Consciousness and the Scientific Method

Willard Miranker
Yale University
Department of Computer Science
TR-1230

Consciousness and the Scientific Method

Willard Miranker
Yale University Department of Computer Science

Abstract: Duality, that is, metaphysics is observed to be a common constituent of the scientific method. This motivates formulation of a dual construct to frame a testable theory of consciousness. Three related examples of such theories are presented. They variously employ Hebbian dynamics and the basic notion of measurement in science as metaphysical hypotheses that bridge the so-called explanatory gap between first person consciousness and third person science. The theories require a ramification process in neural systems (brains) from a basic or atomic form of sensation toward the more elaborate forms of consciousness (feelings, qualia...). This process, an extension of the Renormalization methodology of statistical physics is described.

1. INTRODUCTION

Modern science can claim its origin to be the work of Galileo, Kepler and Newton. They produced the great discoveries on which much of what we do in science is based. Galileo conducted experiments, Kepler made observations and simulations (computations), and Newton developed models and analytic tools. All deduced laws of nature. These giants are the fathers of modern physics.

1.1 The role of metaphysics in science

Prior to their work, science, that is the explanations of nature, was largely metaphysical. God created the universe (time, space and matter); he placed the planets where he chose, set them orbiting and maintains their motion. This creationist metaphysics filled all of our culture, and it even fills large parts of it to this day. However in making their contributions, Galileo, Kepler and Newton did not eliminate metaphysics from science. They eliminated the creationist metaphysics and replaced it with one of their own choosing. I want to claim that the introduction of the Newtonian metaphysics was motivated by the inability of the existing methodology to frame an adequate theory and explanation of the natural phenomena with which scientific inquiry was then concerned. (Moreover this practice of appending new or replacement metaphysics to science when explanatory barriers are reached has since been repeated many times.)

Newton's laws of motion and of gravity are based upon assumptions that time and space are absolute, indefinable quantities/qualities that comprise some Galilean structure (Euclidean space-time). Newton had no idea what time, space and matter were. (Nor for that matter for the most part, have we.) He further postulated the existence of something called force, and most remarkably, he postulated the existence of action at a distance of this force (the force of gravity), an action with no known mediation (known neither to him nor to us). He credited the work of Galileo and Kepler as the foundation upon which his inventions were built[1]. Newton also invented the Calculus, the mathematical tool with which his framework could be used to deduce the explanations of the observations and calculations of Kepler (Kepler's laws of planetary motion) and of much of the ensuing development of science as well. In fact those explanations and that development along with experimental confirmation is what provides the validity of Newton's inventions, as it should.

We tend to overlook the metaphysical component of this great work for what it is, as if admitting to the metaphysics is tantamount to admitting to a taint on science. For two hundred years, the Newtonian model sufficed. Then in short order, toward the end of the 19th century, two new sets of experiments were performed that defied explanation within the Newtonian framework. The first or these was the Michelson-Morley experiment that showed the remarkable constancy of the velocity of light, that is, its constancy independent of the velocity of an observer. (This experiment also caused the abandonment of the concept of the Ether, yet another metaphysical construct.) The second was Planck's discovery of the quantal property of the emission of black body radiation. To explain these newly observed phenomena required the introduction into science of new metaphysical assumptions about the nature of the universe.

This remarkable property of light led to the development by Einstein of the special and then the general theory of relativity. Time and space could no longer be viewed as absolute. Clocks run at changeable rates, and so, the very notion of simultaneity is altered. Not only that, but space is now known to be curved by gravity. This shattered the absolute (and comfortable) Newtonian view of space-time.

---

[1] Isaac Newton in a letter to Robert Hooke, February 5, 1675 in reference to Galileo's and Kepler's work in physics and astronomy, where he wrote, "If I have seen further, it's because I stood on the shoulders of giants".

To deal with the quantal property of black body radiation, a collection of scientists produced a profoundly strange metaphysics upon which the remarkable theory of quantum mechanics was developed. This theory which has been used to explain many newly discovered and often mysterious properties of nature is, in fact, a work in progress. Among the names associated with this new theory are Planck, de Broglie, Heisenberg, Schrödinger, Bohr, Dirac and von Neumann. One of the remarkable properties that the quantum theory treats is the dual wave/particle nature of atomic particles. For example when a photon seems to behave as wave, we cannot attribute a location to it. Bohr characterized this conundrum by saying that a photon is not a thing, and we are not necessarily entitled to ask the question, 'where an object is?' So in nature we find objects that are things and objects that are not!

In all three cases ((i) Newtonian, (ii) relativity, (iii) quantum), a common characteristic is the need to deal with phenomena that are outside of human experience and therefore in each case, motivating the invention of new metaphysics to bridge the gap between the phenomena and our experience. These challenges to our experience include, respectively, (i) action at a distance (ii) great velocities and mass and (iii) extremely small scale.

It is somewhat misleading to have suggested that the two hundred-year period between Newton and Einstein was itself without the discovery of unexplainable phenomena and the consequent introduction of additional metaphysics into the scientific corpus. We have already mentioned the Ether. Some others are (a) unmediated attraction/repulsion at a distance of electric charge and the Coulomb law (b) the same for magnetism (c) the invention of electro-magnetic fields. The metaphysics corresponding to these cases does not displace Newton's but rather augments it in kind, and so we hardly notice their metaphysical quality.

So the pattern: unexplainable phenomena, introduction of new metaphysics, expansion of our scientific understanding by reduction within the newly augmented framework is a familiar and repeated one. This process is hardly at an end, since today, motivated by analytic results stemming from the quantum theory, new experiments produce results outside of our current experience. These are certain non-locality results (what Einstein termed 'spooky action at a distance') and certain non-causality results as well. (See Peat, 1990.)

We must also recognize that each implemented appeal to metaphysics to augment our model of nature is a step of dualism in science. Indeed we are obliged to regard our universe, or at least our description of it as multiple.

1.2 The explanatory gap

Now we come to consciousness and the explanatory gap. The argument goes like this.

> Consciousness is a first person phenomenon[2]. We experience it as an internal quality, but it is not externally observable[3]. Science is a third person discipline based on external observation. So science as it is presently constituted cannot explain consciousness. That is, there is indeed, an explanatory gap.

In spite of this gap, the reductive approach is still widely pursued in the study of consciousness, more or less with the following tack. Using existing scientific laws, we shall bit-by-bit discover more and more about the brain, its neural circuitry and its behavior. Doing so will likewise bit-by-bit shed more and more light on aspects of consciousness until its mystery just goes away. While such an approach might very well succeed, it tends to limit consciousness to be epiphenomenal.

Why is the reductive approach so seductive? Probably because it promises an explanation of nature that is within our experience (within our current understanding of science), an explanation that we can not only accept intellectually, but one we can feel and be comfortable with. I suppose that we tend to hope that consciousness will turn out to be a 'thing' after all[4].

However there is the second approach (that we have laid out). This method too has also worked well, more that once, and as we have seen, in the most profound inquiries about nature. Applied to consciousness, this approach is likely to give it a causal aspect. So let us append a new metaphysical construct to science and employ it to explain consciousness, employ it to bridge the explanatory gap. One of the notable, but failed attempts to do this was Descartes'. Descartes introduced his metaphysical res cognitans, and with it, he framed an explanation of consciousness. His hypothesis is now largely viewed as quaint, mostly because it has not found scientific validity. It has not provided

---

[2] By consciousness we mean the experiencing of qualia (sounds, colors, pains…) and feelings (fear, love, hunger...), what Chalmers, 1995 has called the hard problem of consciousness.

[3] I have for this reason called consciousness and internal property of matter.

[4] Indeed many consciousness reductionists are uncomfortable with the possibility that quantum concepts will find a place in the study of consciousness.

for experiments that explain and confirm its character as a bona fide step for the scientific understanding of consciousness[5]. At least not yet!

1.3 Consciousness creationism

Finally I append a comment on what I call consciousness creationism, a term I use to describe a widely held notion that consciousness is uniquely a capacity of man, or possibly in a primitive version, a capacity of some higher animals also. To convey my view of the nonsense of this position consider the following (Llinas, 2001, p. 212).

"We have known for a long time that single cells are capable of irritability, that is the ability to respond to stimuli with a behavioral response consisting of either moving away from, or approaching an object or another cell. In the latter case the relation may be that of hunting for food or fleeing deleterious or threatening conditions. These observations should remind us that there are in single cells certain abilities related in a primitive way to intentionality, and thus to what may be considered a primitive sensory function. If we are allowed to consider that qualia *represent a specialization of such primitive sensorium*, then it is a reasonable conceptual journey from there to multicellular phenomenon of 'corporate feelings' manifested by higher organisms. If this is something we can live with, then we will understand that, *qualia must arise from, fundamentally, properties of single cells,* amplified by the organization of circuits specialized in sensory functions."

Analysis of this suggested circuit organization amplification process that is based on neural net models can be found in Miranker, 2001. See Sheets-Johnstone, 1998 for discussion and examples of consciousness without the need for 'brains'.

1.4 Outline

We shall give three examples of metaphysical constructs that are more characteristic of the scientific method than the res cognitans. We shall then show how these constructs are employed, each to develop a theory of consciousness. We shall also suggest how each might lead to testable experiments that facilitate our understanding of consciousness. Experiments that could ground these constructs as acceptable science. Following this we shall formalize the ramification process of information processing in a neural net model of the brain. This provides the basis for the development of consciousness from its primitive form in the metaphysical constructs of the three examples toward its more developed form in terms of feelings and qaulia that we experience (and as identified in the quote of Llinas in Section 1.3).

---

[5] Contrast this with the formidable body of experimental successes flowing from Newton's metaphysics.

The first of these constructs, summarized in Section 2, is a sort of action at a distance methodology based upon Hebb's law for the development of synaptic strength (Miranker, 2000). We shall note that Hebb's law defines an unmediated process, one that we shall take as a primitive sensory function (the metaphysics). The second construct, summarized in Section 3 (Miranker, 2002) is an extension of von Neumann's theory of observation (of quantum measurement) that he used to frame his theory of quantum mechanics (von Neumann, 1995). This extension is to the measurement (observation) of the information processed by neuronal structures as opposed to the direct measurement of those structures and their activity, as in the customary quantum theory approach.. Finally in Section 4 we give a third construct that combines aspects of the first two, and we also give a summary of the development in Miranker, 2001 of the sensory circuit organization process, the ramification of neural information processing needed to proceed from primitive levels of sensations to the elaborate features of consciousness. This last discussion also shows how the observation/measurement notions of consciousness of Section 3 expressed in terms of quantum physics may be formulated in terms of classical physics. We comment on pan psychist implications of our constructs.

2. CONSCIOUSNESS IS AN INFORMATION STATE

We propose that consciousness corresponds to an information state that accompanies neural processing. The information state is associated with, but is different from the action potentials in terms of which conventional neural processing is conducted. The information state is a dual representation of the unconscious information conveyed by those action potentials. The information state varies in value, depending on details of the neural processing, and when it acquires adequate strength, the information state emerges as the consciousness associated with the neural processing that it parallels. Thus this approach is a coupling of what we shall call internal (the awareness aspects of the information state) and external (the information processing performed by the neural circuitry) properties of matter. This is a duality between primal (the external) and dual (the internal) properties of matter. That is, a duality between externally observable and externally unobservable aspects of nature.

We start by reviewing the standard Hebbian synapse, and we interpret its dynamics as a *quantum of consciousness* or an *atom of awareness*. Its strength, typically called the synaptic weight, and which we shall denote by the symbol s characterize the neuronal synapse. Hebb's law states that this strength changes according to correlation between the input at that synapse and the subsequent output of the neuron itself. The strength increases if the correlation is positive and decreases otherwise. This idea is modeled by the following simple equation given in terms of the input and output signals in question and ds/dt, the time rate of change of the strength (which we shall for convenience here after abbreviate as $\dot{s}$). We shall denote the input and output (the afferent and efferent signals) as v^a and v^e, respectively. Then the equation expressing Hebb's law is the following.

$$\dot{s} = H(v^e, v^a)$$

2.1 The metaphor of experience, atomic awareness, case against epiphenominalism

We interpret the Hebbian dynamics as an atomic awareness (i.e., as a primitive form of consciousness). The mediation of the value of $H(v^e, v^a)$ is unknown, so that the Hebbian dynamics is taken as a postulated and irreducible property of nature (in category analogous to the law of gravity, say)[6]. We say that the synapse experiences the neural input and output, and the nature of the experience is a tendency to change the value of the synaptic strength. We regard $\dot{s}$ as a quantifier of the experience (of what we shall call the *atom*ic awareness) in the synapse.

The metaphor of experience suggests that the atomic awareness is a potentiator of the synaptic dynamics that are expressed by the equation above. Thus atomic awareness is a fundamental quality, and not just an epiphenomenal consequence of neural function. Indeed, the synaptic dynamics that are implicated in neuronal information processing, in memory storage...are themselves, at least in part, a result of the atomic awareness.

This experiential viewpoint, often hidden, is not new in science. It has an analog in the context of gravity. We might say that one mass experiences the presence of a second, and the nature of that experience is a tendency to change the distance between the masses according to Newton's law of gravity. That is, it is the gravitational law that is experienced. As far as we know, the law of gravity is metaphysical. It is a postulated and irreducible property of matter. The gravitational force acts at a distance with no known mediating agency. It is interesting to note that in the Principea, Newton himself

---

[6] This characterization motivates a metaphysical view of the Hebbian dynamics. Should these dynamics become expressible in terms of underlying biochemical processes in the future, we should apply the words metaphysical and irreducible alternatively to those underlying processes as appropriate, leaving this presentation otherwise essentially unchanged. The underlying processes would also replace Hebbian dynamics as the entry point of consciousness (i.e., of the subjective) into our third person, objective science. Conceivably, the discovery of such underlying processes could furnish the ground for a reductive explanation of the atomic awareness described here, potentially of consciousness itself.

uses the metaphor of experience. He points out that the water in a bucket knows if the bucket is rotating or not (Gribben, 1998).

This gravitational analogy can be drawn even closer to the Hebbian by considering Keplerian motion (Corben and Stehle, 1950, p.79). Take the case when an elliptical orbit is executed by one of two masses with a distance r to the other. Between the apses of the orbit, the sign of dr/dt is invariant. We might say that the mass pair experiences the sign of dr/dt. That is, the pair experiences the separate attractive and repulsive stages of motion as separate sensations[7]. We see that dr/dt quantifies the experience of the pair of masses, and correspondingly, that the experience potentiates dr/dt (i.e., potentiates the workings of the law of gravity). Note the *pan psychist* character of these remarks.

2.2 Specification of the atom of awareness and the information state

We shall express consciousness in terms of a time varying information state that we denote by **I** and that corresponds to a collection of neurons (a cell assembly), but one that may vary in time. **I** is a vector with a component corresponding to each neuron in the assembly. Motivated by the discussion in Section 2.1, we make the following postulate.

> $\dot{s}$ *is an indicator of an atomic awareness associated with the corresponding afferent synapse.*

That is, $\dot{s}$ quantifies the degree to which the synapse is aware of the input and output activities v^a and v^e. Next let I denote the component of the vector **I** associated with any particular neuron. Then we further postulate that

> I *is a function of that neuron's vector* $\dot{s}$ *(one component for each afferent).*

Compare this hypothesis with the discussion of Keplerian motion in Section 2.1. Referring to the single celled organisms in the quote of Llinas in Section 1.2, we are motivated to specialize I further, so that it is a function only of the sign of $\dot{s}$, *viz.* $I = I(sig\,\dot{s})$. The value of $I(sig\,\dot{s})$ is to be built up out of the components of $sig\,\dot{s}$ (one component corresponding to each afferent synapse). For this purpose we introduce a vector σ with binary valued components σ_j (where j denotes the j-th afferent).

---

[7] This metaphoric attribution of awareness, even of consciousness, to the inanimate (such as the masses), as described here, is but a small step from a literal attribution of consciousness in the most primitive animate objects (see Sheets-Johnstone, 1998).

$$\sigma_j = \frac{1 + sig\,\dot{s}_j}{2}.$$

σ_j takes the value 1/0 corresponding to the postulated attraction/repulsion in the *j*-th afferent synapse[8], that is, corresponding to the sign of $\dot{s}$. We shall call σ the *sensation* at the *j*-th synapse.

Hebb's idea that the strength of a synapse s_j increases/decreases (that the sensation σ_j is attractive/repulsive) may be embodied in the equation for the synaptic dynamics with an appropriate choice of the Hebb function *H*. (See the appendix for such details.)

2.3 Neuronal level awareness, the mirroring, tuning

Now we define *I* to be an appropriate signed average of the neuron's synaptic sensations [9] σ_j. We take *I* to be the information associated with the neuron. Since *I* is the resultant of all of the neuron's atoms of awareness, it is a representation of the awareness at the neuronal level. Then we call *I* the *neuronal awareness.*

It can be shown that *I* has the value 1/0 when the neuron fires/not-fires, and moreover, approaches the values 1/0 (with *I* increasing/decreasing, as the case may be, that is, as the correlation among the excitatory/inhibitory synaptic activities increases/decreases.) We call this behavior a *mirroring* of the efferent neuronal activity by the neuronal awareness. We shall refer to the increase/decrease of *I* with correlation as a *tuning* of the totality of afferent activity of a neuron. Note that this mirroring embodies the consciousness duality that we have formulated.

2.4 The consciousness hypothesis

Since the cell assembly's information state **I** is composed of the neuronal awarenesses *I* of the individual neurons in that assembly, we shall call **I** an awareness state. The steps leading to this are:

---

[8] Returning briefly to the Keplerian metaphor, we could identify the separate attractive and repulsive sensations there with the two values $(1 + sig\,\dot{r})/2 = 1$ or 0. Indeed this suggests the terminology *sensation* for σ.

[9] Let $\sum_{+/-}$ denote the sum over the excitatory/inhibitory afferents of a neuron. Then for that neuron, *I* is defined as follows.

$$2I = 1 + \frac{1}{n_+}\sum_{+}\sigma_j - \frac{1}{n_-}\sum_{-}\sigma_j.$$

a) The development of a primal/dual interpretation of (neural processing via action potentials)/(the mirroring information of those action potentials by the state **I**).
b) An experimental method for demonstrating that the information conveyed by **I** is the information content of conscious experience.

Step (a) has been discussed, so let us now sketch step (b).

What is needed is a wiring diagram of a sufficiently large portion of the brain and the ability to probe simultaneously the activity of an enormous number of synapses. If we understand the brain as unconscious circuitry, the theory presented here (i.e., the computation of **I** as it depends on the tuning of afferent activity...) will predict what is being consciously experienced by the possessor of a brain as a result of such measurements. We must confirm with probes and external computations what unconscious processing is being mirrored by **I.** The theory can then be verified (falsified) by asking the brain's possessor a simple question!

Since we are conscious of one experience at a time, a key question of detail that must be addressed in performing this experiment concerns the way in which **I** will display the preference or choice (for the emergence into consciousness) among the many different signals that are being processed simultaneously in the brain. One possibility is to inspect **I** for its largest components (corresponding to the most highly tuned afferents), and to imagine a competition among the scenes for coming into consciousness with a winner-takes-all outcome. Finally for the reasons developed, we are able to conclude with the following claim.

The emergent information state **I** *is consciousness.*

3. A QUANTUM STATE MODEL OF CONSCIOUSNESS

We begin with the development of a quantum-like model of information processing in the brain. Then we show that consciousness can be identified with a well-known aspect (namely the collapse of a quantum measurement) of such a model.

3.1 The collapse of the wave function

Now consider two parts of the brain[10] in interaction[11]. One will be the (quantum-like) system, and the second will play the role of a measuring apparatus, namely an observer of the information processing of the first part.

---

[10] While the relevant brain parts are separate in terms of organization, they could be quite

Now consider for the moment a classical system of electrical circuits, modeling the brain. Suppose that these circuits consist of two parts, where the first part passes its output to become the input to the second. On the classical view, the second part is a passive receiver of information, in the sense that the flow of information and influence is unidirectional. On a quantum-like view of the two brain parts, the first part is the passive one. It is the second part, which takes an active role in order to acquire inputs, namely by making a measurement of the state of the first part. Continuing, we expect the second part to cause a change in the first part as it makes its measurement and takes its information. (Recall that the corresponding change in physics caused by a quantum measurement is an internal (i.e., non-observable) aspect of matter, being characterized by the collapse of the wave function of the state being measured; of the first part here.)

The collapse of the quantum mechanical wave function is a fundamental process, an unmediated and non-reducible property of nature. We shall say that the wave function knows that a measurement is being made, and that it responds to this information by collapsing. This is a fundamental form of awareness, a primitive consciousness[12]. The composition of the putative form of awareness in the quantum model under consideration is three things taken together: (i) that a measurement of information in one brain part by a second is being made, (ii) the value (i.e., the outcome) of that measurement, and (iii) the associated collapse of the wave function of the brain part being measured. Then we shall make the metaphysical hypothesis that the collapse of the wave function for a part of the brain (is not caused by consciousness as speculated in customary QM (von Neumann, 1955) but) is the emergence of awareness/consciousness[13].

The measurement itself is what the observer (Part II of the brain) comes away with. So we take the value of the measurement, the input to the observing brain part, to be what we shall call the primal aspect of the measurement, because it is an external, i.e., an

---

intermeshed spatially. In an example discussed in Section 3.3, the two parts comprise and share a single neuron.

[11] Again we stress the euphemistic use of the word brain. One of the so-called brain parts could very well be a sensory organ such as the retina or cochlea. Indeed as an anonymous referee has pointed out, one of these parts could be an element of external reality.

[12] We can expect that the manifestation of the collapse in the brain model here will play a role analogous to that of the unmediated and irreducible Hebb's law for synaptic change that furnishes the metaphysical synaptic level atom of awareness in Section 2.

[13] Whether this awareness is a primitive form or a more fully developed form of consciousness is a matter of the neuronal scale of Part I being measured, as we shall see.

observable aspect of matter. The collapse of the wave function of Part I from ψ_I^- to ψ_I^+, say[14], is a dual aspect of the measurement process (ψ_I^+ is an internal, i.e., non-observable aspect of matter). We use this terminology, because ψ_I^+ is an encoding of the value of the measurement as well. Indeed the value of the measurement could be extracted from ψ_I^+ by making a measurement of the latter (i.e., of Part I) immediately after the collapse in question. In QM performing a measurement corresponds to the application of an associated operator. The value of the measurement (say the number pointed to by the needle of a meter) can only be one of the eigenvalues λ_m of that operator. This duality within the quantum process of observation and collapse is characterized in Figure 3.1.

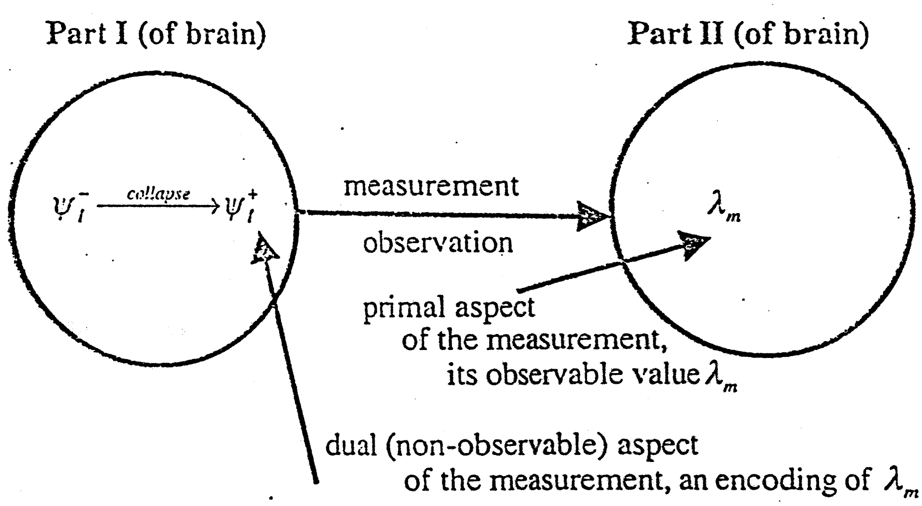

Figure 3.1: The duality between internal and external properties of matter

Then QM requires for the post measurement state (the collapsed state), that we have

$$\psi_I^+ = \lambda_m \psi_m .$$

ψ_m is the eigenstate[15] of the measurement operator corresponding to its eigenvalue λ_m.

---

[14] We use the superscript -/+ to denote the wave function immediately before/after collapse (measurement).

[15] Here are some analytic details about this: Let $\{\lambda_i, \psi_i\}_1^n$ compose a so-called complete set of eigenelements of the measurement operator. Then for the pre-measurement state,

To clarify the primal/dual aspects of our quantum state model of consciousness, we Summary: Part II of the brain makes an observation/measurement of Part I. This causes the wave function of Part I, ψ_I^- to collapse to ψ_I^+, and the value of the measurement, λ_m, becomes the 'classical input' to Part II, or the value of the input, so-to-say. So the primal neural circuit information, λ_m, (a collection of frequencies, neural activities or action potentials, say) is passed on to Part II (or we may say is extracted by Part II from Part I by means of a measurement). This information is the conventional, observable encoding (an external property of matter) of the scene (an image, a color, a sound, a pain…) being processed in Part I. While this information is observable, for instance, by means of a set of voltage probes, we stress that it is unconscious (i.e., it is not available as an internal experience to the possessor of the brain in question).

The state ψ_I^+ (an internal property of matter) is the quantum state of Part I, post measurement. ψ_I^+ is equal to the collapsed wave function $\lambda_m \psi_m$. Since the state ψ_I^+ is given explicitly in terms of λ_m, it encodes the primal scene (the conventional circuitry inputs to Part II) as well. So, ψ_I^+ is a dual state with respect to the property represented by the result of (the value of) the measurement λ_m. We shall say that the dual state (an internal property of matter) mirrors the primal state (an external property of matter. Compare Section 2.3.) Along with the other features of the measurement process, the associated unmediated collapse of ψ_I^- into ψ_I^+ is a manifestation of awareness or experience on the part of the measurement process. *Metaphorically we shall say that the wave function knows of the occurrence of the measurement process and expresses this experience of it by collapsing.* (Compare Section 2.1.) So the collapse/experiencing is taken to be the emergence of consciousness[16]. While the quantum measurement process is a part of the consciousness of the possessor of the brain in question, we stress that it is not observable to anyone else. (This is analogous to the non-observability of the wave

---

we have $\psi_I^- = \sum_{i=1}^{n} a_i \psi_i$, for some constants $a_i, i = 1, \cdots, n$. The quantity λ_m emerges as the value of the measurement with probability $|a_m|^2 / \sum_{i=1}^{n} |a_i|^2$, $m = 1, \ldots, n$.

[16] While this metaphor may seem novel, it has been used implicitly in quantum mechanics proper since the 1920's. Newton uses the metaphor in classical physics in the Principea where he proposes that the water in a bucket knows if the bucket is rotating or not. (This as quoted by Gribben, 1995, p.227.) What is novel here is the replacement of consciousness as the causal agent of the collapse by the attribution of consciousness to the collapse itself.

function in quantum physics. There is current work (Hardy, 1992, Folman and Vager, 1995) attempting to change this.)

A winner-takes-all feature supplies an explanation for why we are conscious of only one experience at a time. It comes as a result of a competition (by means of inhibitory neuronal connections) among a set of possible brain Part II's. Namely, a competition for what is to be measured in Part I (alternatively, which Part II gets to make a measurement), a competition for which experience is to spring into consciousness.

3.2 The atom of awareness, the awareness hierarchy, measurement scales, qualia

Key to explaining qualia is awareness/consciousness levels. These levels form a hierarchy embodied in the collapse process, the discrimination within the hierarchy being supplied by the scale on which the wave function ψ_I of Part I is viewed (i.e., on the scale on which the measurement by Part II is made). For instance, Part II might focus on the wave function of a synapse (only), perhaps the finest scale, and the collapse of such a *primitive wave function* could be taken as the *atom of awareness* (the primitive form of consciousness) in this quantum model. Or Part II might focus on the wave function of an entire neuron, and the collapse of such a wave function could be the neuronal level awareness (i.e., consciousness at a neuronal level) in the quantum model, etc. *The attribution of consciousness to the collapse of the wave function of a neuronal assembly is based on (is a ramification of) the metaphysics (namely on the 'knowing of' the occurrence of the measurement) attributed to the atomic constituents of this awareness hierarchy.* For instance, if a brain part is processing the color red (of course, as unconscious, but externally observable information), the quale red is experienced, because the movement of that information from that brain part corresponds to an awareness (i.e., to a measurement and collapse) of a higher, complex form in this awareness hierarchy. In Section 3.4, we formalize these ideas with a specific quantum model of qualia.

3.3 Quantum effects in neuronal information processing

Do quantum effects occur in neuronal circuitry, in brain parts? Such effects can occur in at least two ways.

A. Quantum effects in the physics of the matter comprising neuronal structures
B. Quantum effects in the information processing performed by the neurons

The possibility of quantum effects in neuronal matter are discussed in Penrose, 1989, 1994 & 1997 and Hameroff and Penrose, 1996. They place these effects in the microtubules that comprise the cyto-skeleton of the neurons.

In Miranker, 1997, it is shown that the processing of information in a computer (in particular, performing the computer arithmetic) can be represented as a quantum measurement process.

Using B, we shall show how neuronal information processing itself may be represented as a quantum process. For reasons of clarity we consider the basic McCulloch-Pitts neuron with n input synapses. We suppose that an input synapse (like any instrument, natural or artificial) can not discriminate between inputs to it that are too close in value. To model this we introduce a screen of values called R, a discrete set of numbers that form the totality of (descriminatable) synaptic values. These are like the values on a digital meter. Denote the n synaptic inputs by the vector. Each x_k is the weighted product, (synaptic strength) x (input to that synapse). We introduce a wave function or state corresponding to the inputs, namely

$$\psi(\xi) = \sum_{k=1}^{n} B_k e^{i\frac{\xi}{h}}.$$

Each $B_k = B_k(\xi)$ is zero everywhere except on an interval having x_k as its midpoint. The value of B_k on that interval is a specified constant that depends on the input x_k. The neuron sums the individual synaptic inputs and fires (or not) depending on a customary threshold process. In Miranker, 1997, a summation operator, called S, is explicitly derived. In terms of this operator, the quantum measurement (corresponding to the weighted summation of inputs performed by the McCulloch-Pitts neuron), denoted by the (customary QM) symbol $(\psi, S\psi)$, has the following value.

$$(\psi, S\psi) = \sum_{i=1}^{n} x_i + \cdots$$

That is, the quantum measurement is indeed the required weighted sum plus a small error.

So when Part II of the brain is taken to correspond to the case of a single neuron (here a McCulloch-Pitts neuron), the operator S represents the action of that neuron making a quantum style measurement of Part I that mirrors the neuron's conventional circuit input weighted summation process. Part I is the neural assembly that in the conventional sense of neural circuitry supplies the totality of required input activities to that McCulloch-Pitts neuron comprising Part II.

3.4 Qualia

As an application of the theory presented here, we develop a model of qualia. For definiteness, consider audition. A mechanical disturbance in the air impinges on the eardrum and is transduced into electrochemical signals in the cochlea. These signals are then processed as primal quantities (e.g., as voltages in the nervous system that are externally measurable). A quantum measurement process characterizes (mirrors) this. In particular, an operator Q called the (phenomenal) sound operator is invoked by a neural assembly instantiated to make that measurement. (This assembly is a Part II.) When it acts, phenomenal sound is created. There is no change in the primal processing (in the physical circuitry), which is augmented by the appearance of the quale sound. There is a competition among neural assemblies for making the measurement, the winner producing the phenomenal result. The competition might be in terms of the inputs, the outputs, or both… Since *a quale is not a thing* (it is not a primal quantity such as a voltage), we do not attribute a location to it. The same picture prevails for other sensory inputs.

Let W denote the operator corresponding to the synaptic weights of the entire neural assembly (the Part II) in question. More specifically, it is a matrix composed of those weights corresponding to exogenous inputs to the assembly. Using the summation operator S introduced in Section 3.3, we form the product SW. This product denotes the weighted sum operator, namely the operator that converts all of the assembly's exogenous inputs into the weighted sums that are the total exogenous inputs to the neurons in the assembly that receives the exogenous inputs in question (Part II). Let ψ denote a vector of wave functions corresponding to those exogenous inputs. (The neural circuitry that supplies these inputs represents a Part I.) The quale will correspond to the quantum measurement operator Q. Then the value of the measurement is given by the expression $(\psi, Q\psi)$, where

$$Q = W^T S W.$$

So the neural assembly (Part II) is specified for the measurement in question. Specification means that the synaptic weights corresponding to the operator W are developed (through learning, training, perhaps also genetically…) to produce the quale.

4. THE RENORMALIZATION OF INFORMATION

Who or what is experiencing the feeling that characterizes our consciousness? This age old question expresses the intractability of the problem with which students of this subject have always been confronted[17]. All of the proposed solutions, among them the homunculus or some other extra material reality have their well-known shortcomings[18] (Chalmers, 1996, Penrose, 1994, Searle, 1994), Stapp, 1996...).

Here we note that this experiencing is a highly ramified form of a common process in both nature and culture, namely an act of making a measurement. A measurement has two aspects of interest for us. One is a device (the measuring instrument), and the second is an observer of the action of the device. We shall characterize neural activity as a measurement process. The unconscious processing/transmission of signals by neural circuitry embodies both the quantity to be measured (the signals) and the measuring instrument (the neuron). It is the Hebbian dynamics that plays the role of the observer, as we shall see. These dynamics specify adjustments to be made to the neuron's synaptic weights, those adjustments reflecting correlation between (i.e., observation of) the inputs and output of the measuring device. We take this Hebbian observer aspect of the neuronal measurement process to be a basic form of sensing. Naturally this basic cellular level phenomenon bears only the most modest resemblance to what we experience personally as consciousness. As we shall propose, the latter arises out of a sequence of ramification steps of the measurement process. This sequence is constructed using a finite hierarchy of layers of large numbers of inter-communicating neural circuits and their assemblies. This finite construction is a *recurrent process of renormalization of information*. It is fundamentally different from a conventional homunculus approach with its unbounded regress dilemma, since it is initiated with the basic level of sensing in place.

4.1 Network of McCulloch-Pitts neurons

The information flow in a network of N McCulloch-Pitts neurons is described by the following input/output dynamics.

---

[17] Early writing on consciousness is found in Aristotle

[18] Appeals to extra material reality, derided in consciousness studies by many as dualism, are positioned in key places throughout science and are typically not explicit. The many examples include (i) the Shrödinger wave function which, having no existence in reality, can be neither measured nor observed, (ii) unmediated action at a distance exemplified by Newton's law of gravity and also by the Coulomb law, (iii) any of the counter factual effects of quantum mechanics, for example the non-locality of space/time (what Einstein termed 'spooky action at a distance') illustrated by both the double slit experiment and the interferometer experiment. The reason for the success of these examples from science and the failure of Descartes' concept of the *res cognitans,* for instance, in the study of consciousness is properly Darwinian.

$$(4.1) \quad S_i(t+1) = \text{sgn}(\sum_{j=1}^{N} w_{ij} S_j(t) + h_i^{ex}), \quad i = 1,\dots,N,$$

where

$$(4.2) \quad \text{sgn}(x) = \begin{cases} 1, & x \geq 0 \\ -1, & x < 0. \end{cases}$$

Here S_i is the output of neuron i. The synaptic connections from each of the N neurons to neuron i are characterized by the vector of synaptic weights $w_i = (w_{i1},\dots,w_{iN})$. h_i^{ex} represents the exogenous input to neuron i.

4.2 Hebbian dynamics

Information (e.g., a memory trace) is taken to be encoded (associatively) in the neural net by means of its synaptic weights. As the neurons conduct information processing according to (4.1), the synaptic weights change according to Hebbian dynamics. Namely,

$$(4.3) \quad \frac{dw_{ij}(t)}{dt} = \mathcal{H}(S_i(t), S_j(t-1)),$$

where $\mathcal{H}$ is the so-called Hebb function. Hebb's proposal may be expressed by saying that $\mathcal{H}$ is to have the sign of the correlation of its two arguments.

4.3 The measurement/observation process

Using Figure 4.1, we characterize the pair of interrelated dynamical systems (4.1) - (4.3) as a measurement/observation process. The input data is the vector $S(t) = (S_1(t),\dots,S_N(t))$. The measuring instrument is the neuron. The value of the measurement is the neuronal output $S(t+1)$ specified by (4.1). The observer of the measurement[19] is the process that generates the Hebbian dynamics specified by (4.3).

---

[19] The act of measurement in quantum mechanics (associated with the collapse of the wave function) is caused by the consciousness of a human observer when the latter notes the value of the measurement produced by some instrument (this according to von Neumann, 1955. See also Wigner, 1961).) The picture here has some similarity to this point of view in the sense that the synaptic weights and the process of change characterized by the Hebbian dynamics are an aspect of the conscious observer observing himself.

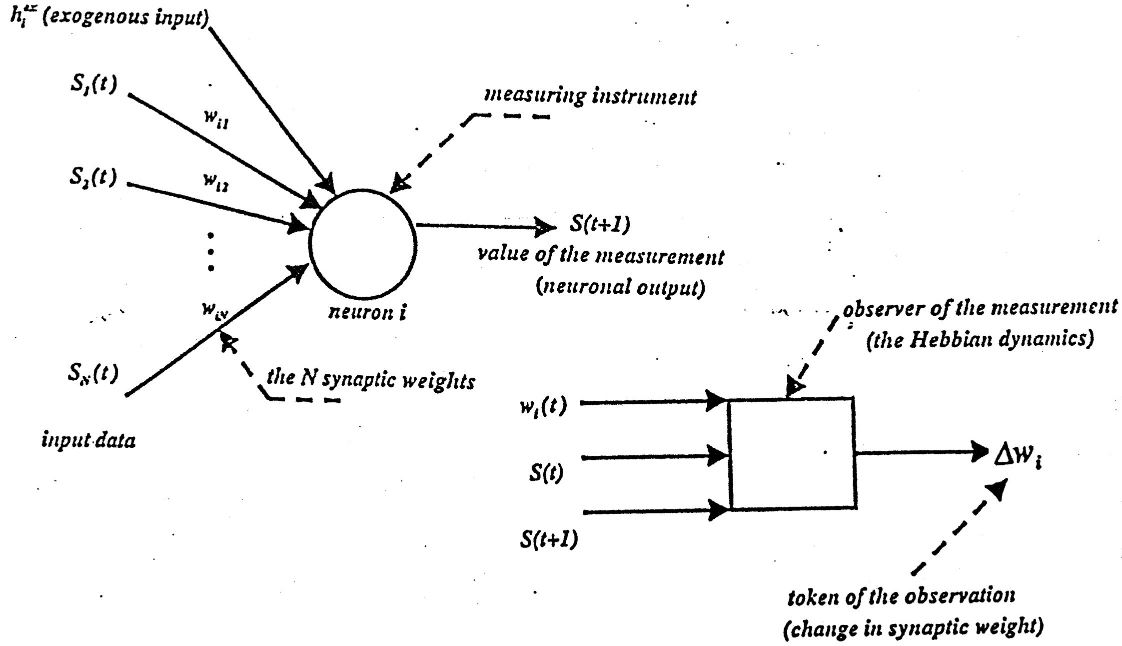

Fig. 4.1: The neural input/output dynamics (upper left) and the Hebbian synaptic dynamics (lower right) showing the token of observation

Discretizing the time and indexing successive time instants with n, we describe the Hebbian observation process in terms of the incremental change Δw_{ij} in synaptic weight.

(4.4) $\Delta w_{ij}(n) = \mathcal{H}(S_i(n), S_j(n-1)).$

4.4 The token of awareness and causality

We introduce the following terminology.

Δw_{ij}, as given in (4.4), represents the degree to which the j-th input synapse of neuron i is aware of the measurement as a result of the observer process.

Although Δw_{ij} is only a token of the awareness which itself is the entire process represented by the Hebbian dynamics (4.4), we shall sometimes refer to it as the awareness itself[20]. It is this awareness quality (which bridges the so-called explanatory

[20] In an earlier work (Miranker, 2000)), a quantity related to this concept of awareness was called an atom of awareness, and it was viewed as a non-reducible and causal quality of matter. Here we avoid such a stand and allow that this awareness feature might very well be a reducible phenomenon (causal or epiphenomenal, as the case may be). In the

gap[21]) that is the starting point for a theory of consciousness. So it is appropriate to inquire whether this quality (and consciousness itself) is causal or epiphenomenal[22].

If synaptic weight change is explainable by known natural laws, as it occurs in dependence on the two arguments $S_i(n)$ and $S_j(n-1)$ of the Hebb function in (2.4), then the awareness seems to be epiphenomenal. If on the other hand, the weight change occurs because the correlation between these two arguments is based on an unmediated process (like action at a distance), then physical matter (that which records the synaptic weight) is being acted upon by an unmediated phenomenon, the awareness. In this case awareness seems to be causal. New laws of physics[23] are not mandatory to explain this causal quality, since the correlation in question might be expressible in terms of a known unmediated action at a distance phenomenon such as Coulomb attraction/repulsion. Indeed since neurotransmission and synaptic weight change are electrochemical in nature, the Coulomb law is a very likely player in the explanation at the cellular level of this first person phenomenon that we call consciousness.

4.5 Expected value of observation

A simple form for $\mathcal{H}(x,y)$ that characterizes the correlation properties of synaptic weight development enunciated by Hebb (cf. (4.3) f), is the product form, namely

$$(4.5) \quad \mathcal{H}(x,y) = \kappa(x - \langle x \rangle)(y - \langle y \rangle),$$

where κ is a constant and $\langle\ \rangle$ denotes the mean value. We may take $\kappa = 1$. Now combining (4.5) with (4.4) and taking expected values, we find that

$$(4.6) \quad \langle \Delta w_{ij} \rangle = \langle S_i S_j \rangle - \langle S_i \rangle \langle S_j \rangle.$$

---

end, of course, the fundamental laws of physics in terms of which the Hebbian dynamics might be expressed are themselves non-reducible. Compare footnote 3.

[21] The explanatory gap is the name given to the disconnect between our third person science and our first person consciousness (Chalmers, 1996).

[22] Epiphenominal aspects of consciousness, frequently referred to as emergent properties, are hardly superficial. They play a principle role in both daily existence and in evolution (Sheets-Johnston, 1999).

[23] Penrose, 1994 and Chalmers, 1996 speculate that new laws of physics are needed to explain consciousness. Penrose offers comments on the possible nature of such new laws.

So the expected value of awareness (literally, the token of observation (cf. Section 4.4)) is proportional to the neural output correlation function. (In the case of modeling magnetic spins S_i, the right member of (4.6) is referred to as the spin correlation function.) The terminology introduced in Sections 4.2-4.6 (concerning awareness of the neural processing by a synapse, that processing and that awareness interpreted as a measurement process) motivates the following definition.

Definition. We say that during a step of neural processing, the *ij*-th synapse experiences a sensation σ_{ij}, where

(4.7) $\sigma_{ij} = \langle \Delta w_{ij} \rangle.$

We stress that while the members of (4.7) are tokens of awareness (and may sometimes for convenience be referred to as the awareness), awareness is properly a property characterizing the behavior of the *ij*-th synapse (indeed the entire Hebbian dynamics process) as an observer of what we have called the neuronal measurement process. (See Figure 4.1.)

4.6 Neural dynamics as a communication process, mirroring

The average over a neuron's synaptic awareness tokens will be a derived token of interest to us, namely

(4.8) $\mathcal{I}_i = \frac{1}{N}\sum_j \sigma_{ij}.$

It may be shown that $\mathcal{I}_i$ is proportional to the expected value of neuronal output $\langle S_i \rangle$. (See Miranker, 2001, Section 4.) So the subject of the cellular level awareness represented by the token $\mathcal{I}_i$ is the neuronal output itself. We call this a mirroring of the expected value of the neuronal output by $\mathcal{I}_i$.

4.7 Mutual information

Since many different patterns of input synaptic activity can correspond to each one of the two neuronal output values, $S_i = \pm 1$, all we can expect to tell about the neuronal input pattern from the value of the output is the average (counting signature) of the inputs. The mirroring expressed by (4.8) tells us this. So $\mathcal{I}_i$ is the mutual information[24] of neuronal

[24] Mutual information is the uncertainty about an input that is resolved by knowledge of the output. See Haykin, 1999, Sect. 11.2.

input/output dynamics (subject to noise) interpreted as a step in a communication process (an information transmission process).

4.8 Neuronal feeling, the observer

Since $\mathcal{I}_i$ is an appropriate signed average of the sensations σ_{ij} (cf. (4.7), (4.8)) of the input synapses of neuron i, it is a token of sensation of the entire neuron, which we shall refer to as the feeling of the neuron. (We stress once more the need to differentiate between a quality and its token.) $\mathcal{I}_i$ may be positive or negative, and this suggests that neuronal feeling (that is the mutual information of the neural measurement process) may be attractive or repulsive.

Indeed this feeling $\mathcal{I}_i$ is expressed in terms of the (externally/third person) measurable but unconscious neuronal output activity, that activity encoding the neural information being processed. So in this sense neuronal feeling is a non observable reflection of that information redounding directly from the so-called observer aspect of the measurement process which is neural dynamics (cf. Figure 4.1), in particular redounding directly from the Hebbian synaptic dynamics.

4.9 An analogy between feeling and magnetization, a field of consciousness

Except for the inclusion of the Hebbian dynamics, the neural modeling discussed here is a well-known variant of the mean field analysis of the Ising model of magnetization. This analogy describes the feeling (defined here) associated with the neurons in a net as corresponding to the magnetization surrounding the dipoles composing a magnet. Thus our model suggests that consciousness is a field of feeling associated with a neural net. In what sense, if any, this field might surround the net (as the magnetic field surrounds the magnet), we can not yet say. Perhaps like the quantum mechanical probability amplitude, it is a field without existence in reality, a field that can be neither measured nor observed externally.

4.10 Renormalization and the ramification of feelings

When the renormalization process of statistical physics is applied to the Ising model of magnetic spins, a block structure of spins replaces the original individual spins. This process is iterated, resulting in a hierarchy of the basic magnetic phenomena at a set of increasing scales and ranges (what we here call ramification). The focus of the renormalization in physics is to produce a theory for specifying the phase changes of the material, a process that is manifest at the higher scales. Our appeal to this methodology translates it to produce successively higher order (longer-range) notions of sensation.

The changes of phase that ultimately arise are interpreted as novel states of consciousness (such as feelings, qualia...) arising at the highest scales of ramification.

Borrowing from physics[25], we note that the analog of sensation in the Ising model increases in value (strength) as the renormalization hierarchy is ascended. (See Miranker, 2001, Section 5 for details. There an example is shown where the sensation strength doubles for each five iteration steps.). This suggests that ramification of sensation is accompanied by increases both of complexity and of strength, two features that serve the viewpoint that ramification is a route toward consciousness. Details of the renormalization process for neural nets including an explicit form for the higher level token of sensation may be found in Miranker, 2001, Section 6. We note that the renormalization of neural nets proceeds without the customary geometric regularity restrictions of the analogous development for magnetic spins. In particular, a block of spins is replaced by an assembly of neurons whose configuration may be arbitrarily specified (as by some relevant processing organization requirement. Moreover called neighboring neurons are defined by their connectivity and not necessarily by their proximity.

In Section 7 of Miranker, 2001, the novel phases of information that the renormalization methodology characterizes are shown to contain the so-called spurious states of the neural net model. Examples of these states are presented[26].

REFERENCES

Amit, D., Gutfreund, H., Sompolinsky, H. (1987). "Statistical mechanics of neural networks near saturation". Annals of Physics, **173**.

Chalmers, D. (1995) "Facing up to the problem of consciousness", J. Consciousness Studies, **3**.

Chalmers, D. (1996). *The Conscious Mind: In Search of a Fundamental Theory,* New York, Oxford University Press.

Corben, H. and Stehle, P. (1950), *Classical Mechanics,* (New York: John Wiley).

Folman, R and Vager, Z. (1995) 'Empty wave detecting experiments', Found Phys Lett **8**, pp. 55-61.

---

[25] Renormalization theory is attributed to L. Kadanoff (Kadanoff, 1966) and K.Wilson (Wilson, 1971). For an accessible presentation of that theory, see Goldenfeld, 1992.

[26] For exposition of the use of methods of statistical mechanics in neural nets, see Amit, Gutfreund, Sompolinsky, 1987.

Goldenfeld, N. (1992). *Lectures on Phase Transitions and the Renormalization Group*, Reading, Addison-Wesley.

Gribben, J. (1998), *Schrödinger's Kittens and the search for Reality*, (Boston: Little, Brown and Company).

Hameroff, S. and Penrose, R. (1996), 'Conscious events as orchestrated time-space selections', JCS, **3**,pp. 36-53.

Hardy, L. (1992), 'On the existence of empty waves in quantum theory', Phys Lett A **167**, pp.11-16.

Haykin, S. (1999). *Neural Networks, a Comprehensive Foundation*. Saddle River, Prentice Hall.

Kadanoff, L. (1966). *Physics*, **2**.

Llinas, R. (2001), *i of the Vortex,* The MIT Press.

Miranker, W. (1997), "Interference effects in computation," SIAM Review, **39**, pp. 630-643.

Miranker, W. (2000), "Consciousness is an information state", J. Neural, Parallel and Scientific Computation", **8**.

Miranker, W. (2001), "The renormalization of information", Yale/DCS/TR-1215 (submitted).

Miranker, W. (2002), "A quantum state model of consciousness", J. Consciousness Studies (in press).

Peat, D. (1990). *Einstein's Moon: Bell's Theorem and the Curious quest for Quantum Reality*, Chicago: Contemporary Books).

Penrose, R. (1989), *The Emperor's New Mind* (New York: Oxford University Press).

Penrose, R. (1994), *Shadows of the Mind* (New York: Oxford University Press).

Penrose, R. (1997), *The Large, the Small and the Human Mind*, (Cambridge: Cambridge University Press).

Searle, J. (1994). *The Rediscovery of the Mind*, Cambridge, MIT Press.

Sheets-Johnstone, M. (1998). "Consciousness; a natural history", JCS, **3**, pp. 194-210.

Stapp, H. (1998). "The hard problem: a quantum approach, J. Consciousness Studies, **3**.

Von Neumann, J. (1955), *Mathematical Foundations of Quantum Mechanics* (Princeton: Princeton University Press).

Wigner, E. (1961). "Remarks on the mind-body problem", in *The Scientist Speculates*, ed. I.J. Good, London, Heineman.

Wilson, K. (1971). "Renormalization group and critical phenomena I & II", Phys. Rev **B4.**