COGNITIVE BASIS OF THE CONCEPT OF CONSCIOUS SELF

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

SARPER ALKAN

IN PARTIAL FULFILLMENT OF THE REQUIREMENS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
THE DEPARTMENT OF COGNITIVE SCENCE

DECEMBER 2005

Approval of the Graduate School of Informatics

_____

Assoc. Prof. Dr. Nazife Baykal
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

_____

Prof. Dr. Deniz Zeyrek
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

_____ _____

Assist. Prof. Dr. Bilge Say
Co – Supervisor

Assoc. Prof. Dr. Erdinç Sayan
Supervisor

Examining Committee Members

Assoc. Prof. Dr. Ayhan Sol (PHIL, METU) _____

Assoc. Prof. Dr. Erdinç Sayan (PHIL, METU) _____

Assist. Prof. Dr. Bilge Say (COGS, METU) _____

Assist. Prof. Dr. Samet Bağçe (PHIL, METU) _____

Assist. Prof. Dr. Erol Şahin (CENG, METU) _____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Surname : Sarper ALKAN

Signature : _____

ABSTRACT

COGNITIVE BASIS OF THE CONCEPT OF CONSCIOUS SELF

Sarper Alkan
MS, Department of Cognitive Science
Supervisor: Assoc. Prof. Dr. Erdinç Sayan
Co-supervisor: Assist. Prof. Dr. Bilge Say

December 2005, 55 pages

Contemporary developments in cognitive sciences have uncovered strong correlations between mental events and nervous activity. Despite their achievements, cognitive sciences are still unable to provide an adequate explanation for the concept of conscious selves. There are two main reasons behind this inability. The first one is the mismatch between the distributed nature of the nervous information processing and the unified nature of the consciousness. The second one is the fundamental differences between conscious experiences and the objects and events in classical physics. This thesis aims to provide a basis for a theory for consciousness and conscious selves by using certain findings of modern physics, neuroscience and philosophy. The argumentation is based on the biological necessity of having neural mechanisms that act like a self and quantum-theoretical approaches to consciousness. Consequently, it is shown that, the concept of conscious self is just a concept that we use to encompass several related concepts and experiences rather than being an ontological reality that is assumed by our common-sense intuitions.

Keywords: Self, Consciousness, Subjectivity, Agency, Free Will, Quantum Theory

ÖZ

BİLİNÇLİ BENLİK KAVRAMININ BİLİŞSEL TEMELLERİ

Sarper Alkan
Yüksek Lisans, Bilişsel Bilimler Bölümü
Tez Yöneticisi: Doç. Dr. Erdinç Sayan
Ortak Tez Yöneticisi: Y. Doç. Dr. Bilge Say

Aralık 2005, 55 sayfa

Bilişsel Bilimler alanındaki çağdaş gelişmeler zihinle ilgili olaylar ile sinirsel aktivite arasında güçlü bağlantılar ortaya çıkarmıştır. Ancak, bilişsel bilimler bütün bu başarılarına rağmen hala bilinçli benliklerin doğasını açıklamakta yetersiz kalmaktadır. Bu yetersizliğin arkasında iki önemli sebep vardır. Birincisi, bilincin bütünleşik doğası ile sinirsel bilgi işlemenin dağınıklığı arasındaki uyumsuzluktur. İkincisi ise bilinçli deneyimler ile klasik fizikteki olaylar ve objeler arasındaki farklılıklardır. Bu tez, fizik, sinirbilim ve felsefeden yararlanarak bilinç ve bilinçli benlikler için teorik bir temel oluşturmayı amaçlamaktadır. Tezde savunulan fikirler, temel olarak benliğe benzer davranışlar gösteren bir sinir sistemine sahip olmanın biyolojik gerekliliği ve kuantum teorisi ile ilgili bilinç yaklaşımlarını kullanmaktadır. Sonuçta, bilinçli benlik kavramının günlük yaşamdaki sezgilerimizin işaret ettiği gibi varlıkbilimsel bir gerçeklik olmakan ziyade sadece birbiri ile ilişkili bazı kavram ve tecrübeleri kapsaması için kullandığımız bir kavram olduğu gösterilmektedir.

Anahtar kelimeler: Benlik, Bilinç, Öznellik, Karar Verme, Özgür İrade, Kuantum Teorisi

ACKNOWLEDGEMENTS

I express my gratitude for my supervisor Assoc. Prof. Dr. Erdinç Sayan for his guidance and support and his faith in me. I am thankful my co-supervisor Assist. Prof. Dr. Bilge Say for her helpful comments and the support she gave me despite her medical problems. I am also thankful for Assist. Prof. Dr. Erol Şahin, Assoc. Prof. Dr. Ayhan Sol and Assist. Prof. Dr. Samet Bağçe for their helpful comments. I offer my special thanks for our institutional secretary Sibel Gülnar for her support and for her patience for answering my every silly question. Lastly I thank for my parents for their faith in me and for Begüm Mülayim for her support.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

“I am a graduate student of the Department of Cognitive Science in METU. You are reading introduction of my thesis. My previous argument is valid if you are reading this.” So far, everything is all right, and almost trivial. What I did is that I have made an assumption that if anyone reads what I have written, then something, which can be called “you”, is reading it. By doing that I have set a property of the thing which can be called “you”: it can read, which is an ability that only exists in persons. We believe that we are persons and are living with persons or groups of persons who can be called: I, we, you, he, she and they. Our language is based on person terms. We all know (at least experience) what is it like to be a person. But, we still don’t know what a person is. What is that “I” we are all talking about? Do I consider myself as myself, the same thinking being, in different times and locations as in the Lockean (Locke, 1975) sense of “person”? Or, am I only a bundle of experiences, which only lasts for the duration of the conscious awareness? Am I in charge of my body and act by my own volition or do I act as a consequence of a causal chain of events? Can there be a scientific explanation of the self or is it destined to remain untouched by any scientific approach?

In the philosophical literature, the debate over the nature of the self continues since the time of the ancient Greek philosophers. It is understandable that self has been discussed since that time by almost every philosopher. First, the answer of this question is very important from the viewpoints of law, politics, and psychology, and of course for almost every branch of philosophy. “What am I?” is a kind of question that everyone asks in their lives. If we can make an explicit definition of the self, then we can rid laws of vague terms about persons. Second, any

significant step taken in understanding the concept of the self will have direct implications on the solution of the problem of consciousness. On the other hand, implications of such a discovery, which are currently not well-conceived, can be seen on the area of information processing as well. If we can form a strong argument in defining the nature of the self, then the principles governing the formation of the self can be used for information processing purposes as well. Like the genetic algorithms, which are based on the theory of evolution, algorithms that imitate self in robots can then be formulated based on the theory of the self. Inducing self-like properties in robots may be crucial for the future developments of autonomous robots.

Many philosophers of mind, including Descartes believed the self to be identical with or to be linked with an immaterial soul. But, the epistemological problems with this view have caused it to lose its power in the last century. One of the problems is the problem of interaction. If anything immaterial, like a spirit, is our self, then how does it interact with our material body? This problem is still unsolved and is still a headache for the *dualists*.[1] On the other side of the discussion, materialism has its own claims and has its own difficulties as well. The weakest point for any materialist approach seems to be the phenomenal properties of the conscious events. A defense for the materialist viewpoint about its main problem is presented in the following section.

1.1. Methodology and research questions

As a basic assumption, throughout the thesis, conscious events will be considered as brain events. This physicalist and *monist*[2] approach will be the basis of our analysis of conscious selves. In many scientific publications we can see the correlation between the brain events and conscious events (Metzinger, 2003; LeDoux, 2002; Llinás, 2000; Damasio, 1999; Northoff & Bermpohl, 2004; Vogeley & Fink, 2003) and there is no reason that they would not be equivalent and the assumption of their equality provides us with a rather specific area to work on

[1] Dualist viewpoint claims that mind and body are separate entities each of which resides in a different reality. Mental events are thought to be associated with an immaterial soul while the body is thought to reside in the physical realm.

[2] Monism is the opposite of dualism. It rejects mind-body dissociation and claims that they both exist in the same realm. This realm can be a realm of thoughts in the case of idealism or the physical realm as in the case of physicalism.

which is the brain. At this point some may say that conscious events have phenomenal properties[3] associated with them. Since there is nothing in the physical world that corresponds to these phenomenal properties, there cannot be equivalence.

A thought experiment about a neuroscientist, named Mary, is used in the philosophical literature as an argument against physicalism (Ramachandran and Blakeslee 1999, p. 230). In this thought experiment, Mary is a future neuroscientist. She knows everything about the brain that neuroscience can provide at the pinnacle of neuroscientific developments. She knows everything about color perception, for example how the light stimulates the receptors in the eyes, how lights of different wavelengths are processed in the brain and so on. Despite her knowledge, however, she is deprived of one thing. She is raised in a black and white environment and she has never seen any color apart from the shades of gray. So, in spite of her enormous knowledge about the brain and color perception, when she goes out of the environment she has been raised in, she will learn something new: the raw experience of colors (i.e. the color *qualia*). Due to this discrepancy, some claim that there is a fundamental difference between the physical events and conscious events; so there cannot be equality.

The thought experiment about neuroscientist Mary may seem to be intuitively plausible. The main point of the thought experiment is that Mary cannot learn the raw experience of the colors without directly experiencing them and as she experiences them, she learns something new. This argument might sound an absolute rejection of the equality of the physical and the mental if you miss the circularity in the argument. Let us ask how Mary is educated in the field of neuroscience. During her education she might have acquired her knowledge through books, audio records, and video records and from human lecturers. If we disregard the visual information about the raw experience of colors that she might have acquired, because she is prevented from having that information, we can see the common means that she gained the information through: the language and the

[3] The phenomenal properties are referred to as qualia (singular: quale) in philosophical terminology. Qualia are the raw experiences that are associated with the conscious experiences such as redness, smell of a rose or sound of a lightning. For a review see Levin (1998).

raw experiences other than colors. Since she cannot get the information about what redness looks like, by using the other raw experiences she might have experienced, she could have acquired that information only through language.

Ramachandran and Blakeslee (1999, pp. 227-232) criticize language as an improper means for carrying information about the raw experiences. If we look at the elements of language, we can observe that the raw experiences serve as a basis for the language. Colors, sounds, smells, tastes and other feelings such as pain, anger and their compositions form the objects of our thoughts. Without sight and touch, we would not be aware of shapes, without shapes we would not form images of objects, and without objects our understanding of the world would be seriously impaired. Even the concepts of abstract objects are formed by our raw experiences. It would be impossible for mathematics to reach the level it has reached without studying it with the help of graphical descriptions. Apart from that, the number theory is ultimately based on the numbers of the objects that we perceive by our raw experiences.

If raw experiences are the basis of our understanding and the language is constructed on them, how can we use the language to explain raw experiences? No matter what the explanation is, it is bound to be a circular one. You cannot explain redness if the basis of your explanation is redness itself. Even if we found that the raw experiences are strongly correlated with physical phenomena (such as large scale quantum coherence in the brain) we would not be satisfied. We can still say that knowing the large scale quantum coherence in the brain does not grant us the experience of redness. But there is another way. Ramachandran and Blakeslee (1998, p. 232) state that if a cable can be drawn from the color processing areas of a person who can see red to the color processing areas to one who cannot, the second person might be able to experience redness as well. Similarly, if we could construct a device that can generate the experience of redness which we could couple to our brain, we would have got around the inability of language in explaining the raw experiences. We could perceive redness and we can perhaps understand "what is it like to be a bat"[4] by using a proper

[4] "What is it like to be a bat?" is the famous question asked by Thomas Nagel (1974). In his works he claims that it is impossible to know what it is like to be in another's consciousness. And he supposes "an organism has conscious mental states if and only if

device. As a result it can be seen that the thought experiment does not deny the mental-physical identity argument; it only points to the inability of language in carrying the information about the raw experiences. There are many things that we seek an explanation for and we do not experience directly. As we do not perceive the interaction of electrons directly, we may not perceive the experiences of other persons or animals, but that should not prevent us from searching for a physical explanation.

The arguments in this thesis will be based on physics. But, at many points I shall refer to subjective experiences in order for presentation of arguments. Even though a physicalist viewpoint will be used, there is no reason to stop using mental terms. Mental terms are necessary for our understanding of consciousness until they are replaced by better ones. Mental terms can be vague, such as “single experience”. But they will be used in such a way that their vagueness does not pose a problem for the arguments presented herein. As the discussion proceeds, vagueness of the mental terms which are directly related to self will diminish and their meaning will be clarified.

The approach to the problems about the self will be based on the necessity of having (or being) a self. If we can figure out why a self is needed, then a consistent theory about the self can be proposed by receiving support from those necessities.

The main objective of the thesis is to provide a coherent view of selves and conscious selves. The main problems about them fall under three categories: The first category is about the basic problems concerning the concept of self: What is a self? What is a conscious self? What is a person? The second is about the causal roles of the selves: Why a self is needed? Is there a causal role for a conscious self? The last is about some other aspects of the self experience: How can some other aspects of the self experience (such as subjectivity, free will, agency and personal identity) be explained within the concept of the self? More problems will arise as we proceed in our discussion.

---

there is something that it is like to *be* that organism – something it is like *for* the organism” (Nagel, 1974, p 166).

1.2.Organization

The thesis is build up on three parts: the need for a self, theories about conscious selves and the unified theory of the self. Each part will be presented in a separate chapter. The organization and the contents of the chapters are outlined below.

In Chapter 2 we will look into the meaning of the self and try to figure out the evolutionary significance of being a self. Emergence of a self in evolutionary history can provide answers that we seek. If we can find the ways in which being (or having) a self provides a better chance for survival then we may better understand the nature of conscious selves as well.

In Chapter 3 we will examine the basic theories about the conscious self. Two main theories about conscious selves will be presented and their strong and weak points will be discussed.

In the fourth chapter we delve deeper into the main problems about conscious selves such as unity, subjectivity, agency, free will, persons, and personal identity. The problems and the proposed answers will be presented in order to build up a coherent theory of conscious selves.

The last chapter will be the conclusion of the arguments that are given in the thesis. The solutions to the problems about selves will be displayed and directions will be presented for the future research.

CHAPTER 2

THE NEED FOR A SELF

The self and consciousness seem to be inseparable if one only considers conscious selves. According to that point of view, without consciousness, the self does not have the medium to operate in. But, if we are to understand the conscious self we must investigate its origins and we must ask why there needs to be a self.

2.1. What is a self?

Cambridge Dictionary of American English (2005) defines the self as: “who a person is, including the qualities such as personality and the ability that make one person different from another.” On the other hand Compact Oxford English Dictionary (2005) uses the definition: “a person’s essential being that distinguishes them from the others.” The inclusion of *person* in those definitions may induce circularity for the purposes of this thesis and it needs to be removed. But there is another thing which is common in these definitions: the self seems to be something that is dissociated from the others, and the others are what are not included in the self.

According to the above characterization, the self can be anything as long as the others exist. It can be a country, an ethnic group of people, a race, a species or a living organism. But this dichotomy is not sufficient to fully encompass the meaning of ‘self’. For example, think of a pencil. Just because there exist other things and other pencils, a pencil cannot be said to be a self. So what is the difference between a country or a living organism and a pencil in the respect of being a self? The difference lies in the “active participation” of the selves in maintaining their dissociation from other things of the same or different sort. A

country protects its borders and tries to maintain its integrity, while a living organism fights with diseases and other factors that can disturb its well being. To perform these activities, the self must distinguish of what belongs to it and what does not. If a self fails to keep the distinction, it will certainly face with the danger of "destruction." A country can be invaded or a living being can succumb to a disease. So, in addition to being an individual among the others, I argue that, a self must be able to distinguish itself from the others or distinguish others from itself.

The logic of the above description of self can be questioned by inquiring the meaning of the terms "active participation" and "destruction." A pencil can be said to be "actively participating" in keeping its integrity (and distinguish itself from the others) via the bonds between the atoms that the pencil is composed of and resist "destruction." A reply to the above claim requires another aspect of a self to be revealed: A self must either be a living organism or be composed of living organisms. In that respect, first, "destruction" of a self means death (or nonexistence of the future generations) for it or its components. On the other hand, destruction of a pencil is not an issue of life and death. Second, the "active participation" of selves is for "keeping their distinction from the others." This sometimes, but not always, involves preserving their molecular integrity. For example, a self may sacrifice some of its parts to provide an increased chance of survival for the other parts of it. Countries may sacrifice soldiers and living organisms may sacrifice some of their cells. Furthermore, the boundaries of the distinctions are not fixed for the selves. The boundary of a pencil can be said to be the outer layer of atoms and is rather fixed, but for a self it can not only change in geometrical form, but can also change with respect to the discriminations that it makes. Membrane of a cell can change in shape and composition and also the passage of molecules that is allowed by the membrane changes according to the needs[5] of the cell.

Finally I suggest a broad definition of a self as follows: A self is either a living being or is composed of living beings. It is an individual among others and it actively dissociates and discriminates itself from the other beings in order to have an

[5] The needs are the increased chance of survival and reproduction which are defined by the evolution.

increased chance of survival and reproduction of itself or its components.[6] The reason for me to use such a broad description of the self at the beginning is to start with a self concept that has some evolutionary significance. Furthermore, the above specification does not necessitate consciousness for now. But I will show that for some organisms consciousness will have to be involved in this discussion in the following sections.

2.2. The living selves

Since our aim is to analyze the concept of conscious self, let us focus our attention to selves that *are* living beings where the conscious selves seem to exist. Labeling all of the living beings as selves may seem trivial. But, as mentioned in the introduction, our problem is not an ontological one. Saying that to be a living thing is to be a self is ontologically trivial but epistemologically it is not trivial. We can know that all living things are selves but still we do not know the particular mechanisms that they achieve to effectively dissociate and discriminate themselves from the others.

For unicellular organisms determining their way of being a self is relatively easy. They all possess a boundary between themselves and the outside world: the cellular membrane. With it, they dissociate themselves from the outside world. The membrane acts as a boundary that separates the living cell from the ever changing conditions of the environment and keeps the conditions inside the cell in a range that allows the continuation of the cell's life. Here we can call the cell a self and its membrane is the boundary of the self. Note that membranes are not impenetrable walls. They allow the passage to the some of the materials in and out with the help of genetically determined mechanisms.

For the multicellular organisms, determining their way of being a self is much more difficult. Single cellular to multicellular transition should involve a change from being one-as-a-self (a single-cellular organism) to group-as-a-self (a multicellular organism) (Llinás, 2000, p. 75). To make that change, evolution had to solve[7]

[6] Llinás (2000) also calls the whole living organisms selves but he does not give an explicit notion of self like which is described here.

[7] By saying that "evolution had to do something," I do not mean that evolution is a goal directed process. This usage of language is the result of the backward thinking that I applied to better understand the concept of self. A more proper way of saying that would

three basic problems. One is about discrimination, the second about dissociation and the last one is about communication. Certainly solutions of these problems are not necessary for the appearance of the first multicellular organisms or the first ones would not be favored by evolution. But, solution of these problems will grant the organisms a distinct evolutionary advantage over the others.

The first problem is solved with what we know as the immune system. Since the organism is made of groups of cells, a physical boundary would not work as effectively as it did in a unicellular organism. So the way of discrimination is changed to a more subtle one. The organism detects and eliminates outsiders (those do not belong to the self) by using the agents of the immune system.

The second part of the solution is the skin. With it the organism can have a basic dissociating barrier between itself and the outside world. But skin is more than just a barrier. It is flexible and can respond to the needs of the body. When a part of the body is damaged or a limb is cut-off, the damaged region is covered by the skin.

The last problem is the problem of intercellular communication. Without a proper way of communication, the evolutionary favorability of multicellular life is quite limited. This effect can be seen in the evolutionary history. Llinás (2000, p. 74) states that after the appearance of the first single-cellular eukaryotic[8] life forms, it took 2 billion years for the first forms of multicellular life to appear. Following that, after the appearance of the first animal, the formation of the whole animal kingdom took only 700 million years. The result is the nervous system that the members of the animal kingdom have. While there are other branches of the tree of evolution where nervous system is not used, in actively moving organisms that employ muscle cells, nervous system eventually emerged.

---

be: “for the life of multicellular organisms to be more favorable in the mechanics of evolution, mutations related to the solution of three problems of multicellular life had to occur.” But this kind of language would only overcomplicate the discussion.

[8] Eukaryotic organisms are which basically possess cells with distinct nucleus (DNA is encased in an intracellular membrane), organelles and cytoskeleton.

Llinás (2000) connects the evolution of the nervous system with the evolution of movement more specifically with the evolution of the muscle cells. He shows that the first neuron evolved as an interneuron between two muscle cells (Llinás, 2000, pp. 78-81). Additionally he gives the example of a sea squirt (*Ascidiaceae*) which, in its life cycle, goes through two stages. In the first stage, the animal is a "free-swimming" larva which is "equipped with a brain-like ganglion" with approximately 300 nerve cells (Llinás, 2000, p. 15). With its primitive nervous system, it can swim through water with the help of its "life sensitive patch of skin", a balance organ, and a primitive spinal cord. When it finds a suitable place, it buries its head into the selected location and passes to the second stage of its life cycle. In this stage it continues its life bound to the location it had chosen. It filters the water passing by for nutrients and it also digests most of its nervous system! The only remaining part of the nervous system is what is required for the "simple filtering activity" (Llinás, 2000, p. 17). The sea squirt needs its nervous system as long as it actively moves through the water and as soon as it gives up the ability to move, it also gives up its nervous system.

What can be the reason for such a strong connection between having a nervous system and having the ability of active movement? Llinás (2000) claims that "prediction is the ultimate function of the brain" and he defines prediction as the "forecast of the future events" (p.21). If you have the ability to move without the ability of prediction, your actions are at best futile if not hazardous to you. You can bang your head on a hard surface if you do not know where you are going to. This reasoning also applies to much simpler animals. If the sea squirt had not predicted the results of its actions, its journey could have easily ended in the mouth of a predator.

In addition to prediction, coordination is also one of the most important functions of the nervous system. Movement requires a smooth coordination of muscle activation. Groups of muscle cells should be activated in a synchrony. Without coordination and synchrony, muscle activation can only result in a tremor. But how does the nervous system predict the future and coordinate the movements of the body? To perform these activities the nervous system must act like a self. In the next section we will see the how and why the nervous system behaves like a self and discriminates information.

2.3. The nervous system

While the nervous system cannot be dissociated from the living selves (or living organisms) that they belong to, their nature deserves much more attention for our purposes. In the nervous system, the way of discrimination and dissociation is more at the information processing level than ever before. We have seen the immune system that employs information to discriminate and eliminate objects that does not belong to the self. But, the nervous system goes one step forward. It should use information to discriminate information. I will build my argumentation step by step.

First, the nervous system needs to discriminate what belongs to the body and what does not. This is important for keeping the distance between the harmful objects and vulnerable body parts as well as for effectively controlling the body. Furthermore, the discrimination should not be a static one either. It should be able to be shaped according to the needs of the body. Just like skin, new discriminations should be made as the shape of a body part changes. In sum, the nervous system should form the information of what belongs to the organism and what does not. I will call this attribute *the ownership*.

Second, the sensory inputs coming from all sensory organs should be taken into consideration and a priority decision should be made before selecting the next action. The sensory inputs should be somehow unified and their importance should be judged. After that, the actions should be commanded by one center to prevent any conflicts. I will call this attribute *unity.* In fact unity is an important aspect of consciousness and it is a topic of ongoing discussion in the field of neuroscience (Section 3.2.3).

Third, I argue that, in order to predict the outcomes of its actions and possible results of the outside events, the nervous system should either calculate possible outcome of the current event or recall the outcome of a similar event in the past. Calculations can be made for slowly occurring events. But for the fast events, calculation of the outcome may not be fast enough for the preparation of an appropriate response. Memorizing the outcomes of previous events provides a good solution to this problem. If the nervous system memorizes an event, it can use the memory to predict the outcome of a similar event. By using the prediction,

the nervous system can determine the next action accordingly. In making the decision, the nervous system should be able to use the appropriate memory. To do that the sensory inputs about the current situation should be able to be matched the old memory. So, it is important for the sensory inputs and the old memories to have a similar coding. Without a similar coding they cannot be compared. So the memories should be able to be added to the *unity*.

Fourth, the nervous system should dissociate the results of its stimulation that it has given to the body from the results of the external events. For example, birds cannot fly if they flap their wings in the same way when the wind blows in the different directions with different strengths. The nervous system of a bird should be able to distinguish how much its muscles affect the shape of its wings and how much the wind affects it. The results of internal stimulation should be stripped of the external effects before they are stored in the memory. Otherwise an objective judgment cannot be made when the memory is recalled in different conditions. Also this dissociation is useful if we think of the simultaneous activation of many muscle fibers. It is important for the nervous system to determine whether a group of muscle fibers are contracted due to their activation or due to the activation of the fibers around them. If the latter is the case, next time, activation would also be sent to the inactive fibers to make them cooperate. Furthermore, the field of gravity presents a constantly changing force-field according to the orientation of the body with respect to the field. The nervous system should be able to discriminate the effects of gravity in order to have the body move efficiently in different situations. Lastly, we all know that muscles can grow tired or can become strong. So the nervous system has to check every time how much stimulation causes how much action. As a result, we can see that it is important to discriminate the effects of the stimulation given by the nervous system from the other effects. I will call this attribute *agent discrimination*.

So far we have discussed the meaning of the self in general. We have seen the ways by which the selves maintain their distinctions from the others and why do they do that. Furthermore we have looked into the realm of living selves. We have seen how a single-cellular organism behaves like a self and why transition from the one-as-a-self to group-as-a self was so difficult. Next we looked at the nervous system, which is one of the solutions to the problems of being a group-as-a-self.

Finally we have seen how and why the nervous system should behave like a self and we saw the main attributes such a system should have. In the next chapter we will see two main approaches to the concept of conscious self. In the fourth chapter we will see the integration of the nervous system and conscious selves that we think we are.

CHAPTER 3

THEORIES OF THE CONSCIOUS SELF

Among the many theories on the nature of the conscious self there are two philosophical approaches that encapsulate all the others: One is “the ego theory” and the other is the “bundle theory.”

3.1. The ego theory

The ego theory has its origins from our common-sense perceptions. In our daily life we tend to think that there is an “I” who is the *subject* of the experiences. This theory suggests that all conscious experiences occur to a self (ego, soul or homunculus) which is also the *agent* of the actions.

3.1.1. Cartesian dualism

There are many religions that support this common-sense theory and we can see its roots in the philosophical literature in the writings of René Descartes. In his famous thought experiment Descartes doubts his knowledge about his senses and his thoughts and tries to find what he cannot doubt (Descartes, 1969). Think of yourself. You are reading this thesis but you cannot be sure of it. Your perceptions might be failing you. According to him, you can doubt the existence of this thesis, of body, even of your abstract thoughts. The one thing that you cannot doubt is the existence of yourself as the thinking entity. No matter what you are thinking of the thinker is you.

After making the thought experiment, he concluded that he can doubt anything, even his physical body, but not himself as a thinking being. So he divided himself into two parts: one is immune to doubt and the other is doubtable. The part that is

immune to doubt he called *res cogitans*, the thinking part of him, and the other is *res extensa*, the physical extension of him or his body (Descartes, 1969).

Such a view supports our common-sense reasoning about the nature of the self. It proposes an unearthly mind (*res cogitans*) that thinks, remembers, senses and acts. But there is a grave problem in this approach: the problem of interaction. If there is an immaterial mind that is distinct from the physical body, then how can it interact with the body? And if there *is* an interaction, how one can claim that mind is not physical? These questions remain unanswered since the time of Descartes. Descartes himself was aware of the problem and he tried to answer it by pointing to possible location of the interaction in the brain. But the problem is not about the location of the interaction. It is about how that interaction is possible if mind is an immaterial entity (Moody, 1993). Without a reasonable solution to the problem of interaction, which seems impossible, dualist theories will have not much success in explaining mind.

3.1.2. Dualism of Eccles

John Eccles and Karl Popper (1977) propose a solution to the problem of interaction in the framework of their dualist-interactionist theories. Within the theory, they separate the reality essentially into two parts. One part (World 1) contains physical states and events and the other (World 2) contains mental states and events. In addition to that, World 2 is divided into three parts. The first is the outer sense which includes the perceptions about the outside world (sight, hearing, etc.). Second is the inner sense that includes the inner perceptions (thoughts, feelings etc.). The third part is what they call "…the self or the ego that is the basis of the personal identity and continuity that each of us experiences through our lifetime…" (Popper & Eccles, 1977, p. 360). In their view, each of the parts of the World 2 interacts with each other and also each of them interacts with a special part of World 1: the "liaison brain". They hypothesize that the liaison brain is composed of specific areas of the brain which are distributed across the cerebral cortex.

Later, Eccles clarifies the specific means and places of the interaction between World 1 and World 2 (Eccles, 1989). He exploits the quantum uncertainty[9] involved in the release of synaptic vesicles and claims that, at some specific set of neurons, the interaction occurs through the manipulation of the probabilities of the release of the synaptic vesicles. Furthermore, he defines the mind as a quantum probability field which is of neither matter nor energy. He claims that this field "scans" and "probes" a large number of synaptic sites across the cortex. This "scanning" and "probing" action, he says, does not contradict with the law of conservation of energy because energy can be borrowed by the synaptic site and paid back "at once" (Eccles, 1989, p. 189-191).

Henry Stapp (2004) raises two objections against Eccles' theory. First, Stapp questions the necessity of having a "knower" which can interpret the neural signaling of enormous complexity. He claims that having such two (the brain and the self as Eccles describes) information processing mechanisms involves an "uneconomical redundancy in nature" (p. 36). Stapp's second argument is against another aspect of the dualistic self. He argues that if Eccles' claims were true, then the patients with *neglect syndrome*[10] would not reject the *ownership*[11] of some of their body parts, because they would "know" that the body parts are belonging to them. If there were a soul or a self which resides in a mental realm, it should not be concerned with the damage dealt upon some brain tissue. Even if sensory inputs do not come from the organ, the sight of the attached organ should be enough for a soul or a knower to ascribe ownership to that organ. A patient should not be "puzzled" with the sight of an arm being attached to his body, but instead he should acknowledge it as his own when he sees it (Stapp, 2004, p. 166-167).

[9] Heisenberg's uncertainty principle states that the product of the uncertainties involved in some of the properties of a particle must be greater than a constant number. For example position and momentum of a particle is a couple of such properties.

[10] Patients with neglect syndrome reject the ownership of the neglected body parts. Often they attribute the ownership of their limbs to someone else (See Ramachandran & Blakeslee, 1998).

[11] Note that ownership is an essential property of the nervous system (Section 2.3). Also note that ownership that is described there is flexible and might change according to the circumstances and of course nerve damage can impair the ownership resulting in neglect syndrome.

Besides Stapp's remarks, some aspects of Eccles' theory are contradictory among themselves. Eccles suggests that all mental states and events are in World 2 which is a probability field of neither energy, nor matter (Eccles, 1989, p. 189-191). But when he tries to explain the means of interaction, he assumes an energy interaction between World 1 and World 2. Even if the energy supplied from World 2 was instantly taken back, this notion contradicts with the idea that World 2 being a probability field of neither energy, nor matter. If World 2 can supply energy (even if for an instant) then it means that World 2 is of energy. In sum, Eccles fails to solve the problem of interaction.

3.1.3. Physicalist ego theories

Another ego-theoretic approach can be a physicalist one. Scientists may point to some part of the brain or some brain processes, and try to come up with the neural correlates of the conscious self. For a complete theory of the conscious self, however, pointing to a part of the brain or to activation of a group of neurons is not enough. Saying that some pack of neurons is where the self is, is only a little more explanatory than saying that the self belongs to our brain. An ego-theoretic approach must explain how a unified self possible as a subject and an agent in the distributed system of neurons.

3.2. The bundle theory

The bundle theory of Hume basically claims that there is no conscious self in the ego-theoretic sense (as a subject to which the experiences occur). According to this view, our mind is just a bundle of experiences and the composition this bundle forms the self (Hume, 2000, p. 399).

Hume has two main arguments for his bundle theory. First, he claims that by using introspection we can only reach our thoughts, feelings and experiences but we cannot find a subject of those experiences. We cannot come across a self by using introspection. He also claims that thinking and feeling (conscious) self cannot be deduced from the occurrence of thoughts and feelings. He states that mind is just bundle of conscious experiences and as a thunderstorm does not need a subject, the collection of the mental states and events does not need a subject (ego-theoretic self) to occur (Hume, 2000, pp. 164, 165).

Like Hume, many contemporary neuroscientists claim that the mind is composed of a bundle. But this time they are more precise and materialistic (physicalistic) in their claims: Since our brain is composed of neurons, which are necessary for any intellectual ability that humans possess as shown by numerous neurological evidences, the conscious mind must also be generated by them and their interactions. At this point the approaches differ from each other. Some people claim that the conscious self is the conscious unity (constituted via synchronous neural activity) that is created by our nervous system (Llinás, 2000 p. 127, 128) while others refer to the (conscious) self by associating it with the personality in psychological terms and pointing some personality related brain areas (Damasio, 1999) while some does the both saying that personality and the conscious unity constitutes the self and disorders in either breaks down the self (LeDoux, 2002, pp. 301-324). But still no explicit description of the (conscious) self is present in these works. I argue that the main reason for that is: First, the self (or conscious self) is an ill defined concept that stems from folk-psychology (there is no explicit notion of it). Second, many neuroscientists fail to recognize their assumption that mind is formed by (interactions of) a bundle of neurons and continue to search for an ego-theoretic self.

The bundle theory is supported by the lack of regions in the brain that seem to support the role of a command center which we can call the conscious self. Even if a specific area of the brain is found out to be supporting such a role, it will still be composed of a pack of neurons and it will still be a bundle. So the bundle theory again prevails in the field of neuroscience.

There are objections to Hume's bundle theory of mind and the contemporary physicalist theories. One of the objections is concerned with the scientific inexplicability of the subjective states from a materialist viewpoint while the others are directed to the bundle theory itself both in its Humean version and the contemporary ones.

3.2.1. Objection of scientific plausibility

As opposed to the dualist view, materialists claim that everything can be explained in materialistic (or physicalist) terms. They say that if one keeps in mind the success of materialism in explaining many seemingly mysterious phenomena,

then one can see that there is no need to suppose the existence of another reality for the mental. But the materialists have a serious problem in this case, which is the phenomenal character of experience or the *qualia*. For example, one can describe everything that goes on during the process of color perception scientifically but the raw experience of the color. There is nothing red in the physical world. We can talk about reflectance of the surfaces or the wavelength of the light in physical terms but they don't describe the human experience of a color. This is the problem of subjectivity. Scientific research requires an objective approach to events. Since all the phenomenal experience is subjective, it is not possible to explain it scientifically (Nagel, 1974).

The above claims are likely to end the case for materialism if the misconceptions involved in them are not understood. Searle (1998) answers the above claims by making a distinction between epistemic and ontological subjectivity. According to him, saying that the Second World War started in 1938 is epistemically objective, whereas if you say that Hitler was more handsome than Churchill, you are stating your opinion about the subject. By doing that, you make a claim, which is epistemically subjective. Everyone may have a different personal stance about the topic, which has no scientific value unless you are not collecting statistics about people's opinions. However, if you express your observation like "There is a book on the table," it is observable by everyone, which makes the claim ontologically objective. If you are saying that you have a headache or that you have a desire, you are making an ontologically subjective claim which is not observable by anyone else but still it has ontological significance. The difference is in the nature of the observed thing, and if you say that you have a headache, it is not an opinion but a fact. As a result, scientific research on phenomenal experiences of the human mind is epistemically possible even if the subject itself is an ontologically subjective one.

3.2.2. The objection of floating mental states

As an objection against Hume's bundle theory, Carruthers comes with the following argument: "If the mind is merely a bundle of states and events, then it must be logically possible for the various elements of the bundle to exist on their own" (Carruthers, 1985, p. 52).

This is rather a weak argument if you are a physicalist. You can say that floating conscious states are not possible because they need a substrate to exist, which is the proper arrangement of active neural tissue.

Another argument can directly target the physicalist viewpoint: If experiences are generated by a bundle of neurons, then it should be possible for a specific bundle of neurons to generate a single experience but nothing else. It seems to me that, such a thing is possible and we can see similar examples in the split brain cases (Section 4.2.2). In addition to that, even if we imagine a disconnected neural mass which experiences only a single experience, say redness, it will certainly not be in the condition to speak about it. It will be devoid of the facilities of communication. So, we would not be aware of such experiences (until an objective method for determining experiences is found) if they occur in a disconnected neural tissue because it would not be able to express the feelings by using language.

3.2.3. The binding problem

The binding problem is the main problem for both Hume's bundle theory and the physicalist bundle theory. If our conscious-self is a bundle of experiences (conscious mental states and events), how can different experiences form a bundle so that we can have many experiences simultaneously? How can I see the text that I am writing on the screen, simultaneously hear the cars passing outside and feel the cold concrete under my feet?

Hume was aware of the problem. Hume (2000, pp. 170, 171) proposed that the resemblance between the experiences and the causal relationships of them bind the bundle together. He argues that current contents of one's mind resemble the past contents and are caused by the past contents. But his arguments would only explain binding over time. His arguments do not provide an explanation for instantaneous binding of different thoughts or experiences. Carruthers (1985) gives the example of a sound of Beethoven sonata and pain caused by an ingrowing toenail and he says that "There is obviously not the slightest resemblance between pain and the sound of the sonata. Nor there is any causal relationship between them. On the contrary, both of them are caused by external physical events…" (p. 55).

On the surface Carruthers' objections seem to be right. But I propose defenses three against these objections. First, there is a resemblance between the events: both are conscious events and both happen in the brain (or with the activation of the nervous system). Second, even though external events seem to be their initial causes, they both happen in the brain and they could have happened without external events being their initial causes. People can feel pain in their phantom limbs (Ramachandran & Blakeslee, 1998, pp. 39-62) and people can hear sounds which don't have an external source (Stephens & Graham, 2000). Third, relevance between conscious events is a requirement. Comparison between the conscious events needs to be done if a priority decision like allocation of attention is going to be made. If there can be a comparison, then there should be relevance. Without relevance you cannot make a priority decision between a sight you see and a sound you hear. These three arguments show us the possibility of relevance and causal relationships between conscious events.

The binding problem is more difficult for the physicalist approach: If our nervous system generates our experiences, then how is even a single experience generated from the interaction or firing of many neurons?

An answer to the above questions must state a principle for the unification of the experiences and binding of neural firing. Unification of the experiences is an important problem in the field of visual cognition. Treisman (1996) states that at least seven types of binding are required in order to identify a visual object. The most striking one is the property binding. The problem arises because different visual properties (color, shape and movement) related to objects are processed by different areas in the brain. In order to construct a coherent object, the different aspects of the visual information must be bound together.

Binding by synchronous activity of neurons is a proposed solution (Llinás, 2000, Treisman, 1996). According to this view, the perceptual unity is achieved by 40Hz synchronous neuronal activity in the visual cortex. Treisman (1996) gives the example of the thalamocortical (between thalamus and cortex) and cortico-cortical (between different areas of the cortex) synchronous activity measured in experiments regarding visual perception in cats (Gray et al., cited in Treisman, 1996, p. 174). In the experiment, it has been found that "units with spatially

separate receptive field fire synchronously in response to a single object, but not in response to two different moving or two separately oriented objects" (Treisman, 1996, p. 174).

Such a result can be thought to be an evidence for the role of synchronous firing to obtain perceptual unity for a single object, but in this perspective perceptual binding of the whole visual field is impossible. A similar 40Hz synchronous neural firing has not been observed in the subject's brains for two different objects which are presented to the visual field of them. In addition to that, Zeki and Bartels (as cited in Viviani & Aymoz, 2001, p. 2917) suggest that "when two attributes (e.g. color and orientation) are presented simultaneously, they will be perceived at *different times* if the percepts are created by the activity of the cells at different sites. Conversely, they will be perceived at the *same time* if the percepts are created by the activity of the cells at the *same site* …. Consciousness is not the consequence of binding the activities of cells at different sites; rather it is the micro-consciousness (generated by each specialized network) that are generated at *different sites* that require binding." So, there is an asynchronous neural activity for the presentation of spatially separated visual attributes. But how is the unity and binding of the experiences that we perceive possible for different sensory modalities in the face of the temporal asynchrony and spatial separation of nervous activity? We will see a possible answer in the next chapter.

3.2.4. Particularity of the conscious states

Carruthers (1985, p. 57) puts his final objection to Humean bundle theory with respect to the particularity of the conscious states. He asks, if the same experience is shared by two minds, does that make two minds one. He gives the example of Siamese twins who are joined back to back in birth. If they have an exactly similar pain in their back, he asks, would that suppose a single mind or two different minds in the Humean sense?

The answer is easy if you are a physicalist. It is known that pain is not generated at the body parts, but related to *body image* in our brain (e.g. phantom limbs (Ramachandran & Blakeslee, 1998)). So, even if some nerve fibers carry the information to one brain and some to others, there will be two pains in the two brains of the Siamese twins. In addition to that, if two minds (two brains) are

connected somehow, then we can claim that the experiences constituted in the two brains can be coupled via neural coupling and there will be only one unitary mind. We can see such an event in every ordinary human being because they carry two brains in their skulls (one left and one right hemisphere) and coupling of these two give rise to a single mind. We will see the cases of split brain patients in the fourth chapter. The presentation of those cases will make the points above more clear.

CHAPTER 4

UNITY OF CONSCIOUSNESS AND CONSCIOUS SELF

We have seen the need for a being a self for a biological organism. In order to be a group-as-a-self, we have discussed the need for communication. In addition to that, we have explored the conscious selves which we intuitively correlate with an ego which is the subject of the experiences and the *free agent* of the actions. We have seen the problems with the dualist arguments and we have seen that there is no place for an ego-theoretic self in the nervous system. As an alternative theory, we have seen Hume's bundle theory and its modern, physicalist version. Among the other problems, the binding problem has been the strongest objection to any kind of bundle theory. The nervous activation shows neither spatial nor temporal unity that could give rise to the coherent, unified experience that we have. The nervous system processes information in spatially separated regions and only local synchrony in nervous activation has been observed so far (Treisman, 1996). So how the binding problem can be solved? In this chapter first we will see how quantum mechanics plays a role in the solution of the binding problem and then discuss how various aspects of the nervous system that are described in Section 2.3 can be integrated into the bundle theory.

4.1. Binding with quantum coherence

Quantum mechanics shows us the existence of a proper substrate for conscious experiences. John (2001) claims that activity of discrete connectionist networks cannot explain the unity of consciousness. Stapp (2004) assumes a more strong position and argues that classical physics cannot provide an explanation for the

unity of consciousness. It is suggested that quantum mechanics can be used explain consciousness which can also account for the effects of general anesthetics which are thought to prevent *quantum superposition* by preventing electron mobility in the hydrophobic pockets of some selected proteins(John, 2001; Hameroff, 1998). In the following sections we will see how quantum events are related to the brain and in what way quantum mechanics offers a substrate for conscious unity.

4.1.1. Quantum theory and its interpretations

Quantum theory is hard to grasp because its predictions and findings come up with issues that are not parallel with our intuitions about the structure of the world. It is best to start with the results of a two slit diffraction experiment to better understand the structure of a quantum world. I will try to express the experiment and its results as objectively as possible so that we can discuss the different interpretations of quantum theory.

Two slit diffraction experiments are well known cases which are performed by preparing an experimental setup which sends a particle (it can be a photon, an electron or an atom) from a particle gun towards a wall with two small slits in it. The sent particle is then detected by a measurement device (let's say a photosensitive paper) that is located behind the wall. If many photons are sent towards the slits, then they form a diffraction pattern on the photosensitive paper. This pattern is similar to the water waves passing through a barrier with two slits. Furthermore, the shape of such a pattern can be predicted by Schrödinger's wave equation. But if, only a single particle is sent towards the wall with two slits, then its position can only be determined with the measurements done by the experimenter. There is no other way to determine the location of the particle on the photosensitive paper. Quantum formalism (specifically the solution of the Schrödinger's Equation) can only offer a set of probabilities regarding to the position of the sent particle.

The problem here is what happens to the particle during its travel? If it can interact with the both of the slits (suggested by Schrödinger's equation), then how can it be detected only at one point of the detector? What is its trajectory? It is difficult to make any ontological claim about the state of the particle in its travel. This

difficulty, as Schwartz et al. (Schwartz, Stapp & Beauregard, 2004) point out, prevented the founders of quantum mechanics from making any ontological claim about their findings. Thus, Heisenberg, Pauli and Bohr proposed an epistemological interpretation which is known as the Copenhagen interpretation. According to this interpretation the quantum theory is all about knowledge of human subjects and mathematical rules for manipulating that knowledge. Copenhagen interpretation refrains from proposing any ontological argument about the underlying reality of the quantum phenomena and it is only concerned with what is apparent to us. Words of Heisenberg reflect this attitude:

> The conception of the objective reality of the elementary particles has thus evaporated not into the cloud of some obscure new reality concept, but into the transparent clarity of a mathematics that represents no longer the behavior of the particle but rather our knowledge of this behavior. (Heisenberg, cited in Schwartz et al., 2004, p. 8)

Stapp (2004) and Schwartz et al. (2004) accepts a more radical form of Copenhagen interpretation which is first formulated by von Neumann. This epistemological interpretation is exclusively ego-theoretic and uses the assumption of human agency (Process 1) as a fundamental principle. But they left the question of "How a mysterious Process 1 can act over the *cloud of probabilities*[12] formed in the brain?" completely unanswered. So, I will not go in the details of this approach.

Erwin Schrödinger showed the necessity of making ontological claims and the absurdity of the Copenhagen interpretation (or any epistemological interpretation) by his famous thought experiment: "cat in the box" (cited in Hamerhoff et. al., 1996, p. 435). In this thought experiment the cat's life is dependent upon the behavior a quantum particle fate of which is not 'known' by anyone. If no ontological claims are made about the particle, then the cat is considered both dead and alive until a human observer opens the box. So a number of ontological interpretations are proposed to overcome this absurdity.

[12] This cloud of probabilities is said to be formed by the uncertainties involved in the release of synaptic vesicles in the nerve cells.

Before progressing further to the ontological interpretations of the quantum theory, I would like to introduce well known concepts of *quantum superposition* and *wave function collapse*. According to quantum formalism, any observable property (i.e. mass, position) of a quantum system can be described by a single wave function which is a superposition of a number of states (*eigenstates*) that are allowed by the constraints that act upon the system. Solution of the Schrödinger's equation gives the relative probabilities of detecting the quantum system in any one of those states. If the system has two eigenstates each having a probability 0.5 of detection, then in half of the experiments done on the system one state is detected, while in the other half, the other state will be detected. The quantum system prior to detection is said to be in *quantum superposition*. The quantum particle in the Schrödinger's thought experiment is in quantum superposition with two probable outcomes. The transition between from the superposed state to one of the eigenstates is called the *wave function collapse*. Most of the debate about the quantum ontology is on the nature of the quantum superposition and on the conditions under which wave function collapse occurs.

There are three interpretations of quantum theory which have ontological claims. The first one is the Everett's (cited in Barett, 2003) *many worlds interpretation*. This interpretation accepts quantum superposition, but rejects wave function collapse. According to many worlds interpretation, every possibility dictated by the Schrödinger's wave function actually happens however, by creating different worlds. So, in the two slit diffraction experiment the particle strikes every possible location but, we cannot observe the other possibilities because they happen in another parallel world. This interpretation involves huge metaphysical assumptions (infinite number of worlds created by each quantum event) which are hard to deal with. Due to this reason alone I will not discuss it in the rest of this work.

David Bohm's (1981) interpretation has another assumption which is called the *pilot wave*. Bohm suggests that particles in fact follow continuous (classical) trajectories but their possible trajectories are determined by a field of guiding pilot waves. These pilot waves act as a field of potential[13]. Predictions of this

[13] A field of potential can be imagined as a surface that a ball rolls on freely under the influence of gravity. In this case the ball tends to roll along where the steepest descent is possible.

interpretation accords with the findings of quantum mechanics because the possible trajectories formed by the pilot waves are calculated by the Schrödinger's wave equation. Furthermore, nonlocal interactions are possible between the pilot waves which also accords with the *quantum nonlocality*[14]. In this interpretation quantum superposition does not exist for particles themselves. It only exists in the form of the superposition of the pilot waves. So, this interpretation does not have to tackle with the ontology of the wave function collapse. In Chapter 5 I will point out how can this interpretation be used to account for consciousness, but before that we will examine the interpretation of Hameroff and Penrose (1996). Since their theory provides a detailed solution to the problems of consciousness we will examine their theory more deeply.

4.1.2.Penrose-Hameroff interpretation and quantum consciousness

Penrose (1994) proposes a quantum gravity solution to the problem of wave function collapse to overcome the difficulties associated with the Copenhagen interpretation. In his interpretation he assumes that particles do exist at many forms simultaneously when they are in a superposed state. Furthermore, he characterizes the superposed states of the quantum systems as unstable and measures the stability of the system by the gravitational difference between the superposed states of it. Following that, he proposes a rate of reduction (collapse of the wave function) which is calculated by considering the quantum gravity effects on the system. Penrose calls this process *objective reduction* (OR) because it does not require measurements to occur. In that respect this interpretation rejects the notion assumed by Copenhagen interpretation which refrains from making any claims about the reality that is independent of human knowledge.

Hameroff and Penrose formulate their theory about consciousness on the concept of *quantum coherence*. When certain conditions, as described by quantum mechanics, are met, particles become principally indistinguishable from each other and share a unity. They can be described by a single wave function and they are essentially unentangled with the environment (Penrose, 1994). These kinds of

[14] According to quantum no locality, certain particle pairs (entangled particles) can show nonlocal interactions which Einstein calls "spooky action at a distance" in his famous words.

substances are called *Bose-Einstein condensates*. [15] In many ways such a substance seems to be a substrate for the conscious unity that we seek. First, the particles that become entangled share an identity and behave like one large particle (Zohar, 1996). This property makes way for the explanation of the nonlocal unity of our experiences which are formed in a brain with spatially distributed neurons. Second, if the brain can utilize quantum superpositions, it can deal with many different sensory inputs at a time (just like consciousness) (Penrose, as cited in Zohar, 1996). Third, OR can be a decision making process which realizes the results of quantum computation (Hameroff & Penrose, 1996).

4.1.3. A place for quantum coherence

Large scale quantum coherence can be the physical substrate of consciousness. But, where and how can it occur in the nervous system? Among several candidates for a place for quantum coherence in the brain (e.g. DNA, cellular membrane and synapses), *cytoskeletal microtubules* seem to be the strongest one[16] (Hameroff & Penrose, 1996; Penrose, 1994; Woolf & Hameroff, 2001; Hameroff, Nip, Porter & Tuszynski, 2002).

Cytoskeletons of *eukaryotic cells*[17] are formed by interconnected networks of microtubules. Microtubules are tubular structures which are formed by tight, spiral-like arrangements of bean shaped protein structures which are called *tubulins*. A tubulin can assume two *conformational states* or a quantum superposition of those two states which is determined by the *localization* of the electrons in the *hydrophobic pocket* of the tubulin. In other words each tubulin can be in one of the two forms, one stretched and one closed, or it can assume a form in which the two forms coexist in a state of quantum superposition according to Hameroff and Penrose (1996).

---

[15] Examples of Bose-Einstein condensates include superconductors, superfluids and laser beams (Zohar, 1996).

[16] There are long discussions in Hameroff and Penrose (1996) and in Penrose (1994) about how and why the microtubules are considered to be the strongest candidates for the locations of quantum coherence. Also Woolf and Hameroff (2001) provide some additional up-to-date support and also provide more details on exact mechanisms for microtubular quantum coherence. Reciting those arguments here would overcomplicate the presentation, and moreover, they are not critical for our discussion.

[17] Eukaryotic cells are the cells which possess a nucleus and a network of microtubular cytoskeleton. All animal and plant cells fall into this category, and some single cells are also eukaryotic (e.g. euglena).

Hameroff and Penrose (1996) deduced two possible computational works that microtubules can perform by considering their structure. First, they can perform *classical computation*[18] where the information could be stored as the states of the tubulin dimers. Also groups of tubulins can provide binding sites for *microtubule associated proteins*[19] (MAPs) so that information can be hardened. Second, they claim that, microtubules can perform quantum computation via quantum coherence and *self-collapse of the wave function* in tubulins.

4.1.4. Hameroff-Penrose model of quantum consciousness

The next question is where to put consciousness in quantum computation. Hameroff and Penrose (1996) answer this question by proposing a sequential and cyclic model for quantum computation (Figure 1). In their model, they suppose that quantum coherence slowly builds up within microtubules by engaging more and more tubulins until the mass-energy difference of the involved tubulins integrated over time reaches a threshold. According to them, the reduction of quantum coherence can be orchestrated by the modulation of the microtubule associated proteins (MAPs) or by other tubulin modifications. Mediated by genetical and environmental factors, Hameroff and Penrose (1996) suggest that, these modifications can set the possible outcomes of the OR which results in an orchestrated objective reduction (Orch OR).

Hameroff and Penrose (1996) hypothesize that building up of quantum coherence stands for the preconscious processing and the collapse of the wave function stands for conscious experience (see Figure 1). Furthermore, they claim that the time required for the preconscious processing for *normal* experiences is correlated with Libet's work (as cited in Hameroff & Penrose, 1996) which states that there is a 500 msec delay between the direct electrical stimulation of the somatosensory cortex of awake human subjects and their reports of conscious experiences about the stimulation. Hameroff and Penrose also propose that if the number of tubulins

[18] According to Hameroff and Penrose (1996) classical computation in microtubules can be done by the self-organization of the conformational states of the tubulins.

[19] Microtubule associated proteins interconnect microtubules by forming radial links. Also they provide means for material transmission along the microtubules (Insinna, Zaborski, & Tuszynski, 1998). Furthermore, it is known that MAPs take role in synaptic plasticity and learning (see Hameroff & Penrose, 1996).

involved in the quantum computation increases, the quantum gravity threshold can be reached faster, which, they claim, causes more intense conscious experiences.

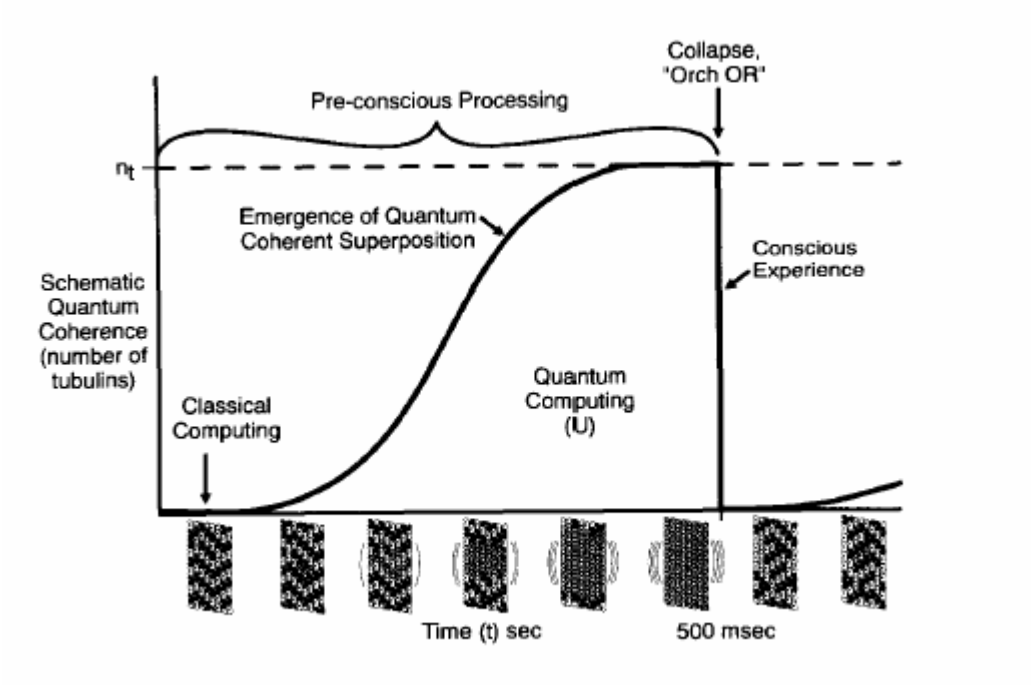

Figure 1 Schematic graph of the process of quantum computation in microtubules (Hameroff & Penrose, 1996, p. 531). Microtubules are depicted under the graph. Each unit in the microtubules stands for a tubulin dimer. The black and white tubulins are the ones that are in one of the two classical conformational states. Grey ones denote tubulins in a state of quantum superposition. In this model, microtubules are thought to perform classical computation in the absence of quantum superpositions. Quantum computation begins with the involvement of tubulins in superposed states. As the system reaches a quantum gravity threshold (area under the curve), it reduces into a classical state. 500 msec preconscious processing is required for conscious experiences according to Libet et al. (cited in Hameroff & Penrose, 1996, p. 529).

4.1.5.Difficulties associated with the Hameroff-Penrose model

Quantum processing might be the physical substrate of the conscious experiences. But the above interpretation by Hameroff and Penrose is problematic for several reasons:

- Hameroff and Penrose propose that conscious experiences occur in a discontinuous fashion. They say that quantum coherence ends with a wave function collapse and conscious experiences are correlated with the collapse of the wave function. It is a generally accepted view that wave function collapse is an instantaneous event. Even if we accept the notion that conscious experiences are built up by frames, we expect each frame to last more than an instant. You cannot construct a building by using

dimensionless frames. So, I argue that, frames of conscious experiences should at least last for a finite time to account for the apparent continuity of our consciousness. Thus wave function collapse is not suitable to associate with conscious experiences.

- In their view *all* the quantum processing is done preconsciously and consciousness comes into the picture when the wave function collapse occurs. By associating conscious experiences with the wave function collapse alone Hameroff and Penrose seem to disregard any causal role for consciousness. I suggest that their view is no different than the classic *epiphenomenalism*[20] and it is far from explaining consciousness and conscious experiences because they claim that conscious experiences occur *after* quantum computation.
- I argue that their correlation of the time required for the preconscious processing in their model with the delay observed in Libet's work can be misleading for two reasons: First, 500 msec of quantum computation is too long for any *normal* experiences. The results of quantum computations cannot be got until the quantum system collapses into a classical physical state. So, the whole 500 msec has to pass for the computation to end. This means that our brain processes information with time steps of 500 msec. But, performing computations with such long time steps would not be tolerable for active organisms like us who often has to predict the near-future. Second, it has been shown that the 500 msec delay observed in Libet's experiment might not be related with the preconscious processing but instead with the transient inhibition of the electrical stimulus by the activity of local inhibitory neurons for a range of electrical stimulus intensity (Pollen, 2004).
- There is an ongoing debate about how long large scale quantum coherence can endure within the human brain and this is the main line of attack for who criticize the theory of Hameroff and Penrose (Schwartz et al., 2004). Tegmark (2000) suggest that quantum coherence in brain cannot last longer than 10^{-13} sec due to the "hot, wet and noisy" environment of the brain. He further argues that this duration is

[20] Epiphenomenalism is philosophical view which suggests conscious events are just shadows or side effects of physical events in the brain which have no causal effects on these events.

incompatible with the minimum time-range of the activation of nerve cells (10^{-3} sec). While Hagan et al. (Hagan, Hameroff & Tuszynski, 2002) claim that microtubules can effectively be shielded from the environment by some mechanisms within the nerve cells, Schwartz et al. (2004) argue that 10 orders of magnitude is a large discrepancy to explain away.

- I argue that self collapse (Orch OR) of quantum coherence is not a necessity for quantum computation. For smaller nervous systems, the self-collapse of quantum coherence might not occur in time for it to be useful for any purpose. Instead of waiting it to collapse, in such systems, decoherence could be forced in a cyclic manner by outside interference. For example a timing system within the organism may remove the shielding around the quantum coherent microtubules occasionally and cause decoherence. By doing that the organism can get the results of quantum computation in time although prematurely. Thus, any time estimation which relies on the self collapse of quantum coherence can fail in biological systems.

4.1.6. Quantum consciousness and bundle theory

Although there are problems with the Hameroff-Penrose model, its basic principles can still provide strong support for the bundle theory for the following reasons:

- Hameroff and Penrose (1996) claim that there are places in the brain where quantum coherence and quantum computation can occur. The strongest candidate is microtubules. They also propose that microtubules can perform classical computation as well as act like a quantum processor in a cyclic manner.
- Neither conscious experiences (elements of our consciousness) nor the particles involved in quantum coherence can be observed from outside. I suggest that, this property of quantum coherence can account for the *subjectivity* of the conscious experiences (see Section 4.1.6).
- Large scale quantum coherence can bind the information that comes from the different parts of the brain.[21] Tubulins in separate parts of the brain can join quantum coherence and represent any sensory information associated

[21] This is the main reason for the need of a quantum mechanical account for explaining consciousness. Quantum coherence can provide the non-local unity that we seek for explaining unity of consciousness.

with them in a nonlocal manner. This aspect of quantum coherence is not only parallel to the conscious unity but it also solves the binding problem for the physicalist version of the bundle theory which is discussed in Section 3.2.3.

- I argue that the tubulins that are involved in quantum coherence might correspond to conscious experiences. For example involvement of the tubulins in the visual system of the brain can give rise visual experiences. In this perspective, when some of the tubulins involved in quantum coherence are removed, some conscious experiences will also be removed and when all are gone there will not be anything that we can call consciousness.

Considering the above remarks, I suggest that consciousness *is* a transient field of quantum coherence that is formed in our brain by utilization of microtubules in some specific set of neurons.

This notion of consciousness is similar to those who correlate consciousness directly to quantum coherence (e.g. Zohar,1996). But it is also fundamentally different from the account of Hameroff and Penrose (1996) because first, it does not assume Orch. OR and second, it does not associate conscious experiences with the wave function collapse. Instead, decoherence is thought to be a means for realization of the results of quantum computation and conscious experiences are considered to be generated by the microtubules that are involved in quantum coherence.

4.2. Other aspects of the self experience

We have seen that quantum coherence can be a solution to the binding problem. In this section we will see how other aspects of the self experience are explained with a bundle theoretic approach that relies on quantum coherence. These aspects include subjectivity, ownership, agency, personal identity and persons.

4.2.1. Subjectivity

Subjectivity is one of the most puzzling aspects of the self experience. The questions about it involve: (1) Why cannot our experiences be observed by other

people? (2) What is the difference between my experience and someone else's experience? (3) Can two people share exactly the same experience?

Question (1) can be answered by using the most basic property of quantum coherence: No one can see what is happening in a field of quantum coherence. For, environmental interference can cause decoherence. Only after the system collapses to a classical state one can observe the results of quantum computation. If our brains employ quantum computation, and if our consciousness derives from quantum coherence, I argue that subjectivity becomes a natural result.

The answer to the question (2) can again be given by using quantum coherence. The difference between the experiences of two different persons lies in the distinct fields of quantum coherence that they have.

The answer to the question (3) is a yes. If proper arrangements are made to connect brains of two people so that the two brains can form a unified field of quantum coherence they can share their experiences. The question (3) then can be asked in combination with question (2). Such a question was already asked by Carruthers (Section 3.2.4.) as an objection to Hume's bundle theory: What if two minds share an experience, would that make two minds one? Again, the answer is: Yes. In fact this is the case which we see in all normal human beings.

The brains of all normal human beings are composed of two hemispheres, one left and one right hemisphere. The hemispheres receive signals from the opposite sides of the body and they command the muscles at opposite sides of the body as well. In normal humans, the two hemispheres are connected with a bridge which is called corpus callosum. In the extreme cases of epilepsy, the nervous bridge of corpus callosum needs to be severed for the treatment of the epilepsy. Striking support for the above claims can be seen in the reports of the patients whose corpus callosum is severed. It seems that when corpus callosum is severed, two separate minds emerge. (Schiffer, 1998; Mark, 1994; Iacoboni, Rayman & Zaidel, 1994)

In the brains of the split-brain patients each hemisphere receives inputs from a separate half of the body and the conscious experiences generated by those

inputs cannot be shared since corpus callosum is severed. The same is true for the visual inputs presented in the left and the right visual field because inputs coming from each visual field are sent to different hemispheres. The inputs that come from the right visual field are sent to the left hemisphere and vice versa.

Such patients cannot name the objects that are presented in their left visual field because the left hemisphere, where the linguistic facilities are located, is not aware of the objects (Schiffer, 1998). Also, many patients report conflicts in the behavior of the left and the right sides of their body. Schiffer (1998) cites several cases where patients find themselves grappling with their *autonomous* left hands or find out their left legs want to go somewhere else. The actions of the *disconnected minor hemisphere*[22] are not only conflicting but they can also be very purposeful. In some cases the left hand of a patient can extinguish the cigarette which the right hand had lit. In others the left hand can put away a dress that the major hemisphere had chosen to wear and take another one. Perhaps the most striking example is about a split-brain patient who overslept and was finally awakened by (seemingly intentional) slaps across her face by her left hand! Schiffer reports that such cases occur just after the surgery that splits the two hemispheres and lasts until "…the hemispheres learn to get along with each other" (Schiffer, 1998, p. 30).

As I have shown in the above examples, two different minds emerge when the corpus callosum is severed. In the normal people the reverse is true. If two different minds can be connected in a special way (presumably one that allows large scale quantum coherence) then the result can be one, unified mind.

4.2.2. Agency, free will and ownership

Free will and agency are two other problematic issues which are associated with our concept of self. We feel that we are the sole cause and *the agent* of our actions. The laws of our society are based on this feeling/assumption and claim that every healthy person possesses free will. But, physical determinism rejects any kind of free will because according to the physical determinism all events are caused, and in a chain of the cause-effect relationships there is no place for free

[22] Disconnected minor hemisphere is the one that lacks linguistic facilities which are needed to interact with other persons. So, usually it is the right hemisphere.

will. Furthermore, Libet (1985) has found that a readiness potential (measured from the scalp) precedes spontaneous acts without preplanning (such as flexing a finger) by 550 msec. In the same experiment Libet also found out that moving a finger follows the *experienced wish* to move it by around 200 msec. So, it seems that the experienced wish comes after the onset of the readiness potential by 350 msec. In the face of these arguments can a causal role for consciousness be possible let alone the existence of agency and free will?

To solve the above problems, we need to first look into our desires. If we can determine the role of desires in our life, then we might figure out a solution.

In our daily life, we tend to perform acts to attain our *desires*. Desires can be genetically determined like desire to eat, desire for pleasure, desire for love or desire to relieve stress. Also some desires can be *derived*[23] from our genetically determined desires such as desire for money, desire for being popular or desire for weekend to come. For example we may have a desire for performing our jobs better which may derive from desire for money which may ultimately derive from desire to eat.

While desires for many things can derive from more ulterior desires, there is a limit for our desires. Imagine a young tennis player. She can desire to be a good tennis player, desire to win a match or she can desire to score an ace. But the role of desires ends there. Desiring alone cannot make her muscles activated in a specialized pattern (which Llinás (2000) calls *fixed action patterns* or FAPs) which results in a strike that can score an ace. Our tennis player has to work hard and must learn the required pattern of activation (FAP) by trial and error.

The need for *agent discrimination* in nervous systems enters into the picture at this point. In Section 2.3 I suggested that the nervous system must act like a self and discriminate the actions caused by the stimulations given by it from the actions caused by the external forces. In order to learn the FAP, she must use her agent discrimination capabilities. By using that she can fine tune her muscle activations and learn the FAP that she needs.

[23] Learning (or developing) new desires can be done by procedures like classical conditioning, but the exact procedures are not important for our discussion.

There is a generally accepted model of motor control (*forward model of motor control*)[24] that can account for agent discrimination (see Figure 2). In this model, external influences on performed actions can be detected by matching the predicted sensory inputs with the actual sensory inputs. So, if there is a sensory discrepancy showing that intended action did not occur, that information can be used to fine tune the FAP associated with the intended action. By this way each time more and more precise movements can be made.

Blakemore et al. (Blakemore, Oakley & Frith, 2003) state that in some pathologic cases (e.g. delusions of alien control, thought insertion) the *forward output model* cannot be formed (Blakemore et al., 2003), where the patients describe their thoughts, movements or speech is being controlled by alien forces (Mellors, cited in Blakemore et al., 2003; Stephens & Graham, 2003). In these cases we can see the incapability of agent discrimination and thus loss of the feeling of agency.

As specific brain areas, lateral cerebellar cortex of the cerebellum is thought to be the place where forward models are formed for movement control (Imamizu et al., 2000). Blakemore et al. (2003) confirmed the involvement of those areas (by using neuroimaging) and shown that both passive arm movements and *deluded passive movements*[25] of hypnotized subjects produce similar activity both in their cerebellum and their parietal cortex. Whereas, the activities measured in those regions were quite distinct from the two previous cases when an active arm movement is performed.

In the light of the above arguments, if we accept that a forward model of motor control is utilized for agent discrimination, it is easy to suggest that agency is felt *after* an action is performed. The reason for that is the predicted sensory feedback can only be compared with the actual sensory feedback *after* performing an action.

[24] For a review of this model and its extensions see Miall and Wolpert (1996) also Webb (2004) provides a survey about the history of it in explaining psychological data.

[25] Deluded passive movements are obtained from the hypnotized subjects by telling them their arms are going to be moved by a pulley system that is attached to their arms. During the experiment, the pulley system was not activated and the subjects started moving their arms while believing that the pulley system is moving them. The same pulley system was attached to the arms of the subjects when performing active movements or real passive movements.

Only after making such a comparison, movements of the body parts can be discriminated according to self-caused-movements and non-self-caused-movements. Some authors claim that these findings accompanied with findings of Libet (1985) provide strong support to *epiphenomenalism* (Oakley & Haggard, 2005). They argue that if agency is felt *after* performing an action and a wish is experienced only *after* a readiness potential detected in the scalp, then consciousness seem to have no causal role in performing the actions.

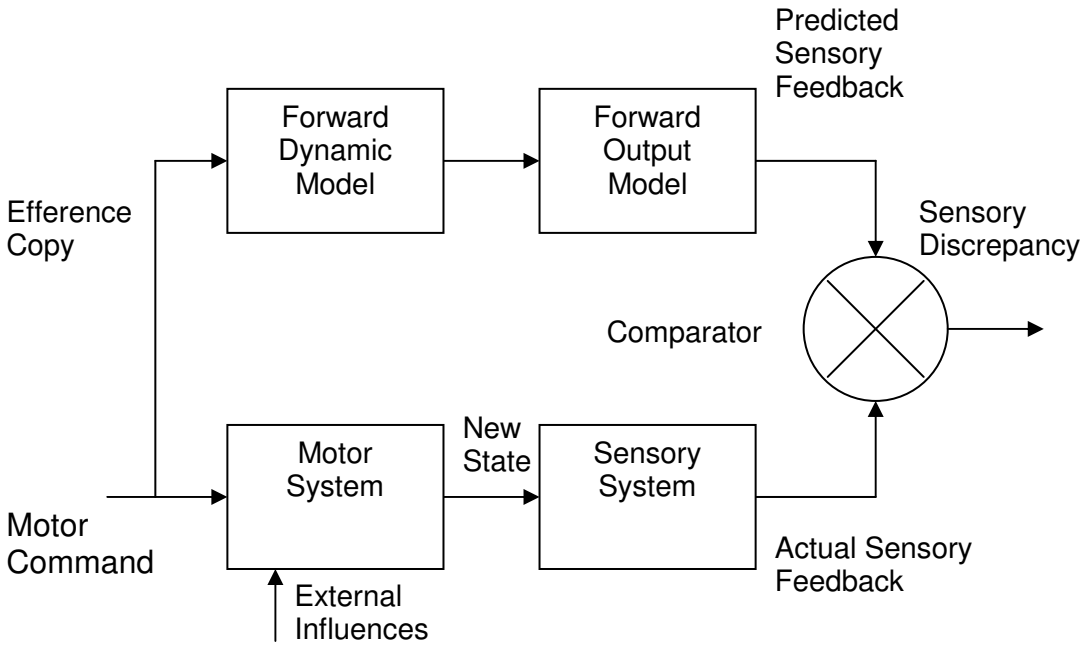

Figure 2 The forward model of motor control, as proposed by Miall et al. This model states that when a motor command is issued, a *forward dynamic model* of the motor command is also formed. Forward dynamic model predicts the consequences of the issued motor command. *Forward output model* predicts the sensory consequences of the motor commands. Then the predicted sensory inputs are compared with the real sensory inputs. In a normal movement, the real sensory feedback is expected to inhibit the predicted sensory feedback. A large discrepancy would mean that there were external influences while executing the motor commands (Miall et al., cited in Blakemore et al., 2003, p.1059).

Although an epiphenomenalist notion is appealing, I argue that another interpretation is possible. To put forth this account, let us go back to our discussion about desires. We saw that desires alone cannot make certain actions performed even when the actions are associated with our body alone. We can say that a desire is just a factor that is used while deciding on actions. Other factors may involve the situation that we are in, our memories, and our abilities. There are

situations in which we act against our desires. If a gun is pointed towards your head, you may do certain things that are against your desires. During a boring class session you may keep yourself in the classroom while having a strong desire to go out. Moreover, while performing mundane actions of daily life, you may not feel any desire at all.

If we go back to the Libet's experiments, by considering the forward model of motor control, we can ask this question: Do we *necessarily* feel an "experienced wish" *before* making a decision? The answer is: No. I argue that our conscious unity can decide on an action by using or not using desires as factors in decision making. Furthermore there is no need for it to generate a "wish experience" before making a decision. The only thing that our conscious unity needs to do is to *monitor* the outcomes of the actions that it had decided upon. So, what Libet (1985) calls *experienced wish* may in fact be something else. We have seen something that can correspond to the experienced wishes. That is the *forward output model*. After making a decision (this can correspond to the readiness potential), our nervous system generates a forward output model (this can correspond to the experienced wish), and then the action is performed. So, there is no problem about mental causation in this kind of situation. I argue that wish, which is different from desire, is experienced *after* the decision is made because it is experienced for making a check for agent discrimination.

Next, we come to the issue of free will. We have seen that the feeling of agency comes after our decisions. Furthermore we think that we live in a world governed by physical determinism, in which all events are parts of cause-effect chains. Can free will have a place in this picture? The answer is both yes and no. If we are asking whether there is any nondeterminism in our decisions, the answer is yes. If our brains can really make quantum computations, it means that our decisions are realized by a wave function collapse. This process is inherently nondeterministic (or noncomputable as Penrose (1994) prefers to say). So, a nondeterministic free will is possible. But, if we are asking for a notion of an absolute freedom of will, which is assumed by our laws and by almost all religions, the answer can be sought in the laws of quantum mechanics. Schrödinger's equation and its transformations are deterministic and computable. As a result, absolute free will is still an illusion if our brain is a quantum computer.

Lastly we come to the issue that how our nervous system makes the discriminations about the ownership of the body parts. Blakemore et al. (2003) claim that the forward model of motor control can be used for determining ownership. They say that by detecting the coincidences between body parts during movements our brain can determine what belong to us and what does not and form a body image. An example of such a procedure has been given by Ramachandran and Blakeslee (1998, pp. 60-62). They claim that by arranging some artificial sensory coincidences it is possible to extend one's body image of his/her nose to several feet. They also claim that in a similar manner it is possible to have one bestow ownership to a table! Note that ownership and thus body image is flexible and it is susceptible to brain damage. Phantom limbs and neglect syndrome are examples for the cases when brain damage break downs ownership.

4.2.3. Persons, conscious selves and personal identity

We have seen two different views about consciousness and conscious selves. One is the ego-theoretic approach which claims that conscious experiences occurs to a conscious self which is the *subject* of the experiences and the *agent* of the conscious acts. Absolute free will is often attributed to such kind of entities (Descartes, 1969; Eccles, 1989; Stapp, 2004). The other view was the bundle theory of Hume (2000) with its physicalistic extensions. This theory suggests that consciousness is formed by bundles of experiences (or by a field of quantum coherence). Here, there is no subject to which the experiences occur. Among the two theories I have suggested that a physicalist version of the bundle theory which includes quantum coherence best explains consciousness. In that respect consciousness can be defined as a transient field of quantum coherence that is formed in our brains that forms a bundle of experiences. Again, there is no reference to a Cartesian-ego here.

Having restated notion of consciousness that is presented here, let us turn to the concept of person. Perhaps the broadest definition of a person is given by John Locke to describe this elusive phenomenon. Locke described a person as "a thinking, intelligent being, that has reason and reflection, and can consider itself as itself, the same thinking thing, in different times and places" (Locke, 1975, 2 (27): 9). At the first look, we can see that this description is implicitly ego-theoretic for it

assumes “a thinking being.” Even if we correlate the thinking being to our conscious unity, still the conscious unity is only a bundle of thoughts and experiences. There is no thinker.

Regarding the whole discussion throughout the thesis we can conclude that there is no thinker. But, then, from where does the feeling of “I”ness stem from?

We all use sentences like: I do this. I did that. I think such and such. Note that the important aspect of all of these sentences is attribution of agency. The attribution of agency even exists for thoughts (Section 4.2.2) and susceptible to breakdowns by brain disorders. As discussed in Sections 2.3 and 4.2.2, *agent discrimination* is an important process for skill development. I argue that a concept of “I”[26] is generated to “relate” the various experiences that are discriminated by agent discrimination procedure and then attributed with the feeling of agency. Thus the concept of “I” can serve as a basis for retrieving the memories associated with it. The same line of thinking can be applied to explain the feeling of “my”ness. We claim that we are in possession of our body and feel that the body parts are ours. I claim that the feeling of “my”ness is the same as the *ownership* attribute that is discussed in Sections 2.3 and 4.2.2. Our nervous system bestows ownership over certain sensory inputs (experiences) and thus forms a body-image. Consequently, the feeling of “my”ness is an essential part of the experienced body-image.

I suggest that the concept of “I” that is described above and the feeling of ownership (‘my’ness) can be taken together to form a notion of self that is similar to the folk-psychological self or our conception of persons. However, first, it is important to note the concept of “I” is only an apparatus for memory retrieval and feeling of ownership is a means for forming a body image. Second, both of them can be broken down by nerve damage (Ramachandran & Blakeslee, 1998, pp. 39-62; Stephens & Graham, 2000; Blakemore et al., 2003). Third, neither of them is display the properties of an ego-theoretic self because the concept of “I” is not the subject of our experiences. Our consciousness is composed of a bundle of

[26] The concept of ‘I’ presented here has very similar characteristics with the Damasio’s (1999) ‘autobiographical self’ and psychological notion of personality. Even though the line of thinking that I have applied here is fundamentally different from his line of thinking, we both might be pointing to the same thing (personality).

thoughts experiences which are in turn generated by a bundle of microtubules and there is no need for a subject. The concept of “I” is only an apparatus for memory retrieval to the memories that are clustered around the concept. In that respect the concept of “I” is just another element of the bundle. Furthermore the concept of “I” is not the agent of our actions. I argue that our actions are determined by the field of quantum coherence (our consciousness) in our brain and they are later checked by agent discrimination procedures. Lastly, as discussed in Section 4.2.2 absolute free will is not possible even if a quantum mechanical account is proposed for explaining consciousness.

Another question begs an answer in the current discussion: Can our conscious unity as a whole be considered as the conscious self? Such an account still fails to meet the requirements of an ego-theoretic self. First, the conscious unity is a transient entity (a transient field of quantum coherence). Second, it is not the subject of experiences. It is just a bundle of experiences. Although our conscious unity is the agent of our conscious acts, I argue that such agency is not accompanied by absolute free will.

Our conscious unity also fails to fit into the broadest definition of the self that I have proposed in Section 2.1. Our conscious unity can make some self-like discriminations such as agent discrimination and ownership discrimination. But it is only a part of a living being and discriminations that it makes is not for dissociating itself from the others but for helping the organism to be a group-as-a-self (Sections 2.2, 2.3).

The last question about the persons is: How do we feel that we are the same beings at different times and spaces (as in the Lockean sense of person)? I argue that it would be absurd for the nervous system of a normal human to generate more than one I concept for different situations. This would only complicate the process of memory retrieval. So, I claim that one single concept of I is a natural course of events. But I don’t reject the possibility of creating more than one I concept (which can be caused by a catastrophic event) and forming different clusters of memories around them. In fact, multiple personality disorder (dissociative identity disorder) can be such an example.

In sum, I argue that ego-theoretical selves dissolve into a concept of “I,” a body image and a conscious unity (as a bundle of experiences) which lacks a subject and an agent in the strong sense (possessing absolute free will). Simply, an ego-theoretic self do not exist. Likewise our conscious unity fails to fit into the broadest definition of the self that is stated in Section 2.1. It only “acts like” a self by making certain discriminations. Lastly, in the absence of a thinker, our concept of persons are very similar to the concept of “I” that is mentioned above. However this similarity suggests that they are *just* concepts.

CHAPTER 5

CONCLUSION

This chapter summarizes the answers the thesis provides to the research questions we have set (Section 5.1), points out limitations of the discussed model for consciousness (Section 5.2), and presents some directions for the future research (Section 5.3).

5.1. The answers

5.1.1.Consciousness

I argue that consciousness *is* a transient field of quantum coherence that is formed in our brains. A possible candidate for quantum coherence is the microtubules in the neurons (Hameroff & Penrose, 1996) so the above notion of consciousness is essentially bundle-theoretic. I suggest that the only requirement for such kind of field formation is to last long enough for any necessary sensory information or memory trace to be included.

My approach differs from Stapp (2004) and Eccles (1989) for I do not assume an ego-theoretic self or a free agent. It also differs from Hameroff and Penrose (1996) because I don't associate consciousness with Orch OR for that approach associates the frames of consciousness with an essentially instantaneous event. Furthermore, the transient field of consciousness stated above does not have temporal requirements as in the model of Hameroff and Penrose (1996). So, the above notion of consciousness does not subject to the attack by Tegmark (2000). Lastly, the above notion fits in any ontological interpretation of quantum mechanics as long as the interpretation accepts the existence of quantum coherence like Bohm (1981).

5.1.2.Self

I have given a definition for a self to start with: A self is either a living being or is composed of living beings. It is an individual among others and it actively dissociates and discriminates itself from the other beings for the increased chance of survival and reproduction of itself or its components.

Apart from Llinás's (2000) implicit reference to a similar notion of the self, self is usually considered to be directly associated with mind in the form of conscious self.

5.1.3.Conscious Self

As a result of the discussion through the thesis, a definition of the conscious self that is closest to our common-sense intuitions is as follows: The conscious self is consecutive frames of consciousness, each of which can attribute ownership and agency to certain experiences and correlate these with memory traces of previous agency attributed experiences by forming and referring to a *concept of I*. Note that this definition rejects the notion of a unified conscious self (such as an ego or a soul) as an ontological reality. It is only a definition that accords with our common-sense intutions by using the concepts that are formed through this thesis.

The above definition of conscious self is different from ego-theoretic conceptions of conscious self because it is neither a subject for experiences nor it is a free-agent in the absolute sense. The above definition does not include these problematic properties.

The concept of conscious self given here is different from other materialistic claims (Llinás, 2000; Damasio, 1999; LeDoux, 2002) about the nature of the self because it is based on the explicit notions of consciousness (as a field of quantum coherence which solves the binding problem) and the concept of I.

It can be seen that conscious self can make certain discriminations like ownership and agency. But, conscious self fails to fit in the broad notion of self given above for two reasons. First, it is only a part of a living being. Second, the discrimination of ownership does not discriminate conscious self from the others but instead it

serves as a mechanism for the living organism to discriminate itself from the others.

5.1.4.Persons

I argue that our common sense notion of person is best related with the concept of I because the concept of I is only a *concept* like our concept of persons. Since it is involved in the memory retrieval of any agency attributed experience, it displays a kind of continuity. But, different from the Lockean sense of person it does not think (Section 4.2.3).

5.1.5.Other aspects of the self experience

Subjectivity is an inherent property of quantum processing. During a quantum computation, contents of the quantum computation cannot be observed because interference collapses the wave function and ends the quantum processing.

Agency is perceived when an action initiated by the consciousness is completed. The feeling of agency provides agent discrimination and allows the conscious self to determine which movements are initiated by it and which does not.

Before the initiation of an action, a wish experience is not necessary. Desires can be felt to initiate an action, but they are just factors in decision making. Sometimes we may act against our desires while at other times no desire can be felt.

Free will is possible in quantum computation due to inherent randomness (or noncomputability (Penrose, 1994)) of the wave function collapse. But, the possibility of an absolute free will seem impossible because the transformations of Schrödinger's equation are deterministic.

5.2. Limitations of the model

The bundle theoretic approach to self consciousness seems to be adequate in general. But, it has several limitations when explaining conscious selves and consciousness itself. The limitations are discussed in the following sections.

5.2.1. Discontinuity of quantum computation

Quantum computation is discontinuous. It is composed of a series of quantum coherence built-ups that are followed by wave function collapses. It is questionable

if frames of quantum coherence can account for the apparent temporal continuity of our consciousness. However this problem can be solved by introducing felt quality of the passage of time. If we think of consecutive frames of consciousness, passage of time can be felt by employing some neural mechanism which calculates the passage of time between the experienced events and introduce it into the current frame of consciousness by a felt duration which connects the memory traces of the past events to the current experiences. In fact, it has been shown that the feeling of time is not absolute and it can be modulated by some drugs or other factors (Çevik, 2003).

5.2.2. Possibility of large scale quantum coherence in brain

Microtubules can be the sources of quantum coherence. But, a large scale quantum field that we expect needs the involvement of many neurons. The quantum coherence should spread through neurons in order for such a large scale quantum coherence to occur. Some proposals have been made which involves quantum tunneling in gap junctions (cytoplasmic bridges between the neurons) (Woolf & Hameroff, 2001). But, some prefers to propose other alternative places for quantum computation, rather that focusing on micrutubular quantum coherence (John, 2001). Furthermore, there are opponents like Tegmark (2000) who claim that large scale quantum coherence cannot occur in the brain for extended durations.

5.2.3. Coupling and de-coupling of quantum coherence

We experience different things at different instants. Our perceptions come and go and sometimes we recall a memory which joins our ongoing conscious experience. So, there must be a decision mechanism which determines what is going to be involved in a field of quantum coherence. This mechanism should effectively couple and decouple some elements when a frame of quantum processing ends. It is still not known how such a process occurs.

5.2.4. The other minds problem

A theory of quantum coherence provides only a correlate for consciousness. We still do not know that if consciousness is equal to a field of quantum coherence in the brain. It seems that we will never be sure, because fields of quantum coherence are closed to outside observation just like consciousness.

5.3. Directions for future research

Consciousness has always been an elusive phenomenon but now its elusiveness coincides with the elusiveness of quantum events. To provide an adequate explanation for consciousness, principles of quantum computation have to be determined. In turn, to determine the principles of quantum computation, our subjective experiences can be used. By using the subjective experiences, we can have an idea about what is going on within a field of quantum coherence. Consequently, a research program that uses subjective experiences as research tools can provide better answers than a program which uses objective methods alone in determining the laws of quantum computation. The first steps of such a cooperative research include: first, the determination of the exact instruments that our brain can use to give rise to a field of quantum coherence, second determination of how a large scale quantum coherence forms by using those instruments, and third, determination of how it is possible to couple and decouple certain regions of the brain to this field of quantum coherence. With regards to propositions of Damasio (1999) and LeDoux (2002) about the involvement of dopaminergic excitation mechanisms in the generation and suppression of consciousness, I propose that such coupling and de-coupling can be modulated by using such mechanisms which are located in the brain-stem. These mechanisms can add or remove some regions of the nervous system to the frames of quantum coherence by promoting or suppressing the ability of microtubules located in those regions to join the large scale quantum coherence.

REFERENCES

Barett, J. (2003). Everett's Relative-State Formulation of Quantum Mechanics. In Edward N. Zalta (ed.). The Stanford Encyclopedia of Philosophy (Spring 2003 Edition). Retreived October 6, 2005, from http://plato.stanford.edu/entries/qm-everett/

Blakemore, S., Oakley, D. A., Frith, C. D. (2003). Delusions of alien control in the normal brain. Neuropsychologia, 41, 1058-1067.

Bohm, D. (1981). Wholeness and the implicate order. London and New York: Routledge.

Cambridge Dictionary of American English (2004). Retrieved June 20, 2005, from http://www.dictionary.cambridge.org/define.asp?key=self*1+0&dict=A

Carruthers, P. (1986). Introducing Persons. New York, NY: Routledge.

Compact Oxford English Dictionary (2005). Retreived June 20, 2005, from http://www.askoxford.com/concise_oed/self?view=uk

Çevik, M. Ö. (2003), Effects of methamphetamine on duration discrimination. Behavioral Neuroscience, 117 (4), 774-784.

Damasio, A. (1999). The Feeling of What Happens: Body and Emotion in the Making of Consciousness. New York, NY: Hardcourt Brace.

Descartes, R. (1969). Meditations. Trans. E. S. Haldane and G. R. T. Ross. Cambridge: Cambridge University Press. (Original work published in 1641.)

Eccles, J. C. (1989). Evolution of the Brain: Creation of the Self. London and New York: Routledge.

Garret, B. (1998). Persons. In Routledge Encyclopedia of Philosophy, Version 1.0., London and New York: Routledge.

Hameroff, S. (1998). Anesthesia, Consciousness and hydrophobic pockets – a unitary quantum hypothesis of anesthetic action. Toxicology Letters, 100-101, pp. 31-39.

Hameroff, S. R., Nip, A., Porter, M., Tuszynski, J. (2002). Conduction pathways in microtubules, biological quantum computation and consciousness. BioSystems, 64, 149-168.

Hameroff, S. R., Penrose, R. (1996). Orchestrated reduction of quantum coherence in brain microtubules: a model for consciousness. In S. R. Hameroff, A. W. Kaszniak & A. C. Scott (Eds.). Toward a Science of Consciousness: The First Tucson Discussions and Debates. (pp. 508-540). Cambridge, MA: MIT Press.

Hume, D., (2000). A Treatise of Human Nature (Norton, D. F. & Norton, M. J. Eds.). Oxford and New York: Oxford University Press. (Original work published 1739.)

Iacoboni, M., Rayman, J. & Zaidel, E. (1994). Left brain says yes, right brain says no: normative duality in the split brain. In S. R. Hameroff, A. W. Kaszniak & A. C. Scott (Eds.). Toward a Science of Consciousness: The First Tucson Discussions and Debates. (pp. 197-202). Cambridge, MA: MIT Press.

Imamizu, H., Miyauchi, S., Tamada, T., Sasaki, Y., Takino, R., Pütz, B., Yoshioka, T. & Kawato, M. (2000). Human cerebellar activity reflecting an acquired internal model of a novel tool. Nature, 403, 192-195.

Insinna, E. M., Zaborski, P., Tuszynski, J. (1996). Electrodynamics of microtubular motors: the building blocks of a new model. BioSystems, 39, 187-226.

John, E. R. (2001). A field theory of consciousness. Consciousness and Cognition, 10, 184-213.

LeDoux, J. E. (2002). Synaptic Self: How Our Brains Become Who We Are. New York, NY: Penguin Books.

Levin, J. (1998). Qualia. In Routledge Encyclopedia of Philosophy, Version 1.0. London and New York: Routledge.

Libet, B. (1985), Unconscious cerebral initiative and the role of conscious will in voluntary action. Behavioral and Brain Sciences, 8, 529-566.

Llinás, R. (2000). I of the Vortex: From Neurons to Self. Cambridge, MA: MIT Press.

Locke, J. (1975). An Essay Concerning Human Understanding (Nidditch, P. H., Ed.). New York: Oxford University Press. (Original work published 1689).

Mark, V. (1994). Conflicting communicative behavior in a split-brain patient: support for dual consciousness. In S. R. Hameroff, A. W. Kaszniak & A. C. Scott (Eds.). Toward a Science of Consciousness: The First Tucson Discussions and Debates. (pp. 189-196). Cambridge, MA: MIT Press.

Macmillan, M. B. (1986). A wonderful journey through skull and brains: The travels of Mr. Gage's tempting iron. Brain and Cognition, 5 (1), 67-107.

Metzinger, T. (2003). *The Self-Model Theory of Subjectivity*. Cambridge, MA: MIT Press.

Miall, R. C., Wolpert, D. M. (1996). Forward models of physiological motor control. Neural Networks, 9 (8), 1265-1279.

Moody, T. C. (1993). Philosophy and Artificial Intelligence. Eaglewood Cliffs: Prentice Hall.

Nagel, T. (1974). What is it like to be a bat? Philosophical Review, 83, 435-450.

Northoff, G. & Bermpohl, F. (2004). Cortical midline structures and the self. Trends in Neurosciences, 8 (3), 102-107.

Oakley, D. A., Haggard, P. (2005). The timing of brain events: authors' response to Libet's "reply." Journal of Consciousness Studies. Article in press. Retreived November 27, 2005, from http://www.sciencedirect.com

Penrose, R. (1994). Shadows of the Mind: A Search for the Missing Science of Consciousness. Oxford: Oxford University Press.

Pollen, D. A. (2004). Brain stimulation and conscious experience. Consciousness and Cognition, 13, 626-645.

Popper, K. R., Eccles, J. C. (1977). The Self and its Brain. London and New York: Routledge.

Ramachandran, V. S. & Blakeslee, S. (1998). Phantoms in the Brain: Probing the Mysteries of the Human Mind. New York, NY: William Morrow.

Robinson, W. (2003). Epiphenomenalism. In Edward N. Zalta (ed.). The Stanford Encyclopedia of Philosophy (Spring 2003 Edition). Ed. Retreived June 25, 2005, from http://plato.stanford.edu/entries/epiphenomenalism/

Schiffer, F. (1998). Of Two Minds: The Revolutionary Science of Dual Brain Psychology. New York, NY: The Free Press.

Searle, J. R. (1998). How to study consciousness scientifically. In S. R. Hameroff, A. W. Kaszniak & A. C. Scott (Eds.). Toward a Science of Consciousness II: The Second Tucson Discussions and Debates. (pp. 15-29). Cambridge, MA: MIT Press.

Spence, S. A. (1996). Free will in the light of neuropsychiatry. Philosophy, Psychiatry and Psychology, 3, 75-90.

Stapp, H. P. (2004). Mind, matter and quantum mechanics (2nd Ed.). Verlag, Berlin and Helidelberg: Springer.

Stephens, G. L. & Graham, G. (2000). When Self-Consciousness Breaks: Alien Voices and Inserted Thoughts. Cambridge, MA: MIT Press.

Tegmark, M. (2000). Why the brain is probably not a quantum computer. Physical Information Sciences, 128, 155-179.

Treisman, A. (1996). The binding problem. Current Opinion in Neurobiology, 6, 171-178.

Viviani, P., Aymoz, C. (2001). Colour, form, and movement are not perceived simultaneously. Vision Research, 41, 2909-2918.

Vogeley, K. & Fink, R. (2003), Neural correlates of the first person perspective. Trends in Cognitive Sciences, 7 (1), 38-42.

Webb, B. (2004). Neural mechanisms for prediction: do insects have forward models? Trends in Cognitive Sciences, 27 (5), 278-282.

Woolf, N. J., Hameroff, S. R. (2001). A quantum approach to visual consciousness. Trends in Cognitive Sciences, 5 (11), 472-478.

Zohar, D. (1996), Consciousness and Bose-Einstein Condensates. In S. R. Hameroff, A. W. Kaszniak & A. C. Scott (Eds.). Toward a Science of Consciousness: The First Tucson Discussions and Debates. (pp. 439-450). Cambridge, MA: MIT Press.