October 8, 2001 LBNL-xxxxx

The Emergence of Consciousness *

Henry P. Stapp
Lawrence Berkeley National Laboratory
University of California
Berkeley, California 94720

Abstract

ABSTRACT

*This work is supported in part by the Director, Office of Science, Office of High Energy and Nuclear Physics, Division of High Energy Physics, of the U.S. Department of Energy under Contract DE-AC03-76SF00098

Introduction

It is widely believed by both scientists and philosophers that consciousness, as we experience it, was not always present in this universe, but emerged gradually from a more purely physical stratum in conjunction with the development of biological systems, and, in particular, nervous systems. But if one assumes that the physical foundation from which consciousness emerged is adequately described by classical physical theory then one is put in a quandry by the deterministic character of that theory. For the dynamical completeness of classical physical theory entails that either the emerging structure causes actual violations of the deterministic laws or that this emerging structure makes no actual difference to the course of physical events. The second option seems implausible, because then there could be no feed back that would allow survival considerations to lead to the gradual development and improvement of this emerging structure of the basis of its fitness. And the close coordination that we can recognized in our own conscious process and the capacities of our bodies to behave in ways that can actualize our mental intentions would have no natural explanation. On the other hand, the laws of classical physics are so tightly and beautifully woven into a rigid rational structure that the idea that something can step in from the outside and simply over-ride them seems ad hoc and unnatural.

We know, of course, that classical physical theory is basically invalid: it has been replaced at the fundamental level by quantum field theory. The indeterministic character of this latter theory obviously opens up the possibility for the gradual emergence of structures that would act strictly within the bounds allowed by quantum theory, and yet exhibit the two-way causal connection between the emerging structures and the underlying physical organism that would permit a rationally understandable development of an emergent structure from rudimentary beginnings to human-type consciousness.

A misunderstanding that has blocked the development of this natural idea has been the notion that the indeterminism of quantum theory is exclusively random. There is indeed one aspect of this quantum indeterminism that is intrinsically random, and any tampering or biasing of that randomness would destroy the very fabric of quantum theory, and replace it by some ad hoc and arbitrary mutant that would not be quantum theory at all, but some mutilated structure that would lack the cohesiveness and beauty that places quantum theory apart from all pretenders.

However, there is a second, and very important, element of indeterminism in quantum theory, and this second element is perfectly suited to accomodate an emerging structure that could arise from the physical organism, and maintain a two-way causal connection to it within the strict confines of quantum field theory. This second element of indeterminism is the general form of the freedom that is called, within quantum theory, the freedom of the experimenter to choose which aspect of nature he wants to probe, or, at a more impersonal level, the indeterminism of the "basis". This "basis problem" is the root of the controversies about how quantum theory should be "interpreted".

The thesis of this paper is that this second element of indeterminism of quantum theory is precisely the ingredient of quantum theory that underlies the process of the emergence of proto-consciousness in the primitive organisms, and then the gradual evolution of such rudimentary systems with proto-consciousness into higher forms such as thinking human beings.

The Principles

The main first principle is that "feelings" are primordial ingredients of nature. They are causally connected to the physical aspects of nature, but are not reducible to, or re-expressible in terms of, the physical aspects. This is in line with Wm. James's position that it is logically possible for experiential-type realities to be prior to any observing entities that "experience" them. Indeed, since we are aware of, or know, the "self", that self must be *within* thought, and part of it, not something that stands behind the thought, or the act of knowing [See MMandQM, p. 21].

The "self" that appears in human experience is assumed, in this theory, to be a slowly changing 'fringe' part of a stream of human conscious experience [MMandQM p. 160]. This "self" is perhaps ubiquitous in human experience, but the concept of a "knower" is asserted to be inessential to the primordial concept of a"feel". Each quantum collapse has a "feel", or perhaps one should say a "proto-feel", regardless of whether there is a larger "feel" that feels, or knows, both the feel and a "self" that "feels" it.

The second key element is the body-world scheme [MMandQM p. 41, 150ff], which is the (human) brain's representation of the body it controls, and the world around that body. There are the "current", "projected", and "historical" body-world schemata. The first one is the system's representation of the current state of the body and world, the second is the system's representation of an intended state, and the third one is the system's representation of the history known to it. These are described in MMandQM.

The third key ingredient is the "top-level process", which is called the top-level 'code' in MMandQM. It accesses and acts upon, the body-world schemata.

The problematic element in quantum theory is the choice of "basis". This choice can be reduced to the choice of projection operator P. In the Copenhagen interpretation this choice is done by the human experimenter. I carry this idea over to the von Neumann-Wigner formulation of quantum theory. In order to have a simple definite model I assume that the operator P(t) is defined to be the possible P that maximizes $TrPS(t)$, where the 'possible' P's are those such that the collapse $S(t) \longrightarrow PS(t)P$ has a definite "feel". The operator P(t) is presented to nature if and only if the system gives its 'consent'. This consent is either given or withheld on the basis of an evaluation of the feel of P(t).

I have conjectured that the operator P acts on the low frequency electromagnetic field, and in particular on the "coherent state" part of that field, and more specifically on the coulomb part of that field in the radiation gauge, in the rest frame of the cosmic background radiation. Then the "feel" of the component picked out by P(t) is essentially the feel of a "quasi classical" state consisting of a host of 'oscillators' distributed over the body/brain of the individual: the 'feel' would be like the 'feel' of a symphony orchestra.

The evaluation of this 'feel', and the action of collapsing the state $S(t)$ to $PS(t)P$, or to $(1-P)S(t)(1-P)$, requires that the state and the 'feel' be aspects of some enveloping reality that has the capacity to do this evaluation and action. No attempt is made in this model to describe this enveloping reality: the model simply describes the rules of operation at the level of the dynamics of the quantum state S(t). The question of exactly what it is that makes the state behave in the way it does is like the question of why Newton's laws hold. Thus the model no more aspires to be a complete description of the totality of reality than Newton's theory was intended, in Newton's view, to be the complete and final last word of science: it strives to be a scientific theory with significant explanatory power.

This is the outline of my vN/W QT model of human mind-brain dynamics. The question here is how this model of the human mind-brain dynamics fits into a general theoretical structure that can encompass the evolutionary predecessors of human beings, and provide an account of how human consciousness, and the brains that support it, could develop, jointly and naturally, from very simple beginnings, via the quantum dynamics described by this model together with considerations of fitness and survival.

The principal observation here is that there is in the model nothing fundamental that cannot be present also in very simple systems. The essential features of collapse based in 'consent' based on 'evaluation' based on 'feel', can be present also in rudimentary systems. In human beings the collapses act on the top-level process, which accesses and acts upon the body-world schemata, which act upon the motor centers, and upon other low-level brain processes, whereas in a primitive system the collapses act directly to select a course of action. Thus although there is a huge difference as regards the complexity of the state selected by the operator P in the case of human beings as contrasted to a rudimentary system, and hence a huge difference in the quality of the 'feel' of the events associated with the collapse, the basic forms are the same for primitive systems and human mind/brains. So the evolutionary advantage that is given to systems that exploit this quantum process can be effective for both rudimentary and complex systems. In particular, the effect of the quantum collapses—in conjunction with the quantum Zeno effect—of keeping the behaviour of a system focussed on a task can be beneficial for both simple and complex systems.