the individual after receiving an input from the environment, processing that input, and finally behaving in a certain way.

It is certainly granted that the China brain presents an inferior analogy system to organic brains in both the quantity and quality of its connections. However, it seems fair to ask the questions whether China brain, as a whole, can have phenomenal consciousness, or qualia. Is there something like being China? One could certainly argue successfully that such system would enjoy empirical consciousness and awareness. However, it does not seem like it would have qualia. Hence, it is logical to postulate that such system lacks phenomenal consciousness. In fact, it would be dubious if one insisted that such system had qualia. This is exactly the importance of this analogy. Whether in the natural world such system would be accompanied with, consciousness is an empirical matter at best. One might argue that China brain does not present the full complexity of the brains connections. However, one has to prove that extra number of connections would allow qualia.

One might argue that biochemistry and molecular biology has to be taken in account. This argument is in need of qualification as well. If one insists that it is biochemistry and molecular biology that are responsible for conscious experience, then one has automatically committed to the position that consciousness is only possible for DNA based life forms. Moreover, consciousness cannot arise from silicon-based life forms. This is a logically and naturally implausible position to take. The reason for this is the logical plausibility of the issue. There is no logical reason to assume that consciousness has to rise necessarily from carbon-based material and not any other material. It might very well be the case that all the cases of conscious individuals that we come across are carbon-based life forms, but there is no reason to reject the logical conceivability that it can be otherwise as well.

As far as natural possibility is concerned, as we stated above, all cases we know of where we suspect the presence of consciousness are carbon-based organic material. The point is, however, that we cannot a priori reject the notion of consciousness being

found in non-carbon-based matter. Although, a posteriori all conscious creatures we know of are carbon-based.

### **2.5.2 Chinese Room argument**

John Searle's Chinese room thought experiment presents the criticism against functionalism's failure to account for the intentional nature of consciousness.<sup>63</sup> Suppose the case of an English speaking person, who does not speak, or understands, Chinese. This person is locked in a room with an in-chute and an out-chute. He also possesses a book containing instructions in English with respect to operation of Chinese characters. Our person obtains occasionally questions through the in-chute. These are questions in Chinese. He uses the instruction book to find the appropriate responses to questions. Moreover, he copies the Chinese symbols corresponding to the answer on a piece of paper and places the paper in out-chute. It is important to note that our individual does not understand the meaning of the questions and the stories, because his instruction book is not a dictionary. The instruction book simply gives the individual the proper formula of what to do when he/she sees a specific Chinese symbol or strings of symbols. Hence, the individual's action is purely mechanical, syntactical, and in no way semantical. Our individual simply manipulates Chinese symbols based on the English instructions provided by the manual.

Form the perspective of somebody outside of the room, who provides the questions and stories to our individual, it would seem that this system understands Chinese and it is an intelligent system. From the third person perspective, we are dealing with an intelligently communicating system, since for all of our intelligible questions we receive intelligent answers. However, the story seems quite different from the first-person perspective of the individual inside the room. In fact, it seems quite wrong to infer that the individual understands Chinese. This points to a very significant point that seemingly

---

<sup>63</sup> Searle, J., (1991) *Minds, Brains, and Programs*, reprinted *In The Nature of Mind*, editor David M. Rosenthal, Oxford University Press, New York.

intelligible behavior does not prove the existence of intelligence behind the veneer. We can apply the same conclusion to qualia. Hence, seemingly purposeful and planned behavior does not automatically entail the presence of content of consciousness behind the behavior. The system seems to be semantic in nature, but it is essentially syntactic.

It is important to note that the individual and the system will pass the Turing Test. However, the Turing Test does not seem to guarantee a phenomenal content in conscious mind. According to the Turing Test, if a machine can trick us into believing it has consciousness, then it passes the Turing Test. However, it seems like the Chinese Room argument proves that the appearance of intentionality is not a reliable indicator of the presence of actual intentionality. A system that passes the Turing Test may be simulating a conscious mind; this does not mean that it really has a conscious mind. This conclusion seems to apply to any physical system, which shows stereotypic behavior.

Furthermore, Searle rejects the notion that consciousness can necessarily emerge from syntactical and computational system with sufficient complexity and appropriate architecture. He asserts that GOFAI systems (good old fashioned Artificial intelligence) cannot give rise to consciousness with content, since they are purely symbol processing systems. Searle (1997) reformulates his Chinese room argument to address this question.

Imagine that you carry out the steps in a program for answering questions in a language you do not understand. I do not understand Chinese, so I imagine that I am locked up in a room with a lot of boxes of Chinese symbols (the database); I get small bunches of Chinese symbols passed to me (questions in Chinese), and I look up in a rule book (the program) what I am supposed to do. I perform certain operations on the symbols in accordance with the rules (that is I carry out the steps in the program) and give back small bunches of symbols (answers to the questions) to those outside the room. I am the computer implementing a program for answering question in Chinese, but all the same I do not understand a word of Chinese. And this is the point: *if I do not understand Chinese solely on the basis of implementing a computer program for understanding Chinese, then*

*neither does any digital computer solely on that basis, because no digital computer has anything I do not have.*<sup>64</sup>

For Searle, in the case that such system is purely syntactic and not semantic, meaning that it processes symbols and does not understand meaning, and then the system is not intentional and *a fortiori* cannot be conscious. We can formulate Searle's position as such:

1. Programs are utterly syntactical.
2. Consciousness is essentially semantic.
3. Syntax is not identical to nor is itself sufficient, regardless of its complexity, for semantics.
4. Consequently, GOFAI systems have no consciousness.

The need for semantics, meaning, and intentionality in modern AI is expressed in terms of 'symbol grounding'.<sup>65</sup> For symbols to generate meaning, they must be *grounded* in world. This means that symbols can become semantic by being linked to events in the real world. This is achieved by internal representation of sensory input:

Such iconic representations first have to be categorized into recurring elementary features (which correspond to perceived features of the world). The association of symbols with such recurring feature categories would allow symbols to pick out the class of features or objects that they 'name', thereby 'grounding' the symbols. Once symbols are grounded in elementary features, the composition of symbols into strings would allow the generation of complex combinations that would inherit their grounding from their elementary constituents. For example, once the symbol 'horse' and 'stripes' are grounded in appropriate feature categories, one can derive 'zebra' ('zebra' = 'horse' and 'stripes'). Connectionist systems might achieve the pattern recognition of elementary invariance in input required for feature or object categorization in a natural, endogenous way.

---

<sup>64</sup> Searle, J., (1997) *The Mystery of Consciousness*, London, Granta, p.11.

<sup>65</sup> Harnad, S., (1990) The symbol grounding problem, *Physica*, D42, 335-346.

Harnad, S., (1991) Other bodies, other minds: a machine incarnation of an old philosophical problem, *Minds and Machines*, 1: 43-54.

Cognitive systems that manipulate symbols according to rules might then become grounded simply by incorporating a connectionist ‘front end’.<sup>66</sup>

Accordingly, words obtain meaning through their correlation with internal representational states. The representation becomes grounded in the brain via causal relationships between internal representations, actions, and external events.

Another fundamental question is whether we can do without qualia and intentionality and circumvent problems they raise. Dennett, for instance, subscribes to a kind of eliminativist position. According to Dennett, concepts like ‘mind’ and ‘consciousness’ are nothing more than attributions we make based on our observations of behavior.<sup>67</sup> For Dennett, these concepts are *useful fictions* of everyday life, but they don’t correspond to anything physically real. Accordingly, consciousness does not really exist. With respect to qualia, Dennett states:

Philosophers have adopted various names for the things in the beholder (or properties of the beholder) that have been supposed to provide a safe home for the colors and the rest of the properties that have banished from the external world by the triumphs of physics: raw feels, phenomenal qualities, intrinsic properties of conscious experience, the qualitative content of mental states, and, of course, qualia...I deny that there are any such properties. But I agree wholeheartedly that there seem to be.<sup>68</sup>

Furthermore, he states:

What science has actually shown is just that light-reflecting properties of objects...cause creatures to go into various discriminative states...These discriminative states of observer’s brain have various primary properties (their mechanistic properties due to their connections, the excitation states of their elements, and so forth), and in virtue of these primary properties, they...have secondary, merely dispositional properties. In human creatures with language, for instance, these discriminative states often eventually dispose the creature to express verbal judgments alluding to the color of various things. The

---

<sup>66</sup> *Ibid.* p. 91.

<sup>67</sup> Dennett, D., (1991) *Consciousness Explained*, London: Penguin.

<sup>68</sup> Dennett, D., (1994), Instead of Qualia, in A. Revonsuo and M. Kampinnenp (eds) *Consciousness in Philosophy and Cognitive Neurosciences*, Hillsdale, NJ: Lawrence Erlbaum, p.129.

semantics of these statements makes it clear what colors supposedly are: reflective properties of the surface of objects or of transparent volumes... And that is what colors are in fact... Do not our internal discriminative states also have some special intrinsic property, the subjective, private, ineffable properties that constitute the way things look to us (sound to us, smell to us, and so forth)? No. The dispositional properties of those discriminative states already suffice to explain all the effects on both peripheral behavior (saying 'Red', stepping on the brake, and so forth) and internal behavior (judging 'Red', seeing something as red, reacting with uneasiness or displeasure if red things upset one). Any additional qualitative properties or qualia would thus have no positive role to play in any explanation, nor are they somehow vouchsafed to us directly in intuition. Qualitative properties that are intrinsically conscious are a myth, an artifact of misguided theorizing, not anything given pretheoretically.<sup>69</sup>

He insists:

Nothing red, white, or blue happens in your brain when you conjure up an American flag, but doubt something happens that has three physical variable clusters associated with it— one for red, one for white, and one for blue, and it is by some mechanical comparison of the values of those variables with stored values of the same variables in memory that you come to be furnished with an opinion about the relative shades of the seen and the remembered colors.<sup>70</sup>

In sum, according to Dennett, the qualia of consciousness have no real existence.

Dennett's argument proceeds in four distinct steps to draw the above-mentioned conclusion:

1. The first step of the argument is to translate the first-person perspective into third-person perspective. In other words, he translates the 'what is like to' experience qualia (color for instance) into how a system might perform a task of discrimination and labeling.

---

<sup>69</sup> Dennett, D., (1994), p. 130.

<sup>70</sup> *Ibid.*, p. 136.

2. Furthermore, the task can be performed by brain without resorting to representations that are themselves qualitative.
3. Therefore, qualia are needed for functional analysis.
4. Hence, qualia do not exist.

The key premise of this argument is the first step, which implies that first person accounts can be translated into third person accounts without losing anything of importance. This seems to be the crucial claim and the source of the circularity of this argument. This is precisely the point of contention whether this translation is possible. The argument assumes the very thing it wants to prove. Hence, it begs the question. Moreover, Dennett fails to tell us why there should be the case that the third person perspective can act *as if* it is first person perspective. What is about third person perspective that lends itself to such pretension? This needs to be explained. Furthermore, the 'as if' must be given an ontological status as well. Dennett fails to do either.

## Chapter Three

### The Empirical Evidence

At this point, we will devote our efforts to a presentation of what neuroscientific research tells us about the nature of consciousness, perception, attention, memory, and self.

#### 3.1 Consciousness Correlates

What is consciousness in the scientific discourse? There are many things to think about consciousness from an empirical perspective. It can be understood as “the normal mental condition of the waking state of humans, characterized by the experience of perception, thoughts, feelings, awareness of the external world, and often in humans...self-awareness”.<sup>71</sup> For Koch and Tononi, “the most important property of consciousness is that it is an extraordinarily informative. This is because, whenever, you experience a particular conscious state, it rules out a huge number of alternative experiences”.<sup>72</sup>

However, in order to understand this phenomenon for empirical purposes, we must be more differentiative. This is exactly what Pinker suggests.<sup>73</sup> According to Pinker, we must differentiate between three distinct phenomena:

1. *Sentience*: this is subjectivity or qualitative character of conscious states. Sentience corresponds to qualia in the philosophical discourse.
2. *Access to information*: this refers to the reportability of conscious experience.
3. *Self-knowledge*: this refers to self-consciousness and the ‘meta’ ability of consciousness to examine its own activity.

As we discussed before, explaining ‘subjectivity’ is the most intractable component of this discussion. As we mentioned before, Chalmers calls this the ‘hard problem’ that is

---

<sup>71</sup> Colman, A.M., (2001), *Oxford Dictionary of Psychology*, Oxford, UK, Oxford Univ. Press, p. 160.

<sup>72</sup> Tononi, G., Koch, C., (2008) The neural correlates of consciousness: An update, *Annals of the New York Academy of Sciences*, 1124; p. 253.

<sup>73</sup> Pinker, S., (1997) *How the Mind Works*, New York: W.W Norton.

“question of how physical processes in the brain give rise to subjective experience”.<sup>74</sup> He further states:

If any problem qualifies as the problem of consciousness, it is this one...even when we have explained the performance of all the cognitive and behavioral functions in the vicinity of experience—perceptual discrimination, categorization, internal access, verbal report—there may still remain a further unanswered questions: Why is the performance of these functions accompanied by experience...why doesn't all this information processing go “in the dark“, free of any feel?<sup>75</sup>

The ‘access to information‘ constitutes the ‘easy problem‘, according to Chalmers. It is relatively easy as far as it lends itself to empirical investigation. Under this category, we can subsume functions such as discrimination and categorization of environmental stimuli, the ability to access one’s mental and internal states, and voluntary control of behavior.

From the perspective of neural networks, consciousness seems to be most likely associated with recurrent processing, also known as re-entrant processing.<sup>76</sup> Recurrent processing encompasses feedback from higher to lower areas of the brain, which results in extensive interconnections between different areas of brain. There are a variety of reasons why this is postulated. Firstly, recurrent information processing scheme is correlated with high complexity, which is to be expected in conscious processing. Secondly, there are a vast number of recurrent loop networks in the brain. Thirdly, consciousness seems to play a fundamental role in learning and re-entrant loops are correlated with the learning process at the level of networks.

---

<sup>74</sup> Chalmers, D., (1995b), The puzzle of consciousness, *Scientific American*: December, p. 63.

<sup>75</sup> Chalmers, D., (1995a), Facing up to the problem of consciousness, *Journal of Consciousness Studies*, 3, p. 201.

<sup>76</sup> Lamme, V.A.F., (2006) Towards a neural stance toward on consciousness, *Trends in Cognitive Science*, 10: 494-501.

There has been a great effort to identify brain structures and areas associated with consciousness. In this respect, visual processing has been extensively studied. Hence, the best evidence is yielded by visual processing areas associated with consciousness. Dehaene et al. (2006)<sup>77</sup> suggest that prefrontal, parietal, and cingulate cortex are associated with conscious awareness of stimuli. In case of visual stimuli, evidence points to the activity of superior parietal and dorsolateral prefrontal cortex<sup>78</sup>.

Further evidence suggests that consciousness is involved in integrated brain functioning. This should come as no surprise, since perceptual experience is holistic in nature. We experience many aspects of a phenomenon simultaneously, its color, its sound, its touch, and its smell for instance. It has been suggested that visual awareness is correlated with integrated networks of prefrontal cortex, occipital cortex, cerebellum, and superior temporal cortex.<sup>79</sup>

Another point of interest is whether consciousness comprises an unitary process. The phenomenon at the center of interest with respect to this question is the ‘split-brain patients’. This is our patients, whose *corpus callosum* has been disconnected due to generalized epileptic seizures, so the seizure is contained in one hemisphere. The *corpus callosum* is the largest bundle nerve axons, which connect the two brain hemispheres with each other. The question is whether the split-brain patients have a unitary consciousness or do they have two separate and distinct streams of consciousness. A further fact of interest is that information crosses the brain and it is processed by the contralateral side of the brain. For instance, the information of left visual field is processed by the right hemisphere and the data from the right visual field is processed by the left hemisphere. The information from left side of the body is processed by the right side of cortex and the data from the right side of the body is processed by the left cortical

---

<sup>77</sup> Dehaene et al. (2006).

<sup>78</sup> Rees, G., (2007) Neural correlates of the contents of visual awareness in humans, *Philosophical Transactions in Royal Society B – Biological Science*, 362, 877-886.

<sup>79</sup> McIntosh, A.R, Rajah, M.N., Lobaugh, N.J., (1999) Interaction of prefrontal cortex in relation to awareness in sensory learning, *Science*, 284: 1531-1533.

hemisphere. Moreover, the Broca and Wernicke areas are responsible for language production and comprehension and they are located mostly in the left cortical areas of parietal lobe and frontal lobe respectively.

Sperry suggested that these patients do have two distinct streams of consciousness.

Each hemisphere seemed to have its own separate and private sensations...the minor hemisphere [the right one] constitutes a second conscious entity that is characteristically human and runs in parallel with the more dominant stream of consciousness in the major hemisphere [the left one].<sup>80</sup>

On the other hand, Gazzaniga, Ivry, and Mangun maintain that the split-brain patients have a unitary stream of consciousness, which is based in the left hemisphere.<sup>81</sup> They call this system *the interpreter*, which tries to make sense of information available to it.

A left brain-system that seeks explanation for internal and external events in order to produce appropriate response behavior.<sup>82</sup>

This [system] generates a causal understanding of events that is subjectively complete and seemingly self-evident, even when that understanding is incomplete.<sup>83</sup>

The case of ‘incomplete understanding’ refers to patients with brain damage.

A further evidence for the Gazzaniga et al. position is that:

No split-brain patients has ever woken up following callosotomy [cutting the corpus callosum] and felt as though his/her experience of self had fundamentally changed or that two selves now inhabited the same body.<sup>84</sup>

---

<sup>80</sup> Sperry, R., (1968) Hemispheric disconnection and unitary conscious awareness, *American Psychologist*, 23: p. 723.

<sup>81</sup> Gazzaniga, M.S., Ivry, R.B., Mangun, G.R., (2009) *Cognitive Neurosciences: The biology of Mind*, (2nd ed), New York, W.W. Norton.

<sup>82</sup> *Ibid.*, G-5

<sup>83</sup> Cooney, J.W., Gazzaniga, M.S., (2003), Neurological disorders and the structure of human consciousness, *Trends in Cognitive Sciences*, 7, p. 162.

This goes to show that from the subjective, first-person perspective, the consciousness is unified in spite of the operation.

Furthermore Gazzaniga claims:

The left hemisphere...is constantly...labelling experience, making inferences as to cause, and carrying out a host of other cognitive activities. The right hemisphere is simply monitoring the world.<sup>85</sup>

Moreover, neuroimaging studies seem to point to a correlation between self-awareness and the right hemisphere.<sup>86</sup> TMS studies of the right hemisphere support this finding. Application of TMS to right prefrontal cortex disrupted self-perspective taking in the subjects.<sup>87</sup> These findings seem to stand in tension with Gazzaniga findings, who proposed a more prominent role for left hemisphere. These findings, however, give more prominence to right hemisphere. In final analysis, it seems the question of dominance of hemisphere with respect to conscious awareness is still open. The unitary and split nature of consciousness needs more evidence as well, but evidence seems to point to the unitary nature of consciousness.

---

<sup>84</sup> Colvin, M.K., Gazzaniga, M.S., (2007), Split-brain cases. In M. Velman and S. Schneider (Eds.) *The Blackwell Companion to Consciousness*, Oxford: Blackwell, p. 189.

<sup>85</sup> Gazzaniga et al (2009), p. 680.

<sup>86</sup> Gorman, M.E., (2007) Hypothesis Testing, Perspectives on thinking and reasoning, In S.E. Newstead, J.B.T. Evans, *Essays in honour of Peter Watson*, Hove, UK,: Lawrence Erlbaum Associates Ltd .

<sup>87</sup> Guise, K., Kelly, K., Romanowski, J., Vogeley, K., Platek, S.M., Murray, M., Keenan, J.P., (2007) The anatomical and evolutionary relationship between self-awareness and theory of mind, *Human Nature – An Interdisciplinary Biosocial Perspective*, 18, 132-142.

### 3.2 Perception: The Visual Information Processing Paradigm<sup>88</sup>

The reason we chose the visual system as the paradigm for sensory processing is that more is known about visual processing than any other sensory modality. The visual information processing begins at the retina and the different types of neurons in the eye. The retina is populated with two types of photoreceptors, the *rods* and the *cones*. Rods respond to light and dark contrast, night vision, and they respond to sudden movements. However, they do not produce the clear and precise vision that a normal eye experiences. This is the responsibility of the cones, which occupy the center of the retina in high density. This area is called *fovea* and it is correlated with clear and crisp visual experience. Moreover, the cones are responsible for color vision by reacting to photons with different wavelength. The photoreceptors, cones and rods, are connected to the optic tract via intermediary neural circuitry provided by bipolar cells, horizontal cells, amacrine cells, and ganglion cells. The sequence is such that the photoreceptors connect to bipolar cells. These cells, in turn, connect to the ganglion cells; and the ganglion cell, which project to thalamus and constitute the optic tract. The horizontal and the amacrine cells provide the function of lateral inhibition between the connections of the other cells. Hence, they care for the fine-tuning of the information. They provide the most basic level of visual processing.

---

<sup>88</sup> For reference for the contents of this section, refer to the following works. Since the content of this section represents general notions and findings in neurobiology, which represent common knowledge, a point-by-point reference is not possible. Hence, I provide a list of works, which I have used in my education and they can direct the reader into further readings for expansion of his/her knowledge base as well as verification of my illustrations.

Baer M., Paradiso M., Connor B.W., (2006), *Neuroscience: Exploring the Brain*, Lippicott Williams&Wilkins.

Kandel E.R., Schwartz J.H., Jessell T.M., (2000), *Principles Neural Science*, Mc GrawHill Professional.

Purves D., Augustine G.J., Fitzpatrick D., Hall W.C., (2007), *Neuroscience*, Palgrave McMillan.

Purves D., Brannon G., Cabeza R., Huettel S.A., (2008), *Principles of Cognitive Neuroscience*, Palgrave McMillan.

Squire L.R., (2008), *Fundamental Neuroscience*, Academy Press.

The optic tract begins at each eye and proceeds to the thalamus, and it terminates at the first part of the visual cortex. The ganglion cells project to a central thalamic structure, called *lateral geniculate nucleus*, or LGN. Here, the information from the left visual field of both eyes and the information from right visual field of both eyes are switched over; and the information is relayed to the visual cortex area V1. Moreover, the forward projection of information from LGN to V1 is complemented, and paralleled, by a cortical feedback loop. This implies that LGN is not a simple relay station, but it is involved in some level of information processing. The cortex seems to exercise some form of suppression on the information projecting from the LGN to V1. The LGN is composed of six layers. The lower two levels are composed of *magnocellular* neurons. The upper four layers consist of *parvocellular* neurons. *Koniocellular* neurons comprise a substructure that is sandwiched between the other two levels. The parvocellular neurons react in a *sustained* manner to a stimulating light pattern, while the magnocellular neurons do the same in a more *transient* way. In other words, magnocellular neurons respond to rapidly changing stimuli, while parvocellular neurons react slowly changing and stable input. Parvocellular neurons, furthermore, respond to color; and they seem to be responsible for vision of fine details. Consequently, destruction of parvocellular neurons of LGN will lead to lack of fine vision, color vision, and high fidelity spatial vision. Destruction of magnocellular neurons leads to the inability to detect rapid temporal changes. Koniocellular neurons are responsible of color vision as well.

Not all the ganglion cells from the retina project to the LGN of thalamus. Some of the ganglion cells project to the superior colliculus, or SC, on the superior part of brain stem. Interestingly, SC provides the most important visual processing center in fish, reptiles, and amphibians. In primates, SC is important for coordination of hand and eye movements, *saccadic* movements, and orienting responses in general. One of the main tasks of SC is to signal the difference between where the eyes are pointed now and where they will be next. This information is relayed to the areas of brainstem responsible for eye

movement, called oculomotor system. However, the patients who have an intact retino-collicular tract, while the V1 is damaged, lack conscious visual experience.

The visual cortex is located in occipital lobe, in back of skull. The organization of the different areas of the visual cortex is such that it produces an accurate map of the retina. Consequently, two neighboring points on retina are represented with two neighboring regions in the visual cortex. This means that the organization of the cortical map is *topographic*. Furthermore, the entire left half of the visual field is mapped onto the right V1 and the entire right half of the visual field is mapped onto the left V1. The central part of the visual field is given much more significance than the periphery. That was manifest in the retinal fovea. The same principle is manifest in the cortical organization as well. This is called *retinotopic* mapping. It is important to point out that the visual cortex contains multiple maps for the position, orientation, and direction of motion of stimuli. Moreover, these maps are functionally superimposed on each other.

As we mentioned above, area V1 is the first station where the information from each eye reaches the cortex. The neurons in the V1 react to light points surrounded by darker contrast, to straight bars of light and dark, to different orientations of these bars, and to moving bars of light. Areas V2 to V5 are, however, more specialized. V2 seems to specialize in orientation and color. V3 deals with shapes. V4 processes colors; and V5 is specialized in motion detection. The flow of information is hierarchical and recursive. This means that information is passed from V1 to V2 and so on. Moreover, at each level the scope of analysis broadens. However, higher levels such as V5 report back to the lower levels such as V4 and so on. This accounts for the complexity and recurrent nature of the visual circuitry.

The information processing at the level of V1, V2, and V3 constitutes the lower level analysis of information processing. Starting at V4 and V5 information is processed at more intricate level and information is relayed to the frontal cortex. In the theatre terminology, at this point the players begin to speak to the audience. In other words, from this level on sensation turns into perception. As we noted before, perception is holistic.

We don't perceive lines, bars, shapes, and colors. We perceive things and people in a coherent holistic manner. Consequently, the higher levels of analysis constitute the essential component of the act of perception.

Two main pathways emerge from V1—the *vision for perception* and *vision for action* pathways. These are also called the ventral pathway and the dorsal pathway respectively. The information from these pathways is later transmitted to the frontal cortex. It is believed that the neural structures that present this higher level processing of visual information are found along inferior temporal lobe (along the bottom of the temporal cortex) and the posterior parietal lobe (the top of the parietal cortex). This corresponds to the almost the rear half of the brain. These two pathways are incidentally are also known as the *what* and the *where* pathway. The *inferotemporal* (IT) cortex, or the ventral, pathway is the *what* processing circuitry. In this case, the stream originates in V1 and passes through V2 and V4 into IT and projects from there into *ventrolateral* prefrontal cortex, where higher order conception and cognition occurs. This pathway is responsible for the analysis of form, contour, color, and for detecting and discriminating objects. Here, through the analysis of shapes, colors, faces, and objects, it is determined what is it that we actually see. Consequently, it is wired for *vision for perception*. The posterior parietal circuitry, or the dorsal pathway, is the *where* pathway. This stream originates in V1 projects through medial temporal lobe (MT) to the posterior parietal cortex (PP). The dorsal pathway projects from PP further into *dorsolateral* prefrontal cortex. This pathway illustrates the relative location of the object with respect to other objects in the space. In general, they are concerned with space, motion and depth. The dorsal stream provides the necessary visuo-spatial cues needed for reaching and guiding the eye, the hand, or the arm. Hence, it is for wired for *vision for action*.

MT provides a closely linked system to the dorsal pathway, which is involved in processing motion. Moreover, MT responds to both real and imaginary motion. A patient with damage to MT shows total lack of motion perception. Patient L.M presents the classical case. This patient lost the functionality of her MT region on both sides of her

brain through vascular disorder. Consequently, she suffers from *motion blindness*, or *akinetopsia*. L.M is able to infer movement by comparing the relative location of objects in time. She has normal form and color perception, spatial acuity. However, her perception is not fluid at all. If I could use the analogy of motion picture, then I would say that we perceive the world as a motion picture and she experiences the world as a slide show.

In the previous section, we hinted at the recursive nature of the visual processing circuitry. We said that the information is constantly funneled back from the higher regions to the lower regions in a sort of reverberant circuitry. This feature of processing turns out to be very crucial for the production of a coherent perceptual experience, and it is the source of one of the most fascinating puzzles in the field of cognitive sciences, namely *blindsight*.

Blindsight seems to be a consequence of damage to the recursive circuitry from the higher levels of visual processing and V1. In other words, the information is not properly funneled back from the higher regions to V1. This damage can be the result of stroke or trauma. Typically, the blindsight patient receives visual information from their eyes relayed through thalamus. However, they report that they are not visually conscious of anything. Ironically, when they are confronted with an object, they report that they don't see anything. However, when they are asked to guess what the object in front them could be they give statistically reliable answers as the nature of the object, its location, and so on. They can even venture reliable guesses about some features of the object and they can track the object if moved. Nevertheless, they report explicitly that they have no visual experience of the object.

Blindsight provides a paradigm for the workings of the perceptual machinery. Processing information starting with low-level information and working to high-level information is called *bottom-up* processing. As we have seen, however, there are also *top-down* influences on perception, in which context influences how things function. These two systems combine in an interactive activation model to provide a comprehensive

perceptual experience. An example of this interactive cooperation is the word-superiority-effect. Consider the statement ‘Mary is zorking the third watch tonight’. The interaction between the top-down and the bottom-up information processing makes it possible for us to decode the sentence correctly into: ‘Mary works the night shift today’. The mistakes and the ambiguities with respect to ‘zork’ and ‘watch’ are thus resolved through a contextual analysis.

Disorders such as blindsight introduce the intriguing phenomenon of knowing without sensory experience. It seems like that knowledge is a function of nonconscious aspect of mind. So, in case of blindsight nonconscious visual knowledge remains intact. The puzzle of blindsight is that the type of information that is not presented in V1 is lost when V1 is damaged. As we saw earlier, V1 circuitry is involved in the analysis of light points surrounded by dark contrast and vice versa, light and dark bars of different orientations, and bars in motion. However, it seems through blindsight cases that V1 is needed for higher level visual processing and experience.

The neurons recognizing objects, shapes, and color are present in the higher regions of visual processing along the bottom of the temporal lobe. Damage to these areas will result in failure to perceive objects. The disorder of this area is called *visual agnosia*. The patient suffering from visual agnosia could be presented with a coffee cup. The patient would see the object and would be able to describe its qualities. However, the patient would not recognize the object *as* a coffee cup. However, if we hand the coffee cup to the patient. Upon somatosensory examination of the coffee cup, the patient would perceive it *as* coffee cup. In the case of visual agnosia, the holistic nature of visual experience seems to be missing. The difference between blindsight and visual agnosia indicates that damage to V1 leads to the loss of visual perceptual consciousness altogether. However, damage to medial temporal lobe will lead to deficit in the holistic nature of visual perceptual consciousness.

A special form of visual agnosia is *propagnosia*. This is the case of selective impairment of face recognition. The patient suffering from propagnosia may be not able

to recognize their own face on a photograph or the faces of family members. However, they can recognize objects such as tables, chairs, and watches. This suggests the mind might have two separate subsystems. One subsystem is designed for object recognition and the other for facial recognition. This implies that the object-recognition system can perform analysis of objects by breaking them down into their constituting parts and features and the other facial-recognition subsystem can perform the task of analysis in a holistic manner.

Another interesting fact about the function of V1 is not involved in dream imagery. Although, dream experience is distinct from waking conscious experience, dreams have their own qualia. It feels a certain way to dream. At one point, it was believed that brain activity during REM sleep was analogous to activity in the awake brain. However, it has been found that during dreaming V1 region is suppressed, while the activity of functionally *upstream* regions, such as *medial temporal* cortex and *fusiform gyrus*, is enhanced. These structures are assumed to provide the visual imagery and the sensation of seeing of the dream state.

At this point, we will say a few words about the function of prefrontal cortex (PFC) in general. From an evolutionary perspective, as organisms evolve, the complexity of their actions increases as well. Moreover, their goals extend further in space and time. Their area from which they draw their resources increases in actuality and potentiality and they plan for future events and anticipate the future consequences of their present actions. PFC evolved as the result of increase in these complexities, since they require planning, decision making, cognitive abilities, memory—as the storage and recall of information—and a feeling of authorship of one's actions and self. Consequently, PFC increases significantly in size with phylogeny. Moreover, PFC is extensively and reciprocally connected to *premotor*, *parietal*, *inferior temporal*, and *medial temporal* cortices, the *hippocampus*, the *amygdala*, the *limbic system* in general, and the *basal ganglia*. Furthermore, PFC is the only region, which directly communicates with *hypothalamus*, which is responsible for the release and regulation of hormones. However,

PFC has no direct connection to primary visual cortex, any other primary sensory cortices, or primary motor area.

PFC is, extensively, connected to the basal ganglia, which is phylogenetically ancient. This is a large subcortical region comprised of *globus pallidus*, *striatum*, and *putamen*. The function of basal ganglia is to mediate purposive movement, order the sequence of motor actions, fine tuning of motor skills, planning of motor action, and motor learning. Basal ganglia, however, connect indirectly, through thalamus, back to cortex. Clinical syndromes involving basal ganglia deficiencies include Huntington's disease and Parkinson's disease.

An important caveat in our discussion of conscious visual experience is the role of waking state in this. Experiencing something visually in the world presupposes being awake at different levels of arousal. It is to a short discussion of this topic that we turn next. Waking consciousness lies on a spectrum, which has coma and full alertness on the opposite ends. Damage to an area in the brainstem, encompassing the *medulla* and the *pons* formation, called *reticular formation* is involved in the arousal level. Patients with damage to reticular formation, such as loss of oxygen supply, fall in to coma. This finding indicates that reticular formation is necessary for waking consciousness. Usually, death follows damage to this area, since respiratory and cardiovascular control centers are located in the vicinity of reticular formation in the brain stem.

Reticular formation is necessary for waking consciousness, but its function is not sufficient for it. Thalamus is an active relay station for information to cortex. Thalamus consists of two lobes, or hemispheres, mimicking the function of the superordinated cortical hemisphere. Each thalamic hemisphere contains relay connections for the major cortical input and output tracts. This means that all major sensory streams, except olfactory tract, project to the corresponding thalamic nuclei before the information is transmitted to the proper sensory cortices. Hence, thalamus acts also as director of traffic to and from cortical centers. This implicates thalamus also in the function of *selective attention*. The major nuclei of each thalamus are separated by white matter, consisting of

myelin sheaths of axons, called *laminae*. The nuclei implicated in waking consciousness are present as islands in middle of laminae, also called *intralaminar nuclei* (ILN). In fact, ILN produce the signature cortical EEG of waking consciousness. Damage to both ILN leads to irreversible coma. However, if the reticular remains unharmed, then the patient will be in a state *persistent vegetative state*. Waking consciousness is a necessary condition for visual perceptual consciousness. Using our theatre metaphor, we could say that any happening in the theatre requires the theatre to be *open*. Nothing will happen in a theatre under lockdown.

### 3.3 Attention<sup>89</sup>

William James maintained that attention is a selective capacity, which leads to conscious perceptual experience. In our metaphor, it is the spotlight that brings an actor to the forefront of the stage. What we are usually conscious of is what we attend to. At any one time, only a limited number of objects can be selected in this manner. This implies that if two tasks require attention to carry out, then they will interfere with one another. James provides a phenomenological definition for selective attention:

My experience is those things I agree to attend to, those items I notice shape my mind—without selective interest, experience is an utter chaos...Everyone knows what attention is, it is the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought. Focalization, concentration of consciousness are its essence.<sup>90</sup>

When we perceive the world, at what point do we begin to regulate the quantity of data that we perceive? Our environment is loaded with information, which is available to

---

<sup>89</sup> For the content, please refer to the works:

Eysenck M.W, Keane M.T, (2005), *Cognitive Psychology: A Student's Handbook 5<sup>th</sup> Edition*, Taylor and Francis.

Quinlon P., Dyson B., (2008), *Cognitive Psychology*, Prentice Hall.

<sup>90</sup> James W., (1890) *The Principles of Psychology*, Dover Publication, New York, p. 380.

our sensory-perceptual system. Moreover, this information seems to persist. Take the example, of how much information we collect when we pass a glimpse on a scene. However, it appears that the perceptual system puts strict boundaries on what information is taken in. On average, the subject can hold four or five independent items in focus. Consequently, there is a *limited span of apprehension*, or perception. A compatible explanation is that the sensory storage capacity is high, but the transfer capacity to working memory is limited. Hence, by the time some items are reported other items fade away from the working memory. There is strong evidence of sensory memory. This storage capacity, or memory system, record sensory experience automatically. Moreover, sensory memory traces persist beyond their duration and they have a very large information capacity. *Iconic* memory is the visual sensory memory. *Echoic* memory is the auditory sensory memory. The limitation on the amount of information we process does not appear to be determined by the available information from our sensory memory system. Thus, these processing limitations must appear at a later point in the information processing. Research suggests that simple features can be processed in parallel.<sup>91</sup> However, more precise discrimination requires more focused, limited, serial capacity process.

One effective way to study attention is to ask the subject to look for something. These studies are based on a simple question: How does the time taken to find the target increase as the number of distractions increase? In some combination of target and distractors the search seems effortless. From a subjective perspective, the target *pops out* to the subject. Generally, this happens when the target is sufficiently distinguishable from the distractors in one of properties of color, size, motion, or form. This means that the ‘pop-out’ is dependent on not only on the local stimulus configuration, but also on more contextual effects as emphasized in *Gestalt* psychology.<sup>92</sup>

---

<sup>91</sup> Eysenck M.W, Keane M.T, (2005), *Cognitive Psychology: A Student's Handbook 5<sup>th</sup> Edition*, Taylor and Francis.

<sup>92</sup> Quinlon P., Dyson B., (2008), *Cognitive Psychology*, Prentice Hall.

In case of some other target-distractor constellations, the reaction time of the subject increases linearly with the number of items in display. This implies a serial search as opposed to a parallel search entailed in the ‘pop-out’ case. More light can be shed on this finding by realizing that abstract knowledge can direct attention. This should not surprise us, since abstract knowledge is a backstage component, according to our metaphor. We also know the interaction between backstage elements constitutes the decisive elements, which directs attention. This is called *top-down, task-dependent, focal attention*, or *volitional-controlled attention*. Top-down attention can be ascribed to specific features. This means that top-down attention is feature based, where the search is conducted throughout the whole visual field in favor of selected attributes. This selected attributes can be one feature present everywhere in the visual field or an extended object.

Some phenomenon, however, do not require focal attention to be noticed. These elements are conspicuous in virtue of their intrinsic qualities in context of their environment. An example would be one horizontal line amid columns of vertical lines. The saliency of an object is the function of bottom-up form of attention operating throughout the visual field. Saliency based attention is controlled by an explicit saliency map. Interestingly, the neurons of this map do not code for color, orientation, form, or such qualities; but they code for conspicuity. This means that they respond to how a stimulus is different from its surrounding and other elements in its context. However, this process is usually short-lived. This mechanism selects the most salient feature currently present in environment and directs attention to it so it is consciously perceived via a gating mechanism. Then, however, after a time the attention is shifted to the next salient feature in the environment.

The top-bottom and bottom-up attention systems combine in a way that saliency-based search and focal selection search comprise two components of the same system. This is done in a manner that the two systems maintain their unique abilities and characteristics. For instance, the top-down system is spatially proscribed, or focal, with respect to spatial specificity. It can select specific attributes with respect to feature

specificity. Its duration can be sustained with effort. It is task-dependent and under volitional control. The bottom-up system is not under volitional control; neither is it task-dependent. However, it is active at all times and in all feature dimensions. It searches throughout the sensory-perceptual field. Moreover, its duration is transient.

The cellular manifestations of attention can be grasped by the idea that attention helps one cellular coalition or assembly to establish dominance over other coalitions of neurons. This biased based competition theory assumes that attentional systems—top-down or bottom-up—influence the competition in favor of the attended stimulus. The function of attention is dependent on the relative *distance* between the stimuli and their *level* in the hierarchy of priorities of the individual. This suggests that we should expect little interference between neural coalitions as long as there is little overlap and competition between them. Bottom-up, saliency motivated system, and top-down, volitionally motivated system, influence the competition between coalitions until only one coalition, or few aligned coalitions, triumph in the anterior temporal cortex. V4 and posterior inferotemporal lobe (PIT) seem to play a key role in the attentional modification of visual perceptual tasks. In damage to these two regions, the subject can still distinguish between isolated targets, but not when the target is implanted in a dense visual field.

Imaging research has shown that focal attention is able to adjust reaction throughout cortex, including V1, V2, V4, MT, parietal and inferior temporal areas of the dorsal and ventral systems, premotor and prefrontal areas, and thalamus. Interestingly, the exact location depends on the exact context of the stimulus. The earliest observed effects of attention have been in the lateral geniculate nucleus, and V1. The source of top-down, focal attention appears to be in prefrontal cortex. Sources of bottom-up, saliency attention, seem to be numerous and they comprise pulvinar nuclei of thalamus, areas such as the lateral intraparietal area in the posterior parietal cortex, and the frontal eye fields.

Damage to any of the above mentioned structures (for instance, damage to right side) leads to pathologies of attention. One such pathology is *spatial hemi-neglect*, or *neglect syndrome*. Patients suffering from this syndrome fail to become aware of objects

on the left side of their space. A typical male patient, for example, would walk into a ladies' toilet, because he read the sign 'Women' as 'Men'. The 'Wo' part was simply never noticed. There is nothing wrong with the eyes of the patient. Neither, is there damage to V1 or other parts of the primary visual cortex, or the motor system. *Extinction* is a milder subspecies of the *neglect* syndrome. In this case, the patient can see a single isolated target in the left visual field. However, when a second target is introduced to the right visual field, it will take hold of attention. Consequently, the stimulus in the left visual field turns perceptually invisible. In other words, the right target eclipses the left stimulus.

Neglect is not just a visual syndrome. It can happen in the auditory and somatosensory modalities as well. In the case of the somatosensory modality, the patient will insist that, for example, his/her left arm does not belong to him/her; and it belongs to someone else. From a subjective standpoint, the patient of neglect is not conscious of the stimuli to his/her left side. This stimulus is simply not consciously represented. This is quite different from hemianopia, or blindness in one field of vision due to damage to V1. In the instance of the hemianopic patient, the subject is conscious of the loss and learns actively to adapt to, and rectify, this loss by turning his/her head to negotiate their environment.

As we saw earlier, parietal lobes are crucial for visual perception. Bilateral loss of parietal lobes would theoretically lead to total *neglect*. However, patients with *Balint's syndrome* show a different characteristics. In this case, there is bilateral damage to parietal lobes. However, the seal of this syndrome is fixation onto one point in the visual field. The patients, in this case, see one thing alone and they neglect everything else. They can describe the focused target, but they are not conscious of anything else. Moreover, the patients have no sense of space or spatial relationships, since the neurons in the posterior parietal lobe code for space, as we have seen before.

Theories of selective attention are concerned both with how people attend to information on some channels at the expense of others, and also with how much

information is processed on unattended channels. According to the *bottleneck model*, the mental apparatus includes a central processing system, which receives inputs from sensory channels and compares them with items stored in the memory system to determine their meaning. Overload of the central processor is prevented by means of a selective filter interposed between the central processor and the outside world that examines incoming stimuli by allowing through those, which enjoy certain attributes, and excluding other qualities. An important attribute of this filter is that it is flexible. Moreover, this filter is sensitive to high priority events that are not currently the subject of attention. For instance, one's own name is a priority signal in a conversation between people in the other side of the room that the information is relevant and important to us.

The workings of the filter occur at different stages. At the first stage, the incoming sensory information is analyzed for its physical attributes such as pitch, speech and non-speech properties, origin in space, etc., if it is demanded by the task, a decision or response can be made on the basis of this analysis. At the next stage, stimuli are checked against a list of high-priority messages maintained in the permanent memory system, such as danger signals or the vocal pattern of an individual's name. Stimuli passing this stage of the filter system receive further processing, leading to the comprehension of the meaning of the message, whereas those of low priority are ignored and typically receive no further processing. This implies that the analysis of a message's content occurs fairly early in the processing in that the check for the high-priority items involves the permanent memory system. Moreover, this is not a passive system, but very active. The early stages of analysis of physical attributes could happen automatically and nonconscious. The priorities at the higher levels are, however, continually aligned with the individual's purposes and expectations.

We could look at attention as a resource that can be allocated to a certain task. The processing that can be done on a specific task is proportional to the resources devoted to that task. Here, we assume that information can be processed the available

processing resources are all being used. Consequently, two tasks that do not demand too much cognitive effort should not interfere with each other. Moreover, if one task is more important than another, then resources can be devoted to the more important at the expense of the less crucial task.

Another function of attention is to bind perceptual features together into coherent representations of objects. The idea is that attention is required to create a unified object representation. For instance, when we perceive a red chair, it is clear that the chair we see has red color. Although, the coherence of the perceived combination of features seems self-evident, but it is not trivial, because we know that different aspects of vision are actually processed by different neural pathways. Consequently, at some point these data are brought together to form the representation of the object.

When subjects search for single features their search time does not increase with display size. This suggests that the subjects search all the members of a set of objects for the presence of a value of a single dimension in parallel with unlimited capacity. In contrast, when subjects search for conjunction of attributes, such as *chairness* and *redness*, their search time increases with display size, suggesting a limited capacity serial search.

In vision, a combination of features is part of the same object if those features occupy the same location in space at the same time. This means that we must combine information about *what* is in the world and *where* it is located. Although our perceptual experience is that we see both what something is and where it is located simultaneously, as we know, the brain separates these two kinds of information early in processing. As we discussed previously, there is a *what system*, or the *ventral stream*; and there is also a *where system*, or the *dorsal stream*. What and where systems appear to operate in parallel rather than sequentially.

Intralaminar nucleus (ILN) and another thalamic nucleus called reticular nucleus (nRt) seem to be involved in the resolution of the *binding problem*. This problem consists of the problem of how do different neurons and neuron assemblies in the visual cortex

combine their information into a single and coherent conscious experience. This is the problem of *temporal binding* of conscious experience. The answer to this problem seems to be a widespread oscillating electrical signal that acts as a pacing rhythm. The *forty-hertz hypothesis* claims that ILN and nRt seem to be the active source of forty-hertz oscillations, which sweeps from the frontal lobe all the way to the occipital lobe in regular intervals.

It appears that V1 is responsible for spatial binding of by acting as a spatial coordinator by holding all the higher-level visual processory areas in tune with a single retinal coordinate system, which guarantees the coherence of the picture. ILN, which is a necessary condition for waking consciousness, and nRt emit the forty-hertz signal to provide the temporal binding of the visual conscious experience.

The discussion above implies that attention is required to take information from many different dimensions and integrate it in a way that we can recognize that all information occurred at the same location. This suggests that there are distinct subsystems that operate in parallel early in the perceptual processing to represent low-priority attributes of objects. Hence, these features are easily detected. In contrast, conjunctions of feature are difficult to detect and process, since the information about each feature contains information where in space it is located. Hence, forming a representation of an object in space requires finding the various features that all occurred at the same location.

We know that saliency attention system is not under the control of perceptual consciousness, but it nonconsciously controlled. However, focal attention requires the knowledge about what can be potential candidates for perceptual consciousness. In other words, to chose one item among many presupposes the ability to think about our own mental states. Consequently, voluntary attention requires *metacognition*, which is self-consciousness. Being conscious of the ability to make a choice is not the same as consciousness of a perceptual phenomenon. One reason for this is that the choices we can imagine are abstract.

### 3.4 The Memory Paradigm <sup>93</sup>

*The direct realist theory* of memory is a counterpart to the naïve realism theory of perception. This view states that we remember an item as it originally was and it appeared to us. The event is presented to us directly by memory, as it was present in the memory. This theory is best understood as a causal theory. It assumes that a causal chain connects the subject to the remembered event. For instance, if I remember my ninth birthday party, then it is so partially because I had a ninth birthday party. This theory also maintains that a mediator agent such as an image does not produce our memorial beliefs. Memories present the past to us directly and not through any sensory medium.

The direct realist view, as stated, needs revision to be a viable theory. The first item of revision is the kind of causal relationship acceptable. Not just any kind of causal connection will do. Propositional belief about an experience, even one's own experience, does not bring about memory of that experience. An example would illustrate this point. Imagine the case of John, who was involved in a car accident. After the accident, John sustains a brain injury. The implication of this brain damage is that John has lost his memory of the events prior to the accident. Among the memories lost is his entire memory of his wedding to his wife. Now, his wife describes that event to him in vivid details and she shows him the movie from that event. At this point, John does have propositional knowledge of that event but it cannot be said that his knowledge is based on a justified memorial belief. The causal chain relating a memorial belief to a remembered event must be unbroken. Propositional knowledge about an event, even one's own action, does not give rise to memory of that event.

The revised version of direct realist portion, although, it asserts a correct causal chain suffers from some of the problems that naïve realism suffers from. Memory is subject to illusion and mistake. This point is illustrated by the fact that different people can remember the same event differently. This is analogous to illusory effects of the

---

<sup>93</sup> For reference, please refer to the references works cited in the sections on the visual system and attention.

perceptual system that it is possible, for example, to see a white paper yellow under yellow lighting or a rectangular object can be seen as different geometric shapes from different perspectives. The same holds for individuals, who experience the same event. They remember that event from their unique perspectives, hence, they remember it differently. Another objection to this theory is that it cannot account for memorial hallucinations. Memories cannot only be distorted, but they can also be totally fabricated. This does not imply moral intention of the subject. It is possible for a subject to have vivid and details memories of events that have never happened.

*The representative theory* of memory is the counterpart of the sense-datum theory of perception. This theory maintains that the sense objects function as intermediary items between the external world and the mind. The sense data present the external world to the mind. Here, memory traces are considered as a kind of filtrate of perception. The representative theory of memory is a form of indirect realism, remembering an event is interpreted as mediated by memory images. It is the memory images that accustom us with the past. We are truly justified in believing in a memory when that memory trace is grounded in a memory image. A true memory of an event is the image of that event, which is derived by a continuous chain of causation, from the experience of the event and the image represents that event to some extent correctly.

The representative theory can account for distortions in memory and hallucinations of memory. To remember incorrectly is to have a memory image that produces a false belief about an event. A hallucinated memory is one that has a vivid memory image but it is not linked causally to an event in the past. However, this theory has certain different problems. For one thing, a memory does not necessarily have to be an image. It is possible to remember an item without having a sensory representation of that memory trace. One example of this type of remembering would be recognition remembering. It is possible to recall an item of a memory trace automatically without an image of the whole memory trace. Another problem with this theory is that false memories do not have to be attributed to a faulty aspect of a memory image. For

example, I can have the false memory of spending my vacation in San Francisco, while I spent it in Seattle. The falsehood of that memory does not have to be reducible to a specific image. I don't have to have an image of the Seattle skyline to prove the invalidity of my memory. I can have simply the false impression of the whole experience, which is faulty while no specific image would point to the falsity of that memory. Retrospective thinking of a memory does not have to be grounded on an image either. For example, I can have memory of a conversation without having a specific image of that conversation.

The problems with the representational theory originate from its analogy to perception. This analogy is not necessarily false, but the representative theory draws its conclusion too intimately from this analogy. Recalling of a memory does not have to be a sensory or imaginal process.

*The phenomenalist theory* of memory is a correlate of phenomenalist theory of perception. Phenomenalism is a direct irrealist theory. It maintains that perceptual objects are directly perceived, but it denies that they are real in the sense that they are mind-independent and they exist apart from the perceivers. These objects are perceptually real. However, they are metaphysically irreal and they don't exist separate from perceivers.

According to the phenomenalist theory, remembering is based on the imaginal content of the experience at hand. To remember an event is to have suitable images of that event. The suitable collection of images presents the basis for the belief about that event. The phenomenism reduces the knowledge of objects to sense data. On the same token, memory of an event is reduced to images of that experience. This theory has the same problems that representative theory has images are neither the necessary, nor sufficient conditions for memory. Just like a collection of sense data does not necessarily imply the existence of an external object, so the collection of images does not necessarily imply the existence of a memory trace. No matter how vivid a memory image seems to be, it does not prove that that memory is true.

*The adverbial and the phenomenological theory* of memory is a corollary of the adverbial theory of perception and the theory of phenomenology, in general. The

adverbial theory maintains that when an object produces a memory trace, then in the subject a sensory experience of that object is perceived. In other words, one perceives an object when one is caused to experience that object in a certain qualitative way. The qualitative peculiar way that we are caused to experience an object determines the nature of our perception. Objects are perceived sometimes as they are and sometimes our perception is distorted. However, our perception is direct and intermediary agents do not mediate it. For example, when I look at a book lying on the table from the distance I see the book not as rectangle but as a parallelogram. Yet the book is seen and it is seen directly. Adverbial theory maintains that perception is perspective dependent and thus it should be described in an adverbial form and not with nouns. For example, my claim should be that I see the book parallelogramatically and not that I see the book as a parallelogram.

The claims of the adverbial theory of memory contain both negative and positive statements. First, it claims that remembering is not direct the way perception is. It also claims that memory is not temporally direct because past events are by definition not present and one cannot see the properties of something at the same time as that thing has them. Remembering is epistemically direct. It is based on a causal chain that is contiguous. Beliefs based on memories are not based on inference. In other words, the propositional knowledge of an event received through testimony or any other indirect source does not constitute memorial belief. Remembering does not have to be based on sense imaging. It is quite possible to remember an event without a complete image of that event.

The adverbial theory of memory also makes positive claims. Active or occurrent remembering is realization of a capacity. This capacity stands as a contiguous causal chain to the event. The realization of this capacity consists of images, expressions, and formation of beliefs with respect to that event. The realizations can also be of other kinds like recognition of elements of an event. Passive or dispositional remembering is an unrealized capacity. This pertains to inactive memory traces that are need of cues to be

activated. Propositional memory, remembering that, is a form of dispositional remembering. To have a memory of the occurrence of an event is to have a set of true beliefs about that event that are linked in a causal chain to the past. However, this belief set is memorial one. This is knowledge that has been sustained in memory. Remembering is knowledge from memory. This theory also makes distinction between propositional memory and event memory. Propositional memory does not require use of images but event memory uses images. However, event memory is a form of 'remembering as' and the event does not have to be entirely as it is remembered. Therefore, it allows for some fallibilism in the memory image. The truth of a memory belief is based on 'remembering that' or propositional memory and not 'remembering as' or event memory. If one really remembers some item or occasion, then one is correct about some aspect of it, or one is able to construct some true beliefs about it. This point is analogous to the point about perception that when one really sees something, then one sees that object as something. Therefore, perception should be described adverbially. According to the adverbial theory, memory is highly perspectival and so it should be described adverbially, since it has qualitative characteristic. It must be reiterated that the memory system is not a passive storage facility, but it is an active re-creative process. Remembering is not recollection, but reconstruction.

In the cognitive sciences the concept of memory is intimately related to the concept of learning. Learning is the process, by which we acquire knowledge about the world and memory is the process by which that knowledge is processed. The cognitive theory of memory considers both the structure of the memory systems and the processes working within that structure. The structure of a memory system refers to the organization of the system; and processes refer to the activities taking place within the memory system.

Learning and memory comprise a series of stages. Those processes going on during the presentation of the learning material are referred to as *encoding*. This is the first stage of processing. Encoding results in storage of information within the memory

system. Consequently, storage is the second stage of *processing*. The third, and the final stage, is *retrieval*, which involves the recovering and removal of stored information from the memory system. At this point it should be emphasized that the structure and processes of memory are closely linked. There is no active structure without processes and there is no retrieval without previous encoding and storage. It is only after the processes operate on the passive structure that the memory system becomes active and useful.

Memory store is a multi process system. First, there are sensory stores each of which holds information very briefly and it is modality specific. This means that each one of the sensory systems have their own memory store. For example, the visual system has its own memory store, called *iconic memory*. Second, there is a short-term store of a very limited capacity. Third, there is a long-term store of essentially unlimited capacity, which can store data for long periods of time.

Information from the external environment is initially received by the sensory receptors, which store their information in the corresponding sensory memory stores. Information is held very briefly in the sensory stores while some of it being attended to. The concept of attention plays a vital role in all memory processes particularly the encoding process. Encoding would not be possible without the engagement of some sort of attention system. Attention, as we discussed above, is the ability to direct the perceptual machinery toward a specific sense data receptive field. There are two types of attention systems: the active, focal attention, or top-bottom system and the passive, saliency attention, or bottom-up system. The active system has certain characteristics, as we discussed before. The active system occurs with intention, it is open to introspection, it takes mental effort, it uses mental energy and capacity, and it operates under the function of awareness. The passive attention system is also characterized by certain properties. The passive system happens without intention, it is not open to introspection, it uses up mental energy and capacity, it takes little effort, and it functions outside the awareness.

The short-term stores process the information that has been attended to. In turn, the information processed in the short-term memory is transferred to long-term memory by a process called *consolidation*. This transfer of information from the short-term stores to the long-term memory is determined by the rehearsal of the information and how well it can be rooted in the already present belief structure. Loss of information occurs at different levels of information processing. Data at the sensory store can decay, if it is not attended to. The information is subject to displacement in the short-term memory and the data in the long-term memory can be interfered with. *Retroactive* interference happens when information at present can interfere with information from the past and there is *proactive* interference, in which information from the past interferes with the information in the present.

At this point, let us explore the structure of sensory, short-term, and long-term memory and then we will discuss the processes of memory. Our senses are constantly bombarded with an enormous amount of information most of which does not get attention. Data in all sense modalities endures for some time after the end of stimulation, assisting the duty of obtaining its most important aspects for further analysis. Iconic store is the sensory memory structure of the visual modality. Icons are formed at the onset of visual stimulus. Therefore, within an ever-changing visual field it is the job of the icons to assist visual perception. The iconic memory has a limited temporal capacity of 0.5 seconds. Echoic memory is the memory system of the auditory modality. The echoic system has a play back mechanism. An example would illustrate this point, when somebody asks you a question and you are busy with some chore, you ask him or her to repeat themselves but then you realize you know the question and you answer it. The echoic memory has a limited capacity and information in it will decay after two seconds.

Short-term memory, also called working memory, is marked by two main characteristics. One, it has an extremely limited capacity; only about seven digits or units can be remembered. Second, the storage of information in the short-term memory is

fragile; any distraction leads to displacement and forgetting of information. The short-term memory has three main functions:

- 1) It holds the memory traces and it transfers them to the long-term memory via the process of consolidation.
- 2) It holds the data for limited time and then the information is discarded by forgetting.
- 3) It retrieves the information from the long-term memory so it can be worked on.

This is the workbench idea. The short-term memory is a searchable area. However, this search is exhaustive and any search in the short-term memory is done in a serial manner.

The short-term memory is composed of three structures. Attentional control system or the central executive actively focuses perception on specific events in the environment. It has a limited capacity as well. It regulates the flow information to rehearsal system for consolidation into long-term memory. The attentional system also regulates trouble-shooting when lower processing systems seem inadequate, tasks requiring planning or decision-making, and situation where poorly mastered response sequences are involved. Articulatory loop is another structure of short-term memory. In the articulatory loop memories for words and numbers are held. Moreover, they are processed for consolidation into long-term memory. Articulatory loop uses a subvocal speech system. This is apparent when items in rehearsal are repeated linguistically in the mind. The linguistic connection of the articulatory loop suggests its role in reading comprehension and understanding of speech. The visual sketchpad is the third structure of the short-term memory. It processes the visual and spatial properties of the memory trace. For example, this structure allows one to remember the face of a friend you met yesterday at your favorite coffee shop.

It is assumed that that the attentional and perceptual processes operate at the time of learning and they determine what information is stored in the long-term memory. They are different levels of processing ranging from shallow or physical analysis of a stimulus

to deep or semantic analysis. *Depth*, here, is defined in terms of the meaningfulness obtained from the stimulus. Furthermore, the depth or level of processing of a stimulus has a substantial effect on its memorability. Deeper levels of analysis produce longer lasting and stronger memory traces than shallow levels of analysis. Another important factor in the strength of a memory is its *elaborateness*. This refers to the amount of processing of a particular kind. *Distinctiveness* also effects the processing of memory. Traces that are distinctive or unique in some way will be more readily retrieved than memory traces that closely resemble a number of other traces.

As maintained before, long-term memory has unlimited capacity. The long-term memory system can be divided in two particular systems. This division is a function of role awareness in the retrieval of memory traces.

Accordingly, long-term memory system can be divided into two distinct systems of *implicit* and *explicit* memory. Implicit memory does not depend on conscious awareness for its retrieval. In other words, the subject does not have to be aware of the details of this memory trace at the time retrieval. Implicit memories are not open to introspection at the time retrieval. Implicit memories are mainly expressed through behavior of the subject. There are different types of implicit memories. *Priming* is a form of implicit memory. Here if the subject is exposed to an item once, then he will remember this item more readily next times. The point is that the subject did not paid active attention to the item the first time. Fear conditioning is another form of implicit memory. Phobias constitute another form of implicit memory. Classical conditioning and operant conditioning are forms of associative learning in which the relationship between two items are learned or the relationship between an item and its consequence is learned. This form of knowledge is also type of implicit memory. It is interesting to remember that in the western philosophy the association of ideas played a crucial role in the acquirement of knowledge. Emotional components of memory traces and trauma memories also have major implicit components. Motor skills, habits, and mastered knowledge of procedures are also types of implicit knowledge.

Explicit memories are those memories that depend on conscious awareness for their retrieval. In other words, the subject is actively aware of the retrieval process and this process is open to introspection. There are two types of explicit memories. There are *semantic* memory and *episodic* memory systems. Semantic memory consists of factual memories and propositions. For example, the knowledge that water is H<sub>2</sub>O is a form of semantic knowledge. The items stored in the semantic memory are independent of time and space. In other words, the memory and knowledge of facts is not stored with respect to time and space of the acquirement of those facts. Episodic memories are narrative memories and they have a strong autobiographical flavor. They are time and space dependent, since narratives are time and space dependent.

There are some crucial differences between episodic memory and semantic memory is that episodic memory has more temporal organization than semantic memory. When we remember an event, we tend to recall the aspects of that event in series, or narratively. Episodic memory also involves an association between a memory and its source. This is the foundation of memory as source of justification in a court of law. This leads us to the another crucial difference between episodic and semantic memory systems. This difference involves the way the truth of an item in memory can be established. An item in semantic memory is determined to be true intersubjectively, or by the consensus of a group as an established convention. In contrast, the individual determines the truth of a fact in episodic memory. In other words, episodic memory is subjectively true and semantic memory can be objectively and intersubjectively true. We will return to the importance of epistemic significance memory in knowledge at the end.

<b>Explicit</b>	<b>Implicit</b>
Late-maturing	Early-maturing
Episodic (particularly time/place)	Non-specific
With conscious awareness	Without conscious awareness
Specific	General
Concrete	Abstract
Strategic	Automatic

Context-dependent	Context-free
Partial retrieval (decay effects)	All-or-none
Weighted for function/meaning	Perceptually weighted for form
Intentional	Incidental
Associative	Non-associative

There are also well-studied cases of pathologies of memory system, called *amnesia*. There are two main types of amnesia. One is called *anterograde* amnesia. In this case, patient cannot learn new information after the injury. The other form is *retrograde* amnesia. Here, there is a loss of memory for events before trauma. The system of memory most effected in both cases is episodic memory. Hence, implicit memory systems and semantic memory are mostly spared.

At this point, we turn our attention to mechanism and process of memory storage in the long-term memory. Human knowledge consists of everything that we know. There is fundamental distinction between objects and the relations between objects. Therefore, any viable cognitive theory has to address both aspects. Concepts have a prototype structure; the prototype is a collection of characteristics attributes or the best example, or examples, of the concept. There is no delimiting set of necessary and sufficient attributes for determining category membership; there may be necessary attributes, but they are not jointly sufficient. Indeed membership often depends on the object processing some set of characteristic, non-necessary attributes that are considered more typical and representative of the category than others. Category boundaries are fuzzy or unclear; what is and is not a member of the category is not well defined. So some members of one category may also slip in other categories. Instances of a concept are ordered in terms of their typicality and category membership is determined by the similarity of the object with the prototype.

Concepts are represented in hierarchies of interconnected concept-nodes such as animal, bird, and canary. Any concept has a number of associated attributes at a given level of the hierarchy. For example, an animal has the attributes *has-skin* and *eats*, while

a bird has attributes such as *has-wings* and *can-fly*. Some concept nodes are superordinate of other nodes. For instance, bird is superordinate node with respect to canary node and the animal node is superordinate to bird. For reasons of cognitive economy, subordinate nodes inherit the attributes of their superordinate concepts; that is, as animal and bird are superordinates of canary, canary inherits their attributes. So a canary has attributes eats, has-skin, has-wings, can-fly, is-yellow, and can-sing. Some instances of a concept are, however, excepted from the defining attributes of its superordinates. For example, ostrich is excepted from the defining attribute of can-fly for the bird category. Various processes search these hierarchies for information about the concepts represented. So in concept verification tasks, for example in determining whether one concept is an instance of another; 'is a canary a bird?' a search must be made from one node to another. This leads to the prediction that the greater the distance between nodes the longer it should take to verify the statement. For example, 'is a canary a bird?' should be responded to more quickly than 'is a canary an animal?'. Similarly, if someone is asked whether a concept has a particular property, for example 'can a canary sing? As opposed to 'can a canary fly?' it should take longer to answer the latter because the attributes needs to be inferred from the superordinate bird node, rather than produced directly from the canary node.

There is clearly more to human knowledge than attribute-like information about single concepts and hierarchies of these concepts such as subordinate, basic, and superordinate levels. There are more complex forms of conceptual organization. Concepts are related to one other in ways that reflect the temporal and causal structure of the world. For example, in order to represent the concept of an event, such as reading the news on the bulletin board, it is necessary to have a knowledge structure that relates the act of reading to the objects involved. The knowledge structures that can represent this type of information are called schemata. Schemata consist of various relations and variable/slots, and values for these variables. The relations can take a variety of forms; they can be simple relations, for example is-a, hit, kick, or they can be more complex causal relations such as enable, cause, prevent, and desire. Variables contain concepts or

other sub-schemata. Any concept that fills a slot usually has to satisfy some test. For example, the argument-slot 'agent' in the relation 'breath' requires that the concept that fills it is an animate object. Therefore, values refer to the various specific concepts that fill or instantiate slots. Schemata, thus, encode general or generic knowledge that can be applied to many specific situations, if those situations are instances of schema. For example, the 'hit' relation could characterize a domestic dispute or a car crash. Schemata can often leave slots open or have them associated with them default concepts that are assumed if a slot is unfilled. For example, we are not sure what instrument John used in 'John wrote on the board' but we tend to assume a default value like chalk or marker.

Memories in the long-term memory are stored by being matched to the correct node/nodes and the proper schemata for that node/nodes. Once a memory trace confronts a schemata, which is concept specific mental structure, there are two possible outcomes: 1) *Assimilation*. Here the memory trace is incorporated into the present and relevant schemata. However, this incorporation implies some change in detail to preserve mental harmony and save cognitive energy, or 2) there is *accommodation*, where an available schemata is refigured to fit some new acquired knowledge. These, as might imagined, happen more infrequent than assimilation since it requires more cognitive energy and resources. This is why people are usually resistant to change. Remembering is not an act of recollection of events and facts. Remembering is an act of reconstruction of events and facts and extracting them from relevant schemata.

At this point, we will devote our attention to the molecular underpinnings of learning and the neuro-anatomical correlates of different memory systems and the corresponding pathologies.

Information about one's life experiences is processed in the various sensory systems of the *neocortex*, such as visual, auditory, and somatosensory cortices. These regions in turn send their information to the *rhinal* cortical regions, where multisensory representations are formed. The *rhinal* regions, which are comprised of *perrhinal*, *parahippocampal*, and *entorhinal* nuclei, then converge on the *hippocampus*. Within the

hippocampus, information coming into the *dentate gyrus* (DG) is processed and sent to the CA3 region, which connects with the CA1 region, which in turn connects with *subiculum*. Outputs of the subiculum are transferred to the rhinal areas, which then can send the information back to the hippocampus or back to the sensory neocortex for additional processing and storage. Consequently, the rhinal cortical areas and hippocampus, also called the *medial temporal lobe memory system*, is convergence zone, areas that receive and integrate inputs from diverse regions.

Working memory can process information from variety of sources. This allows the information to be compared, contrasted, integrated, and manipulated by the executive function of working memory. Hence, in order to perform these mental operations working memory must be able to store information temporarily. It appears that working memory is mediated by neural networks in the *prefrontal cortex* (PFC). Much of what is known about working memory has come from studies of visual processing. In visual system, as we have seen before, several subsystems play important roles in stimulus processing. One is involved in recognizing what an object is, the ventral pathway, and the other is involved in locating its position, the dorsal pathway. In working memory, these two types of representations come together. Consequently, objects have identity and location when we perceive them. The process of sending information to working memory from sensory systems is an example of bottom-up processing, whereas the control of sensory processing of working memory is an instance of top-down processing.

The cellular mechanism of working memory is understood to function as following: inputs to the prefrontal cortex from sensory, memory, or other systems terminate on the dendrites of neurons located in the middle layers of prefrontal cortex. These cells have extensive excitatory connections with other cells in the same layer and also connect with cells in the deep layers. Deep layers are connected with other deep layers, as well as with cortical and subcortical motor regions that control behavior. Deep-layer cells also connect with inhibitory interneurons. Dopamine neurons in the brain stem modulate all aspects of the circuitry in the prefrontal cortex, enhancing or facilitating the

excitation. The extensive excitatory connectivity in this circuitry, and its enhancement by dopamine, might underlie the ability of working memory to hold stimuli as long as the subject remains engaged in the task. The output of motor systems inhibits dopamine cells. This implies that once behavior is produced, the facilitation by dopamine terminates and working memory is released to do other tasks. However, it is important to note that working memory is not the function of one region but of a complex interconnected network in the prefrontal cortex. Some of the involved areas are the *lateral prefrontal cortex*, which is the classic working memory region, the *medial prefrontal cortex* and especially the *anterior cingulate region*, and the *ventral prefrontal cortex* and especially the *orbital region*.

There is a very intimate functional and structural connection between memory systems and the processing of emotions. The best-researched area of brain involved in processing of emotions is the *limbic system* in general and *amygdala* in particular. The amygdala receives low-level information about objects and events from sensory processing regions in the thalamus, and more complex information from sensory processing areas in cortex. There are different ways that hippocampal processing of context can influence amygdala. Evaluation of danger is thought to involve interactions between the hippocampus and amygdala. It has been shown that information processed by the hippocampus can be transmitted to amygdala through pathways originating in the rhinal areas of the cortex and subiculum.

Another point of functional interaction between hippocampus and amygdala is *stress*. In the presence of stressful stimuli, the central nucleus of amygdala activates the *paraventricular nucleus* of the *hypothalamus* (PVN). This is achieved either directly, or through *bed nucleus of stria terminalis* (BNST). *Corticotropin-releasing factor* (CRF) is released by the axons of the PVN into the *pituitary gland* (PIT), which in turn releases ACTH into the bloodstream, where it travels to the *adrenal cortex* of the adrenal gland attached to kidneys. The adrenal cortex then releases *cortisol*, which travels in the bloodstream to various organs and tissue sites in the body, including brain. In brain,

cortisol impairs hippocampal function and facilitates amygdala function. Since hippocampus normally inhibits the PVN and the amygdala normally excites it, the effects of cortisol can lead to a feed-forward cycle where cortisol release leads to more cortisol release.

Although the classical working memory area, the lateral prefrontal cortex, does not have direct connections with the amygdala, two other regions involved in working memory do have direct connections with amygdala. These are the medial prefrontal cortex, especially the anterior cingulate region, and the ventral prefrontal cortex, especially the orbital cortex. Since lateral PFC, medial PFC, and ventral PFC are interconnected; the lateral PFC may have some indirect access to amygdala through the other regions. Moreover, working memory is indirectly influenced by outputs of the amygdala to brain stem arousal system, the reticular activating system (RAS), which release modulatory monoamines in all areas of prefrontal cortex and by feedback from bodily responses initiated by amygdala activity.

Activation of emotional-processing systems by emotional stimuli has two consequences. One is the elicitation of emotional reactions that are automatic and nonconscious responses. The second is the activation of the motivational system that in turn guides actions. These are instrumental responses that are based on either past learning or instantaneous decisions. The *lateral nucleus* of amygdala mediates both automatic reactions, such as freezing, and instrumental actions, such as running away, in the presence of threatening stimuli. Defensive *reactions* seem to involve connections between *lateral nucleus* of amygdala to the *central nucleus* of amygdala. Defensive *actions* appear to concern connections from *lateral nucleus* to the *central nucleus*.

The motivational circuitry of the brain involves prefrontal cortex, *basolateral* amygdala of the limbic system, and the *ventral pallidum*. However, the key role is played by the *dopaminergic* projections from the *ventral tegmental area* (VTA) to the *nucleus accumbens*. The key to the understanding of motivational system is the role dopamine in elicitation of behavior. This happens, for instance, in the presence of threatening stimuli.

The lateral nucleus of amygdala processes the sensory properties of the threatening stimulus. Through the connections to the basal nucleus, information about the threat is sent to the nucleus accumbens. Through the central nucleus, dopaminergic neurons in the ventral tegmental area are activated and these, in turn, release dopamine in the accumbens. Dopamine facilitates the ability of accumbens to process the information from the amygdala. Consequently, an amplified signal is sent to the ventral pallidum, which in turn activates motor systems, which regulate instrumental, or motivated, behavior.

### 3.5 Self

We must make a distinction between the self and personal identity, personality, or empirical self. Personality is a self-concept. The self is a framework, which remains mostly stable across various life circumstances. Baars states:

“The implicit self seems to involve the deepest layers of context—the most basic expectations and intentions that guide our lives. Like any context, self seems to be largely unconscious, but it profoundly shapes our conscious thoughts and experiences. It seems to work behind the scenes of the theatre, pulling invisible stings to control the spotlight, shaping the actions planned and carried out with the aid of the theatre, and to some extent perhaps, the actors themselves.”<sup>94</sup>

Apperceptive self is the *deep* context of our lives. It can be interrupted, when the life goals of an individual are disrupted. Actions are valuable, because they express something essential in human existence. That is *future directedness* of human existence. Humans define their lives in terms of their projects, hopes, and dreams. It is our position with respect to our future goals, which determine whether our life is a success or a failure. Now, these goals don't have to be as grand as fighting for freedom. It could be as simple, or as complicated, as finding a lover for satisfaction of our temporary sexual

---

<sup>94</sup> Baars, *The Theatre*, p. 145.

desires, or buying a cool car. It is this goal that gives meaning to our life and it gives us an experience of being truly alive. The time factor of it is really relative.

On the other side of the spectrum, humans are necessarily historical creatures. Our historicity is a product of our culture, our upbringing, religion, or what we read in the papers two weeks ago. It is our past. Now, our past informs our present action, but it does not determine it. If it did, there would be no anxiety in human life, in a sense of existential anxiety. So, our present moment is the consciousness pregnant with potential tension between where we come from and where we want to go. Our freedom is not *ex nihilo*. It is simply our ability to reject some factors of our past in favor of our future projects. The anxiety is the indicator of the fact that I have to choose. Furthermore, whatever I choose as the overarching principle of my life cannot be proven logically to be the correct one.

Daily routines provide a shelter from this tension. We keep ourselves busy in our daily lives and if by instinct we become aware of this anxiety we manage to set aside this aspect of our existence. The structure of daily lives is so set up that we are concealed from this anxiety or Angst, as the existentialists describe it. This concealment of the anxiety is manifested in oblivious attitude toward this tension and an escape from our *true* self who is aware of this tension. Our daily existence, therefore, becomes a perpetual conformation to customs and habits. Everyday existence would be perfectly fine if one could go through life without experiencing the cracks on the surface of our daily existence. It would be a pleasant life if one could forever ignore the abysmal darkness below the surface of daily existence. But human beings do not only exist but also are capable of conscious contemplation of their existence. Therein lies our dignity as well as our misery. Once a crack starts to develop through the surface of our customs and habits and the foundations of our daily lives are shaken, then the abysmal darkness is exposed and anxiety supplants the calm.

Research in neurosciences suggests that at the highest level of organization, there is kind of self called *the interpreter*. The interpreter involves a narrative. Consequently, it

engages the speaking center of the left hemisphere. The split-brain cases provide an interesting insight into the nature of self. In these cases, the two hemispheres are disconnected from each other through a severance of the *corpus collasum*, a neural fiber connecting the two cortical hemispheres. Gazzaniga states:

We first revealed the interpreter using a simultaneous concept test. The patient is shown two pictures, one exclusively to the left hemisphere and one exclusively to the right, and is asked to choose from an array of pictures placed in full view in front of him the ones associated with the pictures lateralized to the left and right brain.

In one example of this kind of test, a picture of a chicken claw was flashed to the left hemisphere and a picture of a snow scene to the right hemisphere. Of the array of pictures placed in front of the subject, the obviously correct association is a chicken for a chicken claw and a shovel for the snow scene.

Split-brain subject P. S. responded by choosing the shovel with the left hand and the chicken with the right. When asked why he chose these items, his left hemisphere replied, “Oh, that’s simple. The chicken claw goes with the chicken, and you need a shovel to clean out a chicken shed”. Here, the left-brain, observing the left hand’s response, interprets that response according to a context consistent with its sphere of knowledge—one that does not include information about the other hemifield snow scene.<sup>95</sup>

Gazzaniga maintains that the different levels of self cooperate in order for narrative interpreter to be able to compensate for any violation. Consequently, it appears that the narrative function unifies the sense of self. There are, however, cases of *depersonalization* and loss of self. Psychiatric manual describe *depersonalization* as:

An alteration in the perception and experience of the self so that the usual sense of one’s reality is temporarily lost or changed. This is manifested by a sensation of self-estrangement or unreality, which may include the feeling that one’s extremities have changed in size, or the experience of seeing to perceive oneself from a distance...the individual may feel “mechanical” or as though in a dream. Various types of sensory

---

<sup>95</sup> Gazzaniga M. S., (2004), *The New Cognitive Neurosciences*, MIT Press, p. 1395.

anesthesia and a feeling of not being in complete control of one's action, including speech are often present. All these feelings are ego-dystonic (self-alien)...<sup>96</sup>

If we take the self to be the fundamental context of experience, then any pathology of self will follow from the violation of the context. Depersonalization is often generated by great violation of everyday expectation about the world—all cases in that routine, dominant goals and perspectives are profoundly challenged. This is congruence with the idea of that disruption of the self involves deep context violation. For the depersonalization patient, the world seems unreal.

Derealization is frequently present. This is manifested in a strange alteration in the perception of one's surrounding so that a sense of reality of the external world is lost. A perceived change in the size or shape of objects in the external world is common. People may be perceived as dead or mechanical...Other associated features include...a disturbance in the subjective sense of time.<sup>97</sup>

The consequence is a sense of real alienation from the world, in which the self is challenged and perception becomes estranged.

*Fugue* provides another example of self-alienation syndrome. It involves “sudden unexpected travel away from home or customary work locale with assumption of a new identity and an inability to recall one's previous identity. Perplexity and disorientation may occur. Following recovery, there is no recollection of events that took place during fugue.”<sup>98</sup> It usually “follows severe psychosocial stress, such as marital quarrels, personal rejections, military conflict, natural disaster.”<sup>99</sup> This is a form of *psychological amnesia*.

---

<sup>96</sup> Spitzer R L., (1979), *Diagnostic and statistical manual of mental disorders*. (DSM-III), Washington D. C., American Psychiatric Association, p. 259.

<sup>97</sup> Spitzer, *Diagnostic*, p. 259.

<sup>98</sup> Spitzer, *Diagnostic*, p. 256.

<sup>99</sup> Spitzer, *Diagnostic*, p. 256.

Another instance of self pathology is the multiple personality disorder. The psychiatric description of multiple personality disorder:

The essential feature is the existence within the individual of two or more distinct personalities, each of which is dominant at a particular time. Each personality is a fully integrated and complex unit with unique memories, behavior patterns, and social relationships that determine the nature of the individual's acts when that person is predominant...Studies have demonstrated that different personalities may have different responses to physiological and psychological measurements. One or more subpersonalities may report being of opposite sex, of a different race, or age, or from a different family than the original personality... The original personality and all of the subpersonalities are aware of lost periods of time...<sup>100</sup>

In this case, the hidden personalities feel a disturbance in the flow of their experience. Most of the cases of multiple personality syndrome correlate with long-term sexual and physical abuse in childhood and adolescence. Consequently, the childlike subpersonalities, or *alters*, keep vehemently the secret of the abuse. Moreover, the adult *alters* do not know anything about the secret. Usually, the alters take refuge inside a stable, and mature *host*. The *host* know about the *alters* and often take a nurturing and soothing role with respect to the childlike *alters*.

There is also clinical evidence that in the case of multiple personality, there is a breakdown to the unifying function of the self as deep context. Hence, there is a shift in the self. Clinical research indicates that the transition from one personality to another happens suddenly and it is associated with stress within the psychosocial context.

### 3.6 Concluding Remarks

Looking at the empirical evidence reveals some basic tenets. Firstly, the brain and the central nervous system is composed of networks of information processing units,

---

<sup>100</sup> Spitzer, *Diagnostics*, p. 257.

which deal with specific types of information, but they are plastic in nature. It means that once the need arises they can analyze other types of data. Secondly, the direction of information process is bidirectional. In other words, there feedback loops that run the data over and over again and in each turn information is reprocessed. There seems to be feedback loops as well feed forward loops. At each turn the information is either propagated or amplified, or condensed. This seems to add to level of complexity of data and it helps the organism to produce more complex regulated behavior. However, it seems an open question whether this condensation of information can lead to another quality of complexity. Here, I refer to the possibility of condensation of information in a syntactical system for the emergence of semantics. Is condensation of data in loops the source of intentionality?

## **Chapter Four**

### **Final Thoughts**

The purpose of this work was to present a historical and analytical of functionalism and its subsequent derivative theories. With ‘historical’ I did not mean a chronologically ordered accurate presentation of theories of mind and consciousness. With ‘historical’ I mean ‘historicity’. This refers rather to the notion that present forms of thoughts and paradigms are historically rooted. There is a train of thought, a process, which is necessarily temporal in nature and it evolves. Theories have a developmental history. Moreover, they look at future and they are creative. Hence, it is essentially important to understand this process within its proper context: past, present, and future direction. A theory is constantly informed by its past, influenced by its present context, and guided by its future direction.

Functionalism, in my judgment, provides an adequate framework for understanding the mind, especially the connectionist version of it. However, it fails at differentiating between mind and consciousness and it fails to account for the two essential characteristics of conscious states, namely intentionality, and qualia.

Moreover, functionalism describes mind’s function. It does explain the person. Its theory is an isolating abstraction, which takes the person out its context, reduces him/her to a brain function, and accounts for that function without accounting for the phenomenal content of his/her experience. That experience, in final analysis sets the causal chain of mental function in motion, which functionalism purports to explain. We can just pay lip service to this ‘context’. It is a fundamental aspect of our existence. And it is precisely this existence, which needs to be addressed in order to understand the person.

As we stated before functionalism states that consciousness stands in a causal relationship to other things. In other words, consciousness is related to other things by the principle of causality. This view implies that consciousness might have never been directed to its objects. Moreover, consciousness and its objects are logically independent of each other.

Functionalism also postulates a self-contained realm of consciousness. Here consciousness is analyzed in two ways. Either consciousness is identified with a kind of self and its thoughts, perceptions and sensory experiences; in other terms, ego and its cogitations. Or consciousness can be labeled as the stream or the bundle of sensations and perceptions. The deficiency in both methods is their failure to explain the fact that it is of consciousness' nature and essence to be *of* objects. Either they postulate that we make constant inferences from the cognitive life to the outside world, or objects are treated as complex bundles of sensations that are inferred from those very experiences. Neither viewpoint is, however, authentic to our phenomenological experience that our experience of the world is holistic. We don't experience the sense data and then make inferences about the world. We experience the world immediately.

A conscious act can, and should, be analyzed in terms of ego, cogitatio, and cogitatum. It is not that these are separate entities, which stand in logical independence from each other. This distinction is for analytic purposes only. The conscious act is a directed act, in which the consciousness intends an object. The directedness or intentionality is one of the essential features of consciousness, because it cannot be reduced to any other feature of consciousness and neither can it be inferred from more fundamental elements. Intentionality cannot be inferred from the stream of sensations. Moreover, it can explain the holistic nature of experience in that conscious acts have meaning by the virtue of intending an object. In other words, our consciousness of objects is mediated by meanings.

Intentionality of conscious act has immense consequences for our world-paradigm and the humanity's place of in it. The standard view presents the world as a collection of substances and discrete objects. These objects stand primarily in a spatial, temporal, and causal relationship to each other. These discrete objects are distinguished from each other through their intrinsic properties. Furthermore, they exist in logical isolation from each other. This view of the world can be described as a *spectatorial* one, in which human beings act as spectators of the world events. Understanding the world is a matter of being

caused to receive sensory data and perceiving those sensory data and finally making constant inferences about the external world from our perceptions. Humans are just perceivers of the reality and they play no metaphysical role in the articulation of that reality. Knowledge of the world is an empirical one of discrete objects, their intrinsic properties, and their interrelations. Our consciousness is a mirror of the reality and it stands in a causal relationship to this reality. This relationship is contingent and the two entities are logically independent of each other.

The standard spectatorial notion of the world is not necessarily erroneous, but this is barely the point of disputation. The problems emerge when this view is considered as the only correct view and the immediate one. The spectatorial view is the product of the reflective and passive interaction with the world. This standpoint, however, is dependent on another attitude, which is more immediate. This is the belief that the structure, organization, articulation of the world is a function of human agency. The world is not primarily a causal one. In fact, the fabric of causality is discovered in reflection on the world and not the moments of immediate encounter with the world.

The world is characterized primarily as sign-like, purpose-full or instrumental, and negative. These are three modes of our immediate understanding and interaction with the world through immediately. The world is a system of referential-totalities.

We encounter the objects in the world in two distinct ways. One is proximal and the other is reflective. The reflective attitude presupposes the proximal one. We come to discover objects as equipments that will fulfill our purposes and needs. The contents of the world are understood with respect to our projects and purposes and once they are proximally encountered and named then they can be subject of analysis and retrospection. Our proximal encounter with the world is not one of intellectual curiosity but one of concern for our projects and concerns. Therefore, the contents of the world have this feature of *being-for-the-sake-of* something. Our understanding of the world is not primarily a spectatorial one but we understand and discover the world through our agency.

Another feature of the world, which is essentially a human one, is that its contents stand as signs to each other. The world is a 'referential- totality', in which objects serve as signs and symptoms for each other and events. The drop in the temperature and the ice and snow are not the causes of winter, but they are signs that spring is upon us. Our understanding of the world is first through reference, directionality, and intentionality. It is through these features that we build a network of meanings and semantics values. Intentionality is the very essential feature of our consciousness. Once the world is understood proximally through our purposes and the referential system, then that understanding can be subject of the spectatorial attitude and retrospective inquiry. What need or impulse drives us to assume the reflective attitude? The reflective stance is taken when our proximal dealings with the world is subject to *breakdown*. This element of breakdown is the third feature of our proximal encounter with the world. It is important to point out that the sense of breakdown and lack is a proximal one and not its remedy, namely the reflective attitude.

Human existence is marked by two opposing elements that are in constant strife and struggle. On one hand, we exist biologically. We belong to the animal kingdom and as such we have needs and desires that are produced by our physiological functions. On the other hand, our existence is marked by transcendence. This transcendent element of our existence is the function of our consciousness. Self-consciousness makes qualitative difference in our existence. Self-conscious beings are capable of self-reflection and self-evaluation in the light of some overarching idea. In other words, humans are capable of evaluating their status quo in reference to their vision of themselves. Self-consciousness introduces a gap in the fabric of reality. Consciousness is able to step back from its interactions with the world and evaluate those interactions in terms of its ideals. Hence consciousness is capable of transcending its own limits and since those limits are self-imposed consciousness is able to transcend itself. Satisfaction of our basic physiological needs and desires is not all that matters to us because we care about what kind of beings we are. This element of care leads us to reflect on the worth of our wants and desires. In

the light of our evaluation of our needs and desires and our reflection upon their worth, we form higher order desires and needs which are more in line with our vision of ourselves. This is because we are capable of having aspirations toward things that transcend our immediate desires and needs. Hence the higher desires can regulate our striving toward the basic wants in light of the overarching principles we want live our lives. In other words, we take stand on our factual givenness. Taking a stand on our givenness creates the possibility of saying 'no' to what there is and all that there has been. This gap creates a perpetual desire to be filled, for consciousness suffers from the presence of this gap. However, this gap can be closed only by realization of those higher ideals. Hence, human existence is in a tension between our immediate desires and our ideals. There is a conflict between what there *is* and what there *should be*. Our existence is not comprised of a set of enduring qualities over time, but it is a perpetual event of becoming in which the gap between facticity and ideals is attempted to be filled.

The events of one's life resemble a narrative rather than a causal chain of events for two distinct reasons. First, the relationship of our consciousness to the world is not primarily one of causality. The consciousness does not stand immediately in a contingent relationship to the world. The consciousness is necessarily directed and our relationship to the world is one of a referential totality, which is motivated by our lacks and our purposes. We stand in a semantic relationship to the world. The second reason for the rejection of the primacy of the causal relationship is the anticipatory nature of our existence. Future plays a fundamental role in our existence in that our existence is one of goal-directedness. Human existence is always on the way of becoming something. There are no exhaustible sets of properties that can define a human being since our ideals are not realized but they define us as much as any tangible characteristic.

This is not to assert that causal interactions are illusory but the point is that our existence cannot be defined by causal relationship alone. Our existence and identity can only be defined with respect to the accumulateness or our past, and our ideals for the

future. The world only emerges, and things take on their shape through the emergence of values.

The distinction between subject and object dichotomy is based on the view of ourselves as a particular kind of thing or substance as opposed to the objects of the world as different substances. This view, however, is the progeny of the reflective attitude and the spectatorial role of human beings. It ignores the more proximal interaction of human beings in the world as agents with projects that see the world as a referential totality. The point is that our relationship to the world is not one of logically independent objects and subjects. The world comes to exist for us through our activities in the world. The world is steeped in semantics and meaning. The view that human beings are self-contained cognitive centers leads us to view the world as a theatre of spectacles. However, the world is essentially human and humans are essentially worldly. We are not primarily spectators of the world that play no substantial role in its articulation but we play a major role in the articulation of reality in the world.

The separation of mind and body is also the product of the spectatorial attitude this time toward ourselves. Our being as selves is inseparably involved in a 'practical life-world'. We come to understand the world by the meanings imparted by our consciousness through the activities of our bodies in the world. The fact that I can look at the mirror and inspect my body or I can use this body as an instrument that is subject to my will is secondary to the fact that I live through this body. The world reveals itself to my consciousness through this body. Hence the relationship between my consciousness and body is not contingent and I cannot exist logically independent of my body or consciousness. A human being is necessarily an embodied point of view. With these words I suffice to finish this project.

## Bibliography

1. Allison, A., (1983) *Kant's Transcendental Idealism: An Interpretation and Defense*, Yale University Press.
2. Armstrong, D., (1981) The Nature of Mind, *The Nature of Mind*, University of Queensland Press, reprinted in *Readings in Philosophy of Psychology*, vol. 1, ed. Block, N., Harvard University Press, Cambridge.
3. Baars B.J., (1988), *A Cognitive Theory of Consciousness*, Cambridge University Press.
4. Baars B.J., (1997a), *In The Theatre of Consciousness: The Workspace of Mind*, Oxford University Press.
5. Baars B.J., (1997b), Consciousness versus attention, perception, and working memory, *Consciousness and Cognition*, 5.
6. Baars B.J., (2002), The conscious access hypothesis: Origins and recent evidence, *Trends in Cognitive Science*, 6.
7. Baars B. J., Franklin S., (2003), How conscious experience and working memory interact, *Trends in Cognitive Sciences*, 7.
8. Baars B. J., Franklin S., (2007), An architectural model of conscious and unconscious brain function, Global Workspace Theory and IDA, *Neural Networks*, 20.
9. Baer M., Paradiso M., Connor B.W., (2006), *Neuroscience: Exploring the Brain*, Lippicott Williams&Wilkins.
10. Berkeley, G., (1979) *A Treatise Concerning the Principle of Human Knowledge*, ed. K. Winkler, Indianapolis, Hackett.
11. Block, N., (1978) Troubles with functionalism, *Minnesota Studies in Philosophy of Science* 9.
12. Block, N., (1982) *Readings in Philosophy of Psychology*, Harvard University of Press, Cambridge MA.
13. Block, N., (1995), A Confusion about a Function of Consciousness, *Behavior and Brain Sciences*, 18.

14. Brandon-Mitchell, D. & Jackson, F., (1994) *Philosophy of Mind Cognition*, Blackwell Publishing.
15. Chalmers D., (1986) *The Conscious Mind: In Search of a Fundamental Theory*, Oxford University Press.
16. Chalmers, D., (1995a), Facing up to the problem of consciousness, *Journal of Consciousness Studies*, 3.
17. Chalmers, D., (1995b), The puzzle of consciousness, *Scientific American*: December.
18. Chappell, V.C. (ed), (1962), *Philosophy of Mind*, Englewood Cliffs, Prentice Hall.
19. Chomsky, N., (1959), Review of B.F. Skinner's Verbal Behavior, *Language* 35(1).
20. Churchland, P., (1988) *Neurophilosophy: Toward a Unified Science of the Mind-Brain*, MIT, Cambridge Mass.
21. Colman, A.M., (2001), *Oxford Dictionary of Psychology*, Oxford, UK, Oxford Univ. Press.
22. Colvin, M.K., Gazzaniga, M.S., (2007), Split-brain cases. In M. Velman and S. Schneider (Eds.) *The Blackwell Companion to Consciousness*, Oxford: Blackwell, p. 189.
23. Cooney, J.W., Gazzaniga, M.S., (2003), Neurological disorders and the structure of human consciousness, *Trends in Cognitive Sciences*, 7, p. 162.
24. Davidson, D., (1994) *Mental Events*, reprinted in Davidson, *Essay on Action and Events*, Oxford, Oxford University Press.
25. Dehaene, S., Changeux, J.P., Nacache, L., Sakur, J., Sergent, C., (2006) Conscious, preconscious, and subliminal processing: A testable taxonomy, *Trends in Cognitive Sciences*, 10.
26. Dehaene, S., Nacache, L., (2001) Toward a cognitive neuroscience of consciousness: Basic evidence and a workspace framework, *Cognition*, 79.
27. Dennett, D., (1991) *Consciousness Explained*, London: Penguin.

28. Dennett, D., (1994), Instead of Qualia, in A. Revonsuo and M. Kampinnenp (eds) *Consciousness in Philosophy and Cognitive Neurosciences*, Hillsdale, NJ: Lawrence Erlbaum.
29. Dretske, F., (1995), *Naturalizing the Mind*, MIT Press.
30. Dixon, N.F., (1981), *Preconscious Processing*, Chichester Wiley.
31. Eccles, J.C., (1980) *The Human Psyche*, New York; Springer.
32. Eysenck M.W, Keane M.T, (2005), *Cognitive Psychology: A Student's Handbook 5<sup>th</sup> Edition*, Taylor and Francis.
33. Flanagan, O., (1992) *Consciousness Reconsidered*, MIT press, Cambridge Mass.
34. Flew, A., (1978), *Body, Mind, and Death*, New York; McMillan.
35. Gazzaniga M. S. ed., (2000) *The New Cognitive Science*, Bradford Book, MIT Press, Cambridge Mass., 2<sup>nd</sup> ed.
36. Gazzaniga, M.S., Ivry, R.B., Mangun, G.R., (2009) *Cognitive Neurosciences: The biology of Mind*, (2nd ed), New York, W.W. Norton.
37. Gorman, M.E., (2007) Hypothesis Testing, Perspectives on thinking and reasoning, In S.E. Newstead, J.B.T. Evans, *Essays in honour of Peter Watson*, Hove, UK,: Lawrence Erlbaum Associates Ltd.
38. Guise, K., Kelly, K., Romanowski, J., Vogeley, K., Platek, S.M., Murray, M., Keenan, J.P, (2007) The anatomical and evolutionary relationship between self-awareness and theory of mind, *Human Nature – An Interdisciplinary Biosocial Perspective*, 18.
39. Baer M., Paradiso M., Connor B.W., (2006), *Neuroscience: Exploring the Brain*, Lippicott Williams&Wilkins.
40. Harnad, S., (1990) The symbol grounding problem, *Physica*, D4.
41. Harnad, S., (1991) Other bodies, other minds: a machine incarnation of an old philosophical problem, *Minds and Machines*, 1.
42. Hobbes, T., (1997 (1651)), *Leviathan*, edited by R., Tuck, Cambridge, Cambridge

University Press.

43. Hobbes, T., (1970) *Hobbes Selection*, editor Woolbridge F. J. E., New York, Scribner.
44. Hoeffe, O., (1994) *Immanuel Kant*, translated by M. Farrier, State University New York Press, Albany.
45. Huxley, T. H., (1893) *Methods and Results*, Appleton-Century-Crofts, New York.
46. James, W., (1971), *Essays in Radical Empiricism*, ed. Perry, R. B., Dutton, New York.
47. James, W., (1981) *The Principles of Psychology*, Harvard University Press.
48. Kandel E.R., Schwartz J.H., Jessell T.M., (2000), *Principles Neural Science*, Mc GrawHill Professional.
49. Kant, I., (1929) *Critique of Pure Reason*, translated Kemp-Smith N, MacMillan.
50. Kant, I., (1983) *The Great Philosophers: Kant Selections*, ed. L. W. Beck, Prentice Hall, New York.
51. Lamme, V.A.F., (2006) Towards a neural stance toward on consciousness, *Trends in Cognitive Science*, 10.
52. Locke, J., (1984) *An Essay Concerning Human Understanding*, Oxford University Press.
53. Mandler, G., (1975), *Mind and Emotion*, New York, Wiley.
54. McIntosh, A.R, Rajah, M.N., Lobaugh, N.J., (1999) Interaction of prefrontal cortex in relation to awareness in sensory learning, *Science*, 284.
55. Nagel, T., (1974) What is it like to be a bat?, *Philosophical Review* 83.4.
56. Pinker, S., (1997) *How the Mind Works*, New York: W.W Norton.
57. Purves D., Augustine G.J., Fitzpatrick D., Hall W.C., (2007), *Neuroscience*, Palgrave McMillan.
58. Purves D., Brannon G., Cabeza R., Huettel S.A., (2008), *Principles of Cognitive Neuroscience*, Palgrave McMillan.

59. Quinlon P., Dyson B., (2008), Cognitive Psychology, Prentice Hall.
60. Rees, G., (2007) Neural correlates of the contents of visual awareness in humans, *Philosophical Transactions in Royal Society B – Biological Science*, 362.
61. Rey, G., (1991), Reason for doubting the existence of even epiphenomenal consciousness, *Behavioral and Brain Science* 14(4).
62. Rosenthal, D., (1986), Two Concepts of Consciousness, *Philosophical Studies*, 49.
63. Rosenthal, D., (1991), The Independence of Consciousness and Sensory Quality, *Philosophical Issues* 1.
64. Roesenthal, D., (1993), State Consciousness and Transitive Consciousness, *Consciousness and Cognition*, 2.
65. Searle, J., (1991) Minds, Brains, and Programs, reprinted In The Nature of Mind, editor David M. Rosenthal, Oxford University Press, New York.
66. Searle, J., (1997) The Mystery of Consciousness, London, Granta.
67. Sloman, A., (1991), Developing concepts of consciousness, *Behavioral and Brain Science* 14(4).
68. Smart, J.J.C., (1962), Sensation and brain processes, in V.C. Chappell (ed) Philosophy of Mind, Englewood Cliffs: Prentice Hall.
69. Sperry, R., (1968) Hemispheric disconnection and unit yin conscious awareness, *American Psychologist*, 23: p. 723.
70. Stanovich, K.E., (1991), Damn! There goes that ghost again, *Behavioral and Brain Science* 14(4).
71. Spitzer R L, Diagnostic and statistical manual of mental disorders. (DSM-III), Washington D.C., American Psychiatric Association, 1979.
72. Squire L.R., (2008), Fundamental Neuroscience, Academy Press.
73. Tonoï, G., Koch, C., (2008) The neural correlates of consciousness: An update, *Annals of the New York Academy of Sciences*, 1124.
74. Tye, M., (1995), Ten Problems of Consciousness, MIT Press.
75. Watson, J.B., (1913), Psychology as the behaviorist views it, *Psychological Reviews*

20.

76. Wittgenstein L., (1953) *Philosophical Investigation*, translated by G. E. M Anscombe, Oxford Blackwell.