The relation between quantum mechanics and higher brain functions: Lessons from quantum computation and neurobiology

Christof Koch[1,2] and Klaus Hepp[1]

April 2. 2007

[1]Institute for Neuroinformatics, ETHZ and University of Zürich, Switzerland
[2]Division of Biology and Division of Engineering and Applied Science, California Institute of Technology, Pasadena, CA 91125, USA. To whom all correspondence should be addressed. Email koch@klab.caltech.edu

The relationship between quantum mechanics and higher brain functions is an entertaining topic at parties between a mixed, open-minded group of academics. It is, however, also a frequently asked question at international scientific conferences, in funding agencies and sometimes at the end of our lives, when thinking about ultimate truths. Therefore a well-founded understanding of these issues is desirable. The role of quantum mechanics for the photons received by the eye and for the molecules of life is not controversial. The critical questions we are here concerned with is whether any components of the nervous system - a 300^{o} Kelvin wet and warm tissue strongly coupled to its environment - display any macroscopic quantum behaviors, such as quantum entanglement, and whether such quantum computations have any useful functions to perform. Neurobiologists and most physicists believe that on the cellular level, the interaction of neurons is governed by classical physics. A small minority, however, maintains that quantum mechanics is important for understanding higher brain functions, e.g. for the generation of voluntary movements (free will), for high-level perception and for consciousness. Arguments from biophysics and computational neuroscience make this unlikely.

1. Introduction

After outlining the problem in brain science and in psychology that some scholars seek to address through quantum mechanics (QM), we outline two arguments that make this unlikely. Firstly, it is unclear what computational advantage QM would provide to the brain over those associated with classical physics. Secondly, as the brain is a hot and wet environment, decoherence will rapidly destroy any macroscopic quantum superposition.

Quantum Mechanics

Quantum mechanics is, in the framework of this essay, the basic theory of all low-energy phenomena for bodies and brains at home and in the laboratory, e.g. for a human lying in a magnetic resonance scanner in an neuropsychological experiment. Hence, QM is the well-established non-relativistic 'text-book theory' of atoms, electrons and photons, below the energy for pair creation of massive particles (see e.g. Gottfried and Yan 2003). In contrast to classical physics and to that other great edifice of modern physics, general relativity, QM is fundamentally non-deterministic. It explains a range of phenomena that cannot be understood within a classical context: light or any small object can behave like a wave or like a particle depending on the experimental setup (*wave-particle duality*); the position and the momentum of an object cannot both be simultaneously determined with perfect accuracy (*Heisenberg uncertainty-principle*); and the quantum states of multiple objects, such as two coupled electrons, may be highly correlated even though they are spatially separated, violating our intuition about locality (*quantum entanglement*).

We rely on the mathematical formulation by (von Neumann 1932). The only predictions of QM (the best we can make in non-relativistic atomic physics and quantum computation (Mermin 2003)) are - given the dynamical law in terms of the family $\{H(t)\}$ of Hamiltonians of the system for all times t and corresponding propagators $\{U(t,s)\}$ - to predict for any chosen initial state S at time s and any chosen yes-no question P the future probabilities $Tr(PS(t))$ at time t, where $S(t)=U(t,s)SU(t,s)^*$. $Tr(PS(t))$ is the probability for 'yes', while $Tr((I-P)S(t)))$ is the probability for 'no'. The time evolution from s to t is given by a 2-parameter family of unitary propagators U(t,s), the solution of a time-dependent Schrödinger equation. There is a dualism in QM between the *dynamical law* $\{H(t)\}$ of the system and the *choices* S, s, P, t of *initial states and final questions* asked about the system. In poetic language, the dynamical law is given by Nature and the allowed questions are sometimes posed by the Mind of the experimenter. However, the introduction of consciousness in Chap VI of

(von Neumann 1932) is only a critical description of human activity and not a theory of mind, expressed in his often misunderstood statement:

> Experience only asserts something like: an observer has made a certain (subjective) perception, but never such as: a certain physical quantity has a certain value.

It should not be forgotten that, even for a simple system, most questions cannot be implemented in the laboratory of even the best equipped physicist by ideal measurements à la von Neumann.

Higher Brain Functions

Higher brain functions (HBF) are macroscopic control processes whose computational basis is beginning to be understood and that take place in the brains of humans and other animals. Typical HBF include sensory perception, action, memory, planning and consciousness (the neuroscience background for this essay is fully covered in (Koch 2004)). For simplicity, we shall restrict ourselves to perception by the mammalian visual system, and to sensorimotor control of rapid eye movements in mammals. Visual perception and rapid eye movements are strongly linked to each other and can be often studied in isolation from other brain functions. These functions involve many areas of cerebral cortex and its associated satellites, in the thalamus and midbrain, and are only partially accessible to consciousness. As reductionists, we make the working hypothesis that consciousness is also a HBF.

We immediately admit that neurobiology is a young science without a sound mathematical structure, unlike QM. However, neurobiological aspects of consciousness, in particular conscious visual perception, can be studied scientifically using a battery of highly sophisticated neuropsychological tests, invasive and non-invasive brain imaging, cross-checked reports of human and animal (e.g. monkeys or mice) subjects and – last, not least – by the first person's insights of the experimenter. The modern quest to understand the relationship between the subjective, conscious mind and the objective, material brain is focused on the empirically tractable problem of isolating the neuronal correlates of consciousness (NCC), the minimal set of neuronal events and structures jointly sufficient for any one specific conscious percept. Furthermore, scientists and clinicians are acquiring more sophisticated technologies to move from correlation to causation by perturbing the brain in a delicate, reversible, and transient manner (e.g. intracortical electrical stimulation in monkeys or in neurosurgical subjects or transcranial magnetic stimulation in normal observers). There are a number

of excellent textbooks and longer review articles on the neural correlates of consciousness (see Koch 2004 and references therein).

Note that it is not clear at the moment whether the NCC can be clearly isolated and identified. In highly interconnected networks, such as the cerebral cortex, it may be very difficult to assign causation to specific neuronal actors. Furthermore, even if this project is successful, knowing the NCC is not equivalent to understanding consciousness. For this, a final *theory of consciousness* is required (for one promising candidate based on information theory, see Tononi, 2004).

In the following section, we scrutinize past efforts to invoke QM for explaining HBF and point out the many explanatory gaps in this approach. In the third section, we turn to the theoretical and experimental insights obtained in the past decade from quantum computations and argue that QM will also in the foreseeable future be ill-positioned to explain HBF. In the last section we try to show that a classical (i.e. classical physics and engineering based) theory of higher brain functions is on its way towards surprising new insights, even about consciousness.

2. Quantum Explanations of Higher Brain Functions

In this section we discuss the contributions of Eccles, Penrose and Stapp to invoke QM to explain HBF and show that they all take a dualistic stance, without refutable experimental predictions. Although we privately have sympathy with some of their beliefs, their explanations of HBF are incompatible with our reductionistic view. In their joint work 'The Self and Its Brain' (Popper and Eccles 1977), the philosopher Karl Popper and the Nobel prize laureate and neurobiologist John Eccles introduced the framework of three worlds: 'World 1' (W1) the physical world, including brains, the 'World 2' (W2) of mental, subjective states, and the 'World 3' (W3) of abstract ideas, physical laws, language, ethics and other products of human thought. Such a categorization is useful for many philosophical discussions and is related to the three worlds of the mathematical physicist, Roger Penrose, in his book 'Shadows of the Mind' (Penrose 1994): the physical world, the world of conscious perceptions and the world of mathematical forms. From the rich contents of these books we will select parts where QM is invoked for explaining HBF.

Eccles' proposal for 'free will' by quantum computations at cortical synapses

Eccles (1994) undertook the arduous task to link his W1 to W2. In collaboration with the physicist Beck he used QM for developing a theory of voluntary movement, which we will illustrate for rapid eye movements. A subject 'decides' to look in a certain direction. This requires – according to Beck and Eccles (1992, 2003) – that this 'idea' is communicated from the mind in W2 to the frontal eye fields (FEF), a small region in the front of cortex in W1, without violating the laws of physics. Typically, people rapidly move their eyes in a coordinated and highly stereotypical, jumping manner called a saccade, making about three saccades every second of their waking life. Every saccade is accompanied by a macroscopic brain activation involving millions of neurons in a rather stereotyped manner. If during our lives we read one thousand books, those of us who read languages written from left to right -voluntarily make more than one million almost identical saccades of about 2 deg away from the fovea, the point of sharpest seeing at the center of our gaze, to the right (Rayner 1998)!

For the following it is important to know that rapid, millisecond communication between neurons occurs using binary, all-or-none electrical impulses – spikes or action potentials – of about 0.5-1 msec in duration and a tenth of a volt in amplitude. At the nerve endings – synapses – these impulses release one or more packets of neurotransmitter. These molecules rapidly diffuse across the small cleft that separates the nerve ending (pre-synaptic terminal) from the post-synaptic terminal located on the next neuron. Here, the neurotransmitter causes a molecular reaction that eventually leads to the generation of a small, electrical signal, an excitatory post-synaptic potential (EPSP) at an excitatory synapse. Thus, fast communication among most neurons is based on an electrical-chemical-electrical conversion. The brain is exceedingly rich in such synaptic connections, between $10^8 – 10^9$ per mm^3 of cortical tissue.

Beck and Eccles's explanation of the generation of voluntary eye movements is to postulate that at the synapses between certain neurons in the FEF there are low-dimensional quantum systems (Qbits) which control the release (exocytosis; see e.g. Becherer and Rettig 2006) of neurotransmitter, whenever an action potential arrives at the presynaptic terminal, and that these Qbits are coherently coupled by the laws of QM. Now let us follow the authors (Italics by the authors, additions […] by us):

> We present now that the mental intention (the volition) becomes neurally effective by momentarily increasing the probability of exocytosis in selected cortical areas such as the FEF neurons [the supplemental motor area in their example]. In the language of quantum mechanics this means a *selection of events* (the event that the trigger

mechanism has functioned, which is already prepared with a certain probability). This act of selection is related to Wigner's (1967) selection process of the mind on quantal states, and its mechanism clearly lies beyond ordinary QM. Effectively this selection mechanism increases the probability for exocytosis, and in this way generates increased EPSP *without violation of conservation laws*. Furthermore, the interaction of mental events with quantum probability amplitudes for exocytosis introduces a coherent coupling of a large number of individual amplitudes of the hundreds of thousands of boutons in a dendron. This then leads to an overwhelming variety of actualities, or modes, in brain activity. Physicists will realize the close analogy to laser action, or, more generally, to the phenomenon of self-organization.

There are a number of problems with this proposal:

(1) The probability for exocytosis is a physical process that is entirely in W1 and therefore cannot momentarily be increased by volition from W2 without violating physics (Hepp 1972, 1998). QM does not generally predict the occurrence of single events – this is where W2 could act, by influencing when a particular event takes place. However, this does not provide a mechanism for free will, as proposed by Beck and Eccles (2003). The generation of millions of identical saccades during reading is not a single event and involves the probabilities of W1 physics, on which the mind in W2 has no influence.

(2) The coherent coupling of a large number of QM degrees of freedom and the resulting laser-like operation (Haken 1970, Hepp and Lieb 1975) in the 'wet and hot' brain has no physical basis as discussed in section 3.

In section 4 we outline a classical model for generating voluntary saccades during reading.

Penrose's proposals for a quantum gravity theory of the conscious mind

Penrose has, as have many mathematical physicists, a strong belief in the independent existence of a World 3 of mathematical objects and physical laws, which the scientist's mind in World 2 discovers by operations which Penrose believes to be non-computational in the framework of Church and Turing. Penrose's (1994) explanatory scheme of how the mind of a mathematician captures Platonic ideas is a joy to read (we are looking forward to his next book!), but irrelevant in our context, since it relies on specific properties of a yet-to-be-discovered quantum theory of gravitation (QG). In addition, as we shall see in section 3, the proposed neurobiological implementation of QG for

generating consciousness (Hameroff and Penrose 1996) is highly implausible. Finally there is not even an outline of how consciousness as an algorithm of the QG brain arrives at discovering mathematical truths. It is simply asserted.

In order to be neurobiologically more realistic, Penrose discusses illusions in the perception of order of two events in time (which we can observe every morning, when the alarm clock seems to start to ring after it has woken us up). We cannot refrain from quoting his 'explanation' in chapter 7.11 of Libet's (1979, 2004) study of the chronometry of volition:

> *If*, in some manifestation of consciousness, classical reasoning about the temporal ordering of events leads us to a contradictory conclusion, then there is a strong indication that quantum actions are indeed at work!

This is amusing, since psychology knows of hundreds of illusions that appear to violate classical physics as well as common sense (e.g. in the motion after-effect, an object appears to move without changing its position) that can be explained in a completely conventional framework. In the case of apparent violation of temporal order, (Lau *et al.* 2006) recently reported in a careful fMRI study of Libet's timing method that the measuring process affects the neural representation of action and thus also the perceived onset that the method is designed to measure. Furthermore, in (Lau *et al.* 2007), disrupting brain activity a fraction of a second following an external event perturbs the perceived duration of an event that occurred previously. In other words, the conscious perception of any physical event takes time to develop and must somehow be back-dated by the brain. None of this need involve anything but classical physics.

In section 4 we will discuss classically cognitive aspects of temporal order in the attentional blink modeled in the 'global workspace' theory of consciousness.

Stapp's ideas on the Quantum Zeno Effect

Stapp (2003) relies on a literal interpretation of von Neumann's axiomatization of quantum mechanics. He calls the unitary time evolution of a state from its initial state S into S(t) 'mechanical' and the choice of a projector P of a 'yes-no' question 'conscious'. In a collaboration with two neuropsychologists (Schwartz et al. 2005) he explains how the mind acts on the brain during cognitive control of emotions. They discuss an experiment by Ochsner et al. (2002), where fearful faces are shown to a subject in a fMRI brain scanner. This generates measurable emotional reactions and a strong activation in the amygdala, a forebrain structure known for its close link to fear and

fear-associations. In one series of scans one can see that these reactions can be repressed, when the subject receives the cue 'reappraise', and areas in the prefrontal and anterior cingulate cortex 'light up'. Now we shall quote (Schwartz et al. 2005) in their QM explanation of the cognitive control of emotions:

> In the classic approach the dynamics must in principle be describable in terms of the local deterministic classic laws that, according to those principles, are supposed to govern the motions of atomic-sized entities. The quantum approach is fundamentally different. In the first place the idea that all causation is *fundamentally mechanical* is dropped as being prejudicial and unsupported either by direct evidence or by contemporary physical theory. The quantum model of the human person is essentially dualistic, with one of the two components being described in psychological language and the other being described in physical terms.

We hope to give a fair account of the authors' point of view. The two 'worlds' pertain to two sets of objects in orthodox QM, on one side the initial and final choices and on the other side the dynamics, as outlined in section 1. We remark that even in classical physics there is a similar 'psychological' choice of the initial (or final) conditions (the initial data of the positions and velocities of all particles and fields) which are more 'conscious choices' than the dynamical laws (e.g. Newton's equations for the planetary two-body system). We continue with the QM explanation of conscious control of emotions, in the words of the authors:

> When no effort is applied [cue: 'don't control your emotions!'] the temporal development of the body/brain will be [S(t) which is] approximately in accord with the principles of classic *statistical* mechanics, for reasons described earlier in connection with strong *decoherence* effects. But important departures from the classical statistical predictions can be caused by conscious effort. This effort can cause to be held in place for an extended period [t], a pattern [PSP] of neural activity that constitutes a *template for action*. This delay [PSP instead of PS(t)P, i.e. by suppressing the 'mechanical' body/brain evolution by the Quantum Zeno Effect (QZE; Misra and Sudarshan 1977)] can cause the specified action to occur. In the experiments of Ochsner the effort of the subject to 'reappraise' causes the 'reappraise' template [PSP] to be held in place and the holding in place of this template *causes* the suppression of the limbic response. These causal effects are, by the QZE, mathematical consequences of the quantum rules. Thus the 'subjective' and 'objective' aspects of the data are tied together by quantum rules that *directly specify the causal effects upon the subject's brain of the choices made by the subject, without needing to specify how these choices came about.*

We are struck by the boldness of this QM 'explanation', as in all other dualistic theories. A theoretical physicist would like to understand whether the QZE holds for the Hamiltonian of the subject in the scanner. This is a non-trivial mathematical problem (Schmidt 2003) and far remote from what happens in the simple models that can be fully analyzed as described in (Joos et al. 2003): In this book, the authors consider a pure state S in a finite-dimensional quantum system with Hamiltonian H. If, in a time interval [0,t], S evolves under the H-dynamics interrupted by N equally spaced projective von Neumann measurements of S, then the probability P(N) for finding S at time t is about $1 - (D(H,S)t)^2/N$, where D(H,S) is the uncertainty of H in S. P(N) tends to 1 when N tends to infinity.

The neural correlate for 'holding in place a template' is a well-studied function of recurrent networks in cortex. Why should a neurobiologist who is interested in the implementation of voluntary control in the prefrontal cortex believe that the QZE operates in these circuits in tiny gates as in Eccles 'theory', while the same short-term memory operation can be perfectly well carried out in conventional neural networks (e.g. Hopfield, 1982)? In section 4 we will discuss a realistic classical dynamical model of a frontal recurrent network, which can hold templates in time and space.

In this section we have summarized the contributions of three well-known and respected scientists. In particular, we are deeply touched by the religious engagement which Eccles has expressed in his last writings (Wiesendanger 2006). In the published literature we have found many publications (see e.g. Tuszinski 2006) about the relation of QM and HBF, many of which cast serious doubts on our refereeing system. Thus, it is entertaining to see that quantum theory can even arise from consciousness rather than the other way around (Manousakis 2006)!

3. Lessons from Quantum Computation

In the foreseeable future, QM will not give interesting predictions about HBF. The reason is that by decoherence relevant observables of individual neurons, including electro-chemical potentials and neurotransmitter concentrations, obey classical dissipative equations of motion. Thus, any quantum superposition of states of neurons will be destroyed much too quickly for the subject to become conscious about the underlying QM. In Zurek's (2003) formulation of environment-induced superselection ('einselection'), the preferred basis of neurons becomes correlated with the classical observables in the laboratory. Our senses did not evolve for the purpose of verifying QM. Rather,

they were shaped by the forces of natural selection for the purpose of predicting the world. Thus, as QM is fundamentally stochastic, only quantum states that are robust in spite of decoherence, and hence are effectively classical, have predictable consequences. There is little doubt that in the wet and warm brain einselection is important for explaining the transition from QM to the classical. Decoherence destroys superpositions - the environment induces effectively a superselection rule that prevents certain superpositions to be observed, and only states that survive this process become classical (Schlosshauer 2006). However, since at low temperatures there exist macroscopic, long-lived entangled quantum states in certain physical systems, a rigorous understanding of the classical limit is missing. Arguments about quantum measurement and einselection for 'everyday' objects with on the order of 10^{24} particles (Leggett 2002) are based on highly simplified models with very few degrees of freedom of the reservoirs and interactions (Hepp 1972, Blanchard and Olkiewicz 2003). The controversy between (Tegmark 2000) and (Hagan et al. 2002) is symptomatic: Here the estimated decoherence times within microtubules vary by about 10 orders of magnitude, both based on the same approximate one body scattering picture of decoherence (Tegmark 1993). For an alternative view on the nature of the quantum measurement problem see Leggett's thoughtful 2002 review article and his chapter in this book.

Lacking a quantitative understanding of the border between QM and classical physics, it is therefore better to turn to hard experimental facts and abstract computational theory to estimate the importance of QM for HBF.

Quantum computation and information theory are active areas of research and are treated in many reviews and textbooks (e.g. Nielsen and Chuang 2000). This large body of work in the last two decades offers two sobering conclusions. The first lesson is that only a few quantum algorithms are known that are more efficient for large computations than classical algorithms (Shor 2004). Most of the excitement in the field flows from Shor's (1997) quantum algorithm for factoring large integers for data encryption (a problem quite remote from the brain's daily chores). A second, much more modest, speedup when moving from classical to quantum bits is associated with Grover's (1997) search algorithm. In the last decade, no other quantum algorithms of similar power and real world applicability have been found. Applications of quantum computing to cryptology and to the simulation of quantum systems are very interesting, but of no importance for understanding HBF.

The second lesson is that it is very hard to implement quantum computations. In its simplest version, a quantum computer transforms a state of many two-dimensional quantum bits (Qbits*)*

using a unitary mapping via a sequence of externally controllable quantum gates into a final state with probabilistic outcome. Quantum computation seeks to exploit the parallelism inherent in the entanglement of many Qbits by assuring that the evolution of the system converges with near certainty to the computationally desirable result. To exploit such effects, the computational degrees of freedom have to be isolated sufficiently well from the rest of the system. However, coupling to the external world is necessary for preparation of the initial state (the input), for the control of its time evolution and for the actual measurement (the output). All of these operations introduce decoherence into the computation. While some decoherence can be compensated for by redundancy and other fault-tolerant techniques, too much is fatal. In spite of an intensive search by many laboratories, no scaleable large quantum computing systems are known. The record for quantum computation is the factoring of the number 15 by liquid state NMR techniques (Vandersypen et al. 2001). Qbits and a set of universal quantum gates have been proposed in many different implementations, but all solutions have serious drawbacks: photons fly with the velocity of light and interact weakly with one another, nuclear spins in individual molecules are few in number and so are trapped electrons, atoms, ions or Josephson Qbits in present devices. Nanotubes, in particular, have been studied intensively in mesoscopic physics, but no quantum-coherent states in internal regions of microtubule cylinders have been found which could implement the (Hameroff and Penrose 1996) quantum process. This paints a desolate picture for quantum computation inside the wet and warm brain.

4. Classical Theories of Higher Brain Functions

Computational neuroscience is a thriving field, partially populated by (ex)-physicists, that seeks to explain how low- and high-level brain functions are implemented by realistic networks of neurons. Theories of brain functions are different from those of physics, because they are exploring the blueprint of huge (e.g. the average human brain has upwards of 10^{11} neurons with perhaps 10^{14}-10^{15} synapses that themselves contain hundreds of copies of about one thousand different proteins, all of which are assembled in an aqueous environment) special-purpose devices, determined by evolution and learning, that evolved during tens of millions of years, using bags full of tricks. On the cellular level, the theory by (Hodgkin and Huxley 1952) of voltage-dependent processes across excitable cell membranes successfully describes the operations underlying electrical activity in individual neurons (Koch 1998). By a good choice of irreducible components – macroscopic, deterministic and

continuous membrane currents - this theory provides an excellent connection to the underlying molecular level – microscopic, stochastic and discrete ionic channels - and to the local circuit level above. Some neuroscientists believe that building realistic cortico-thalamic circuits on the basis of neuroanatomy and the Hodgkin-Huxley theory is the ultimate framework on which to build cognitive neuroscience. Others (e.g. Churchland 2002), however, think, as we do, that between the realistic microcircuit level and the cognitive level a theory of neural systems is necessary, which describes the specific contributions of multiple cortical and subcortical areas to HBF. In this section we shall discuss three recent examples of such theories.

Rapid object recognition in the ventral stream of visual cortex.

Visual recognition is computationally difficult. Computer (machine) vision is only now, 40 years after its first halting steps of automatically detecting edges in photos, in a position to begin to deal with recognition of real objects under natural conditions. One popular approach, termed neuromorphic vision, takes its inspirations from the architecture of biological vision systems, in particular those of the fly and of the primate.

There is compelling physiological evidence that object recognition in the cortex of monkeys and humans is mediated by the so-called ventral or 'what' visual pathway. It runs from primary visual cortex (V1) at the back of the brain to visual areas V2 and V4 to inferotemporal cortex (IT), and beyond. Neurons along the ventral stream show an increase in receptive size as well as in the complexity of their preferred stimuli (features). At the top of the ventral stream, cells are tuned to complex stimuli such as faces.

Hubel and Wiesel (1965) discovered in V1 so-called simple and complex neurons with small receptive fields (the receptive field of a neuron is the region in visual space from which the neuron can be excited, colloquially "that it can see"). They found that complex neurons tend to have larger receptive fields, respond to oriented bars or edges anywhere within their receptive fields (shift invariance) and are more broadly tuned than simple cells to spatial frequency (scale invariance). Hubel and Wiesel postulated that complex cells are built up from simple cells by a pooling operation.

Poggio and collaborators (Riesenhuber and Poggio 1999, Serre et al. 2005, 2007) have developed a realistic account of the ventral stream that accounts for the type of very rapid (i.e. in a single glance, i.e. < 200 msec) recognition of objects in images that humans are capable of (Thorpe *et al.* 1996). It is a hierarchical model based on simple and complex neurons in V1 and their counterparts in V2 and V4 and is organized in a series of layers of networks, hooked together in a feed-forward manner. The neurons in these networks are described as linear filters and are built out of neurons in previous layers by combining position- and scale-tolerant edge detectors over neighbouring positions and multiple orientations followed by a nonlinear, pooling operation (computing the maximum over all synaptic inputs to the cell). These elementary computational operations are all biophysically plausible. The output of the highest stage is feed into a linear classifier (which can easily be implemented as a thresholded sum of weighted synaptic input). The trained network behaves similarly to humans when confronted with a natural scene that may or may not contain an animal. Humans and this hierarchical network can perform this routine two–alternative forced choice task at comparable levels of performance. The trained network actually outperforms several state-of-the-art machine vision systems on a variety of image data sets including many different visual object categories (Serre *et al.* 2007). The network is capable of learning to recognize new categories (e.g. cars, animals, faces) from examples. Since the source code of this network is available in Matlab and can be compactly described by a set of mathematically simple steps, the theory is 'understandable' and invites extensions.

Models such as Poggio's constitute a very suggestive plausibility proof for a class of feed-forward models of object recognition. It has been successfully tested against firing patterns of neurons in the upper stages of the visual processing hierarchy (area IT) in the alert monkey (Hung et al. 2005). Such networks are steps towards a quantitative theory of visual perception. They illustrate well the desired characteristics of a classical theory of HBF, namely multi-area interaction, biological realism, and realistic performance on real images. All steps are specified in detail and can be implemented by known biophysical mechanisms without invoking any quantum effects.

A microcircuit of the frontal eye fields (FEF).

The way we see the world is strongly influenced by where we look. Only within a small region of the retina, the fovea, can we resolve fine details of the visual input, to which we direct our gaze mainly by a form of rapid eye movements called saccades. When we look at a newspaper, we move our eyes using different strategies: We can scan the page for pictures or head-lines, fixate on an

article, start to read or move on. How does the brain flexibly and reliably transform a visual input from the retina into commands to the eye muscles to tell them to saccade to a particular location according to specific rules? The voluntary control of saccadic eye movements in the foveal scanning of a visual scene is highly sophisticated. Not only can saccades be made to the most salient target (‘visual grasp reflex’), but during reading they are also influenced by top-down rules, e.g. in word backtracking (‘anti-saccade’) or skipping (‘countermanding’). The frontal eye fields (FEF) in cortex are prominently involved in all these saccade-related tasks.

Neuroanatomy shows a striking uniformity throughout the cortex, while physiology has implied many cortical areas with various functions. The cerebral cortex is a six-layered structure with a clear connectivity pattern of excitatory and inhibitory neurons, which has been abstracted by (Douglas and Martin 1991) into a 'canonical microcircuit' model. Key to this basic circuit is that the input is amplified by excitatory feedback in a ‘smart’ way such that the signal is enhanced and interpreted at the expense of noise. Quantitative estimates about the connectivity in cortical area V1 have recently been worked out by (Binzegger *et al.* 2004). It is a challenge to confront these data and concepts with an important cortical task, the transformation from vision to saccades in the FEF.

The microcircuit model of the FEF by (Heinzle 2006, Heinzle *et al.* 2007) implements the main steps of the saccade-generated computations. These start with a representation of the visual saliency of the image in the layer 4 – provided by input from earlier visual areas - to visuo-motor intention in layers 2/3, to premotor output in layer 5, to the interpretation of rules in layer 6 to choose between fixation, or saccading to a salient target or execution of a ‘cognitive’ reading pattern. For simulation speed, the visual image and the premotor layers are represented by one-dimensional arrays of spiking (integrate-and-fire) neurons. The network has many recurrent connections, with competition between neurons carrying saliency signals and those responsible for recognition of complex patterns. This competition generates realistic saccadic patterns, in particular during reading, which has been carefully studied and phenomenologically modeled by psychologists (e.g., Rayner 1998). In the model, the neuronal firing patterns for the experimentally well-studied excitatory and fixation neurons resemble those found in singe cell neurons in monkey FEF by (Goldberg and Bruce 1990) and (Sato and Schall 2003). The model makes specific predictions about the firing pattern of inhibitory interneurons, cells that are difficult to observe due to their relative low numbers and small size. In principle, the letter recognition input could be based on the Riesenhuber and Poggio (1999) model of the ventral stream discussed above.

Variants of such networks can be adapted to mimic cognitive control of emotions, as in the experiments by Ochsner *et al.* (2002) without invoking the Quantum Zeno effect.

These two examples of conventional computational neuroscience models demonstrate how hitherto mysterious HBF could be instantiated by neural networks of thousands of realistic neurons. The extent to which they are actually implemented in this matter remains for future research to elucidate. Yet the larger point is that there appears to be little need to invoke implausible, macroscopic QM effects for their solution.

Towards classical models of conscious perception.

Intrepid students of the mind point to qualia, the constitutive elements of consciousness, as the ultimate HBF. The subjective feelings associated with redness of red or the painfulness of a toothache are two distinct qualia. Since it remains mysterious how the physical world gives rise to such sensations, maybe one of the more flamboyant interpretations of QM explains qualia and their ineffable qualities and, therefore, consciousness.

Fortunately, the problem of consciousness and its neuronal correlates is beginning to emerge in outlines. The content of consciousness is rich and highly differentiated. It is associated with the firing activity of a very large number of neurons, spread all over the cortex and associated satellites, such as the thalamus. Thus, any one conscious percept or thought must be expressed by a wide-flung coalition of neurons firing together. Even if quantum gates do exist within the confines of neurons, it remains totally nebulous how information of relevance to the organism would get to these quantum gates and how this information would be kept in a coherent quantum-state across the milli- and centimeters separating individual neurons within the cortical tissue, when synaptic and spiking processes, the primary means of neuronal communication, destroy quantum information on the perceptual time-scale of hundreds of milliseconds. At the end of a recent discussion (Koch and Hepp 2006) we proposed a *Gedankenexperiment* to test a possible link between QM and HBF.

The main intention of this section is to provide at least one classical *framework* of consciousness (Crick and Koch 2003), not, however, a *theory* of consciousness. The framework should organize a wide range of phenomena related to visual awareness, incorporate low-level visual areas and more cognitive, high level cortical areas in a semi-realistic manner as a network of spiking neurons. For pedagogical purposes, we will briefly consider the global workspace model of Dehaene *et al.* (2000)

and Dehaene and Changeux (2005). The model simulates the attentional blink (AB), a classical perceptual phenomenon: Participants are asked to detect two successive targets, T1 and T2, in a stream of letters (say a red 'X' following the occurrence of a 'O'). If the two targets follow each other very closely or are timed far apart, T2 can be detected with ease. If, however, T2 is presented between 100 ms and 500 ms after T1, the ability to report T2 drops, as if the subject's attention had 'blinked'. A two-stage model is the most favored account of the AB (Chun and Potter, 1995): in the first stage, items presented in a rapidly flashed sequence of letters or images (Einhäuser, Koch and Makeig 2007) are rapidly recognized and (coarsely) categorized, but are subject to fast forgetting. If a target is detected in the first stage, a second, slower, and limited-capacity stage is initiated. When T2 directly follows T1, both targets enter the second stage. But if T2 falls within the period of the AB, it is processed in the first stage, but no second stage processing is initiated since this stage is still occupied with processing T1. Thus, the neural representation for T2 decays. The two-stage concept of the AB has recently found support in event-related potentials (Kranczioch, Debener and Engel 2003) and in functional brain imaging (Marois, Yi and Chung 2004). It appears in this and many other experiments that conscious and non-conscious visual processing follow at first similar routes, but diverge at some point in an all-to-none manner, leading to different dynamical brain states. During conscious processing, various pieces of information about the stimulus, computed locally in different areas of cortex, become available for explicit report and flexible manipulation.

In the global neuronal workspace framework (Baars 1989, Dehaene and Naccache 2001), conscious processing crucially involves a set of 'workspace neurons' which work in synergy through long-distance reciprocal connections. These neurons, which can access sensory information, maintain it on-line and make it available to other areas, are distributed in the brain, but are most numerous in fronto-parietal and inferotemporal areas (Crick and Koch 2003, Lamme 2003). In this framework the AB finds a natural explanation. The first stage of processing corresponds to the 'feedforward sweep' of activity (as in the model by Poggio and collaborators). These regions then receive feedback from higher areas through recurrent connections, leading to contextual modulations in the lower areas and a rapid globalization of the stimulus, with amplification through reciprocal connections. Ultimately this would lead to the global 'ignition' of a broad set of workspace neurons, from sensory to fronto-parietal areas to areas implicated in verbal report or motor control. In the model, powerful inhibition prevents most workspace neurons from firing, while only a subset of workspace neurons exhibit sustained activity. It is this state of global availability that is postulated to be what is conscious in a perceptual process. The model postulates that the 'phase transition' to the conscious perception of a stimulus is possible only if vigilance (i.e. arousal) is sufficiently high. The transition from sleep to

wakefulness by neuromodulators is an obvious enabling condition for conscious processing of sensory stimuli. The first phase transition between alertness and light sleep can be dramatically seen in the eye movement system (see e.g. Henn *et al.* 1984). A weak stimulus should win by directed attention. By 'fatigue' a population of ignited workspace neurons should decrease their activity and allow other groups to access consciousness in an all-or-none manner. In the papers by Dehaene and collaborators, this model has been partially implemented in a biologically realistic and well-documented network. Aficionados are invited to self-reference their brains!

Although these dynamical ideas organize quite well the phenomenology of different levels of consciousness (attention, un-consciousness, and consciousness) and lead to a number of interesting predictions (Dehaene *et al.* 2006), most of the major questions remain, some old and philosophical (such as the nature of qualia, is free will an illusion, the Freudian unconscious, evolutionary efficiency) and some new and testable (proportion of workspace neurons in V1, explicit or implicit representations, unconscious homunculus in prefrontal cortex…) (see Koch 2004). We are not claiming that this model is correct (for an alternative quantitative computational approach, see Tononi 2004). The purpose in discussing this particular implementation of global workspace is that it demonstrates how today's consciousness research takes serious the challenge of mapping subjective feelings and percepts onto brain structures using purely classical neuronal events and elements.

5. Conclusion

Although we have, hopefully, convinced our physics colleagues that classical physics is the superior framework for explaining HBF, we hurry to stress that on the molecular and membrane level there are beautiful biophysical problems where the border between quantum and classical physics has to be drawn. One of us has started a program of finding NCC neurons in genetically modified mice trained on aversive associative conditioning (Han *et al.* 2004) and hopes to characterize experimentally the NCC neurons. The nature of qualia, e.g. the 'MY RED', have not been explained, but e.g. the self-referential 'MY red' is part of a wonderful story not only of perception, but also of what I am going to do about it. To be conscious means to tell to oneself stories which allows us to function better in reality. Dysfunctions in the representation of the self lead to major psychiatric diseases. To understand one's self will help others.

The writing of this manuscript was supported by the NSF, the NIMH, the Swartz Foundation, and by the Moore Foundation. We thank Jürg Fröhlich and Tony Leggett for their thoughtful comments on the manuscript.

6. References

BJ Baars (1988) *A Cognitive Theory of Consciousness*, Cambridge University Press, Cambridge

U Becherer and J Rettig (2006) *Cell Tiss Res* **326:** 393-407

F Beck and JC Eccles (1992) *Proc. Natl. Acad. Sci. USA* **89:** 11357-11361.

F Beck and JC Eccles (2003) in *Neural Basis of Consciousness*, N Osaka (Ed), J Benjamins Publ Co, Philadelphia PA

T Binzegger, RJ Douglas, and KAC Martin (2004) *J Neuroscience* **24:** 8441-8453

P Blanchard and R Olkiewicz (2003) *Rev Math Phys* **15:** 217-243

MM Chun and MC Potter (1995*) J Exp Psychol Human Perception & Performance* **21:** 109-127

PS Churchland (2002) *Brain-Wise: Studies in Neurophilosophy*, MIT Press, Cambridge MA

FHC Crick and C Koch (2003) *Nature Neurosci* **6:** 119-126

S Dehaene and L Naccache (2001) *Cognition* **79:** 1-37

S Dehaene, C Sergent and J-P Changeux (2003) *Proc Natl Acad Sci USA* **100:** 8520-8525

S Dehaene and J-P Changeux (2005) *PLoS Biology* **3:** 910-27.

S Dehaene, J-P Changeux, J Sackur and C Sergent (2006) *Trends Cogn Sci* **10:** 204-211

RJ Douglas and KAC Martin (1991) *J Physiol (Lond)* **440:** 735-769

JC Eccles (1994) *How the Self Controls its Brain*, Springer, Berlin

W Einhäuser, C Koch and S Makeig (2007) *Vision Res.* **47:** 597-607

K Gottfried and T-M Yan (2003) *Quantum Mechanics: Fundamentals*, Springer, Berlin

LK Grover (1997) *Phys Rev Lett* **79:** 325-328

ME Goldberg and CJ Bruce (1990) J Neurophysiol **64:** 489-508

S Hagan, SR Hameroff and JA Tuszynski (2002) *Phys Rev E* **65:** 061901

H Haken (1970) *Laser Theory*, Handbuch der Physik XXV/2C, Springer, Berlin

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

CJ Han, CM O'Tuathaigh, L van Trigt, JQ Quinn, MS Franselow, R Mongeau, C Koch and DA Anderson (2004) *Proc Natl Acad Sci USA **100:*** 13087-13092

J Heinzle (2006) *A model of the local cortical circuit of the frontal eye field*, Thesis ETH Zürich

J Heinzle, K Hepp and KAC Martin (2007) *J Neuroscience*, submitted

V Henn, RW Baloh and K Hepp (1984) *Exp Brain Res* **54:** 166-176

K Hepp (1972) *Helv Phys Acta* **45:** 236-248

K Hepp and EH Lieb (1975) The laser: a reversible quantum dynamical system with irreversible classical macroscopic motion. In: J Moser (Ed.) *Dynamical Systems, Theory and Applications*, Lecture Notes in Physics, Vol 38, Springer, Berlin

K Hepp (1998) Toward the demolition of a computational quantum brain. In *Quantum Future* (P Blanchard and A Jadczyk, Eds), *Lecture Notes in Physics,* **517:** 92-104

AL Hodgkin and AF Huxley (1952) *J Physiol (Lond)* **11**: 500-544

J Hopfield (1982) *Proc Natl Acad Sci USA* **79:** 2554-2558

D Hubel and T Wiesel (1965) *J Neurophysiology* **28:**229-289

CP Hung, G Kreiman, T Poggio and JJ DiCarlo (2005) *Science* **310:** 863-866

E Joos, HD Zeh, C Kiefer, D Guilini, J Kupsch and I-O Stamatescu (2003) *Decoherence and the Appearance of a Classical World in Quantum Theory*, Springer, Berlin

C Koch (1998) *Biophysics of Computation: Information Processing in Single Neurons*, Oxford University Press, Oxford

C Koch (2004) *The Quest for Consciousness: A Neurobiological Approach*, Roberts, Englewood, Colorado

C Koch and K Hepp (2006) *Nature* **440:** 611-612

C Kranczioch, S Debener and AK Engel (2003) *Brain Research. Cognitive Brain Res* **17:** 177-187

VAL Lamme (2003) *Trends Cogn Sci* **7:** 12-18

HC Lau, RD Rogers and RE Passingham (2006) *J Neuroscience* **26:** 7265-7271

HC Lau, RD Rogers and RE Passingham (2007) *J. cogn. Neurosci.* **19:** 1-10.

AJ Leggett (2002) *J. Phys. Condens. Matter* **14:** R415-R451.

B Libet, EW Wright jr, B Feinstein and DK Pearl (1979) *Brain* **102:** 193-224

B Libet (2004) *Mind Time. The Temporal Factor in Consciousness*. Harvard University Press, Cambridge MA

E Manousakis (2006) *Found Phys* **36:** 795-838

R Marois, DJ Yi and MM Chung (2004) *Neuron* **41:** 465-472

ND Mermin (2003) *Am J Phys* **71:** 23-30

B Misra and ECG Sudarshan (1977) *J Math Phys* **18:** 756-763

MA Nielsen and IL Chuang (2000) *Quantum Computation and Quantum Information,* Cambridge University Press, Cambridge

KN Ochsner, SA Bunge, JJ Gross, and JD Gabrieli (2002) *J Cognitive Neuroscience* **14:** 1215-1229

R Penrose (1994) *Shadows of the Mind*, Oxford University Press, Oxford

KR Popper and JC Eccles (1997) *The Self and its Brain*, Springer, Berlin

K Rayner (1998) Psychological Bull **124:** 371-422

M Riesenhuber and T Poggio (1999) *Nature Neuroscience* **2:** 1019-1025

TR Sato and JD Schall (2003) *Neuron* **38**: 637-648

M Schlosshauer (2006) *Ann Phys* **321:** 112-149

AU Schmidt (2003) *J Phys A* **36:**1135-1148

T Serre, A Oliva and T Poggio (2007) *Proc Natl Acad Sci USA*, in press

T Serre, L Wolf, S Bileschi, M Riesenhuber and T Poggio (2007) *IEEE Trans Pattern Recognition Machine Intelligence* **29(3):** 411-426.

PW Shor (1997) *SIAM J Computation* **26:** 1484-1509

PW Shor (2004) *Quantum Information Processing* **3:** 5-13

JM Schwartz, HP Stapp, and M Beauregard (2005) *Phil Trans R Soc B* **360:** 1309-1327

S Thorpe, D Fize and C Marlot (1996) *Nature* **381:** 520-522

HP Stapp (2003) *Mind, Matter and Quantum Mechanics*, Springer, Berlin

M Tegmark (1993) *Found Phys Lett* **6:** 571-589

M Tegmark (2000) *Phys Rev E* **61:** 4194-4206

G Tononi (2004) *BMC Neuroscience* **5:**42

JA Tuszinski (Ed.) (2006) *The Emerging Physics of Consciousness,* Springer, Berlin

LMK Vanderypen, M Steffen, G Breyta, CS Yannoni, MH Sherwood and IL Chuang (2001) *Nature* **414:** 883-887

J von Neumann (1932) *Mathematische Grundlagen der Quantenmechanik*, Springer, Berlin (English translation: *Mathematical Foundations of Quantum Mechanics,* Princeton University Press, Princeton, 1955)

M Wiesendanger (2006) *Prog Neurobiology* **78:** 304-321

EP Wigner (1967) In: *Symmetries and Reflections*, Indiana Univ Press, Bloomington IN

Zurek (2003) *Rev Mod Phys* 7**6:** 715-775