QUANTUM THEORY AND CONSCIOUSNESS: AN OVERVIEW WITH SELECTED EXAMPLES

HARALD ATMANSCHACHER

Received 27 January 2004 and in revised form 9 February 2004

It is widely accepted that consciousness or, in other words, mental activity is in some way correlated to the behavior of the brain or, in other words, material brain activity. Since quantum theory is the most fundamental theory of matter that is currently available, it is a legitimate question to ask whether quantum theory can help us to understand consciousness. Several approaches answering this question affirmatively, proposed in recent decades, will be surveyed. It will be pointed out that they make different epistemological assumptions, refer to different neurophysiological levels of description, and adopt quantum theory in different ways. For each of the approaches discussed, these imply both problematic and promising features which will be indicated.

1. Introduction

Ilya Prigogine, who died in Brussels in May 2003, was one of the great visionaries of the 20th century with regard to both scientific and humanistic perspectives. His work on nonequilibrium thermodynamics brought him the Nobel prize for chemistry in 1977 [61]. Together with his colleagues in Brussels and Austin, he initiated and achieved major insights concerning topics such as dissipative structures, complex behavior, the arrow of time, the relation between determinism and stochasticity, and other fundamental questions of the sciences. In more recent work, he and his colleagues showed that, to a certain extent, all these topics can be addressed within a generalization of traditional quantum theory, based on the formalism of rigged Hilbert spaces.

In addition, Prigogine was particularly interested in interdisciplinary implications of his scientific work [62, 63] and encouraged research extending beyond the limits of established knowledge. When Ioannis Antoniou, deputy director of the Solvay Institutes in Brussels, asked him for the future challenges of science some time ago, he answered [30]: “if I was a young researcher now, I would study the mind-body problem. This is the great challenge of the 21st century.”

The problem of how mind and matter are related to each other has many facets, and it can be approached from many different starting points. Of course, the historically leading
disciplines in this respect are philosophy and psychology, which were later joined by behavioral science, cognitive science, and neuroscience. In addition, the physics of complex systems and quantum physics have played stimulating roles in the discussion from their beginning.

As regards the issue of complexity, this is quite evident: the brain is one of the most complex systems we know. The study of neural networks, their relation to the operation of single neurons, and other important topics could and will profit a lot from complex systems approaches. As regards quantum physics, the situation is different. Although there can be no reasonable doubt that quantum events happen in the brain as elsewhere in the material world, it is the subject of controversy whether these events are in any way efficacious and relevant for those aspects of brain activity correlated with mental activity.

The original motivation for the pioneers of quantum theory in the early 20th century to relate quantum theory to consciousness was basically philosophical. Quantum theory had introduced an element of randomness standing out against the previous deterministic worldview in which randomness, if it occurred at all, simply indicated our ignorance of a more detailed description (as in statistical physics). In sharp contrast to such epistemic randomness, quantum randomness in processes such as spontaneous emission of light, radioactive decay, or other examples of state reduction was considered a fundamental feature of nature, independent of our ignorance or knowledge. To be precise, this feature refers to individual quantum events, whereas the behavior of ensembles of such events is statistically determined. The indeterminism of individual quantum events is constrained by statistical laws.

It is fairly plausible that conscious free decisions (“free will”) are problematic in a perfectly deterministic world (for more details, see the volume by Kane [42]), so quantum randomness might indeed open up novel possibilities for free will. (On the other hand, randomness is problematic for volition.) In this regard, Prigogine’s approach [62, 63] is particularly remarkable since it strives for a fundamentally stochastic formulation of quantum theory in which randomness does not only enter in state reduction but is the basis of quantum physics and physics as a whole.

Other features of quantum theory, which were found attractive in discussing consciousness issues, are the concepts of complementarity and entanglement. In various ways, pioneers of quantum physics such as Planck, Bohr, Schrödinger, Pauli (and others) emphasized the possible role of quantum theory in reconsidering the old conflict between physical determinism and conscious free will. For broad overviews, see, for example, [20, 70].

In this contribution, some popular approaches for relating quantum theory and consciousness will be surveyed and compared. Section 2 outlines two basically different epistemological options for conceiving relations between material and mental states of systems. Section 3 addresses three different neurophysiological levels of description, to which particular quantum approaches refer. After some introductory remarks, Section 4 sketches the individual approaches themselves: Section 4.2, from von Neumann to Stapp, Section 4.3, from Umezawa to Vitiello, Section 4.4, Beck and Eccles, Section 4.5, Penrose and Hameroff, and Section 4.6, dual-aspect approaches such as those tentatively proposed by Pauli and Jung as well as Bohm and Hiley. Section 5 offers some comparative
conclusions and addresses briefly how Prigogine’s perspective might be inspiring for further work.

2. Epistemological background assumptions

The question of the relationship between the material world and its apparently nonmaterial counterpart or complement is one of the oldest, most puzzling, and most controversial issues in the philosophy and history of science. There exists a vast literature addressing its many different aspects from many different viewpoints. In its most general form, the mind-matter distinction comprises not only the distinction between mind and body but also, even more specifically, that of mind and brain. Such dichotomies have been posed in both epistemological and ontological frameworks, and elaborated in quite a number of variants. They range from a fundamental distinction of mind and matter at a primordial level of description to the emergence of mind (consciousness) from the brain as an extremely sophisticated and highly developed material system. Informative overviews can be found in [21, 54, 59].

One important aspect of all discussions about the relation between mind and matter is the distinction between correlation and causation regarding mental and material states. All of empirical science is based on discovering correlations between states and properties of observed systems. By contrast, the concept of causation (or causality) is used to make sense of correlations. This has been extraordinarily successful. Roughly speaking, correlation is a descriptive term with empirical relevance, while causation is an explanatory term associated with theoretical attempts to understand correlations. Causation implies correlations between cause and effect, but this does not always apply the other way around.

In basic sciences, one typically speaks of causal relations in terms of interactions. In physics, for instance, there are four fundamental kinds of interactions (electromagnetic, weak, strong, gravitational) which serve to explain the correlations that are observed in physical systems. Concerning the mind-matter problem, the situation is more difficult. Far from a theoretical understanding in this field, the existing body of knowledge basically consists of empirical correlations between material and mental states. These correlations are descriptive, not explanatory; they are not causally conditioned. It is (for some purposes) interesting to know that particular brain areas are activated during particular mental activities; but this does not, of course, explain why they are. Thus, it would be premature to talk about mind-matter interactions in the sense of causal relations. For the sake of terminological clarity, the neutral notion of relations between mind and matter will be used in this paper.

In most approaches used to discuss relations between material [ma] brain states and mental [me] states of consciousness, these relations are conceived in a direct way:

\[
[\text{ma}] \leftrightarrow [\text{me}]. \tag{2.1}
\]

This provides a minimal framework to study reduction, supervenience, or emergence relations [44, 73] which can yield both monistic and dualistic pictures. For instance, there is the classical stance of strong reduction, claiming that all mental states and properties can be reduced to the material domain (materialism) or even to physics (physicalism). This
point of view claims that it is both necessary and sufficient to explore and understand the material domain, for example, the brain, in order to understand the mental domain, for example, cognition. More or less, this leads to a monistic picture in which any need to discuss mental states is eliminated right away or at least considered as epiphenomenal. While mind-brain correlations are still legitimate, though causally inefficacious, from an epiphenomenalist point of view, eliminative materialism renders even correlations irrelevant.

Direct relations between mental and material states can also be conceived in a less reductionistic fashion. A number of variants of emergence [73] are prominent examples. Concerning the mind-brain relation, mental states and/or properties can be considered as emergent if the material brain is not necessary or not sufficient to explore and understand them (see [15] for a more detailed discussion of physical examples for contextual conditions in weak property reduction or emergence). This leads to a dualistic picture in which residua remain if one attempts to reduce the mental to the material. Depending on whether the states and properties considered are referred to as such or by corresponding theoretical terms, this dualism can be ontological or epistemological. Within a dualistic scheme of thinking, it becomes almost inevitable to discuss the question of causal influence between mental and material states. In particular, the causal efficacy of mental states upon material states (“downward causation”) has recently attracted growing interest [78].

As an alternative to (2.1), it is possible to conceive mind-matter relations indirectly via a third category:

\[
\begin{array}{c}
[\text{ma}] \\
|mame| \\
[\text{me}]
\end{array}
\]

(2.2)

This third category, here denoted [mame], is often regarded as being neutral with respect to the distinction between [ma] and [me], that is, psychophysically neutral. In scenario (2.2), issues of reduction and emergence concern the relation between the unseparated “background reality” [mame] and the distinguished domains [ma] and [me]. This will be discussed in more detail in Section 4.6.

Such an option, although not much emphasized in contemporary mainstream discussions, has a long tradition. Early versions go back as far as to Spinoza and Leibniz. In the early days of psychophysics in the 19th century, Fechner [24] and Wundt [88] advocated related views. Whitehead, the modern pioneer of process philosophy, referred to mental and physical poles of “actual occasions,” which themselves transcend their bipolar appearances [85]. Many approaches in the tradition of Feigl [25] and Smart [69], called “identity theories,” conceive mental and material states as basically identical “central states,” yet considered from different perspectives. Other variants of this idea have been suggested by Pauli and Jung [9, 11, 40, 50], involving Jung’s conception of a psychophysically neutral, archetypal order, or by Bohm and Hiley [16, 17, 35], referring to an implicate order which unfolds into the different explicate domains of the mental and the material.
From a psychological perspective, Velmans [78] recently presented a similar approach, backed up with empirical material, and Strawson [74] recently proposed a “real materialism” which uses a closely related scheme. And, of course, a most outspoken proponent of “dual aspect” thinking is Chalmers [21], who considers the option that the underlying, psychophysically neutral level of description could be best characterized in terms of information.

Before proceeding further, it should be emphasized that many present day approaches prefer to distinguish between first-person and third-person perspectives rather than mental and material states. This terminology serves to highlight the discrepancy between immediate conscious experiences (“qualia”) and their description, be it behavioral, neural, or biophysical. The notion of the “hard problem” of consciousness research refers to bridging the gap between first-person experience and third-person accounts of it, which can be rephrased in terms of mind-matter research. In the present contribution, mental conscious states are understood essentially as states of first-person consciousness. This does not mean, however, that the problem of how to define consciousness precisely is considered to be resolved. Ultimately, it will be (at least) as difficult to define a mental state in rigorous terms as it is to define a material state.

3. Neurophysiological levels of description

3.1. Neuronal assemblies. A mental system can be in many different conscious, intentional mental states. In a hypothetical state space, a sequence of such states forms a trajectory representing what is often called the stream of consciousness. Since different subsets of the state space are typically associated with different stability properties, a mental state can be assumed to be more or less stable, depending on its position in the state space. Stable states are distinguished by a residence time longer than that of metastable or unstable states. If a mental state is stable with respect to perturbations, it “activates” a mental representation encoding a content that is consciously perceived.

Moving from this purely psychological, or cognitive, description to its neurophysiological counterpart leads us to the question, what is the neural correlate of a mental representation? According to standard accounts (cf. [52] for recent discussion), mental representations are correlated with the activity of neuronal assemblies, that is, ensembles of several thousands of coupled neurons. The neural correlate of a mental representation can be characterized by the fact that the connectivities, or couplings, among those neurons form an assembly confined with respect to its environment, to which connectivities are weaker than within the assembly. The neural correlate of a mental representation is activated if the neurons forming the assembly operate more actively, for example, produce higher firing rates, than in their default mode.

In order to achieve a stable operation of an activated neuronal assembly, there must be a subtle balance between inhibitory and excitatory connections among neurons (cf. Figure 3.1). If the transfer function of individual neurons is strictly monotonic, that is, increasing input leads to increasing output, assemblies are difficult to stabilize. For this reason, recent results establishing a nonmonotonic transfer function with a maximal output at intermediate input are of high significance for the modeling of neuronal assemblies [45]. For instance, network models using lattices of coupled maps with quadratic
maximum [43] are paradigmatic examples of such behavior. These and other familiar models of neuronal assemblies (for an overview see [3]) are mostly formulated in a way not invoking well-defined elements of quantum theory. Exceptions are the approaches by Umezawa, Vitiello, and others (see Section 4.3) and, to some extent, Stapp (see Section 4.2). Otherwise, quantum theory is not claimed to play any role at the neurophysiological level of direct neural correlates of conscious mental states.

### 3.2. Single neurons and synapses

The fact that neuronal assemblies are mostly described in terms of classical behavior does not rule out that classically not describable quantum effects may be significant if one focuses on individual constituents of assemblies, that is, single neurons or interfaces between them. These interfaces, through which the signals between neurons propagate, are called synapses. There are electrical and chemical synapses, depending on whether they transmit a signal electrically or chemically.

At electrical synapses, the current generated by the action potential at the presynaptic neuron flows directly into the postsynaptic cell, which is physically connected to the presynaptic terminal by a so-called gap junction. At chemical synapses, there is a cleft between pre- and postsynaptic cells. In order to propagate a signal, a chemical transmitter (glutamate) is released at the presynaptic terminal. This release process is called exocytosis. The transmitter diffuses across the synaptic cleft and binds to receptors at the postsynaptic membrane, thus opening an ion channel (see [41] and Figure 3.2). Chemical transmission is slower than electric transmission.

A model developed by Beck and Eccles applies concrete quantum mechanical features to describe details of the process of exocytosis. Their model proposes that quantum processes are relevant for exocytosis and, moreover, are tightly related to states of consciousness. This will be discussed in more detail in Section 4.4.

At this point, another approach developed by Flohr [26] should be mentioned, for which chemical synapses with a specific type of receptors, so-called NMDA receptors (the abbreviation NMDA refers to N-methyl-D-aspartate, the synthetic agonist that activates...
3.3. Microtubuli. The lowest neurophysiological level, at which quantum processes have been proposed as a correlate to consciousness, is the level at which the interior of single neurons is considered: their cytoskeleton. It consists of protein networks basically made up of two kinds, neurofilaments and microtubuli (Figure 3.3(a)), which are essential for various transport processes within neurons (as well as other cells). Microtubuli are long polymers usually constructed of 13 longitudinal \( \alpha \)- and \( \beta \)-tubulin dimers arranged in a tubular array with an outside diameter of about 25 nm (Figure 3.3(b)). For more details, see [41, Chapter II.4].

NMDA receptors, are of paramount significance. Briefly, Flohr observes that the specific plasticity of NMDA receptors is a necessary condition for the formation of extended stable neuronal assemblies correlated to (higher-order) mental representations, which he identifies with conscious states. Moreover, he indicates a number of mechanisms caused by anaesthetic agents, which block NMDA receptors and consequently lead to a loss of consciousness. Flohr's approach is physicalistic and reductionistic, but it is entirely independent of any specific quantum ideas.

Figure 3.2. Release of neurotransmitters at the synaptic cleft (exocytosis).

Figure 3.3. (a) Microtubuli and neurofilaments, the width of the figure corresponds to approximately 700 nm; (b) tubulin dimers, consisting of \( \alpha \)- and \( \beta \)-monomers, constituting a microtubule.
The tubulins in microtubuli are the substrate which, in Hameroff’s proposal, is used to embed Penrose’s theoretical framework neurophysiologically. As will be discussed in more detail in Section 4.5, tubulin states are assumed to depend on quantum events so that quantum coherence among different tubulins is possible. As will be described in a more detailed manner in Section 4.5, a crucial thesis in the scenario of Penrose and Hameroff is that the (gravitation-induced) collapse of such coherent tubulin states corresponds to elementary acts of consciousness.

Clearly, the neurophysiological level used in this scenario is far from that of neuronal assemblies addressed in Section 3.1. Therefore, the intended acts of consciousness cannot themselves be conscious mental states in the ordinary sense of a mental representation. Beyond the hypothetical status of the scenario as such, this entails additional questions concerning both the nature of tubulin-level consciousness and its relation to ordinary conscious states.

4. Selected examples

4.1. Ways to adopt quantum theory. In the following, (some of) the better known and partly worked out approaches that use concepts of quantum theory for inquiries into the nature of consciousness will be presented and discussed. For this purpose, the epistemological distinctions (2.1) and (2.2) addressed in Section 2 and neurophysiological distinctions addressed in Sections 3.1, 3.2, and 3.3 will be used as guidelines to classify the respective quantum approaches in a systematic way. However, some preliminary qualifications concerning different possible ways to adopt quantum theory are in order.

There are quite a number of accounts discussing quantum theory in relation to consciousness which adopt basic ideas of quantum theory in a purely metaphorical manner. Quantum theoretical terms such as entanglement, superposition, collapse, complementarity, and others are used without specific reference to how they are defined precisely and how they are applicable to specific situations. For instance, conscious acts are just postulated as interpretable analogously to physical acts of measurement, or correlations in psychological systems are just postulated as interpretable analogously to physical entanglement. Such accounts may provide fascinating science fiction, and they may even be important to inspire nuclei of ideas to be worked out in detail. But unless this detailed work leads beyond pure metaphors and analogies, they do not yet represent scientific progress. Approaches falling into this category will not be discussed in this contribution.

A second category includes approaches that use the status quo of present-day quantum theory to describe neurophysiological and/or neuropsychological processes. Among these approaches, the one with the longest history was initiated by von Neumann in the 1930s, later taken up by Wigner, and currently championed by Stapp. It can be roughly characterized as the proposal to identify intentional conscious acts with physical state reductions. Evidently, such an identification is a much stronger statement than a mere metaphor or analogy. Another fairly early idea dating back to Ricciardi and Umezawa in the 1960s is to treat mental states, particularly memory states, in terms of vacuum states of quantum fields. A prominent proponent of this approach at present is Vitiello. Finally, there is the idea suggested by Beck and Eccles in the 1990s, according to which quantum mechanics is relevant for the description of exocytosis at the synaptic cleft.
The third category refers to further developments or generalizations of present-day quantum theory. An obvious candidate in this respect is the proposal by Penrose to relate elementary conscious acts to gravitation-induced reductions of quantum states. Ultimately, this requires the framework of a future theory of quantum gravity which is far from developed. Together with Penrose, Hameroff has argued that microtubuli might be the right place to look for such state reductions. Another set of approaches is based on generalizations of quantum theory beyond quantum physics proper. In this way, formally generalized concepts such as complementarity and entanglement can be applied to phenomena in both mental and material domains. In particular, relations between the two can be conceived in terms of dual aspects of one underlying “reality.” This conception, drawing on the philosophies of Spinoza and Leibniz, has been considered attractive by 20th century scientists such as Bohr, Pauli, Bohm, Primas, d’Espagnat, and others. Some of the proposed scenarios will be sketched.

4.2. From von Neumann to Stapp: quantum state reductions are conscious acts. The act or process of measurement is a crucial aspect in the framework of quantum theory that has been the subject of controversy for more than seven decades now. In his monograph on the mathematical foundations of quantum mechanics, von Neumann (see [82, Chapter V.1]) introduced, in an ad hoc manner, the projection postulate as a mathematical tool for describing measurement in terms of a discontinuous, noncausal, instantaneous, and irreversible act given by

$$\rho \rightarrow \rho' = \sum_{j=1}^{\infty} \langle b_j | \rho | b_j \rangle \cdot P_{b_j}, \quad (4.1)$$

where $\rho$ is the state of the system before measurement, and the $b_j$’s are the eigenstates of the measured observable. Formula (4.1) expresses that, by measuring an observable $B$ with (nondegenerate) eigenstates $b_j$, the state $\rho$ of the system will be projected (hence $P_{b_j}$) onto one of the basis vectors of the representation in which $B$ is diagonal. The probability for this eigenstate is $|\langle b_j | \rho \rangle|^2$. Recall that von Neumann’s formulation refers to measurements of the first kind. The transition from $\rho'$ to a particular eigenstate $b_j$ is often called the collapse or reduction of the wavefunction, as opposed to the continuous, unitary (reversible) evolution of a system according to

$$\rho \rightarrow \rho(t) = U_t \rho U_t^+ = e^{-i\hbar H t} \rho e^{i\hbar H t}, \quad (4.2)$$

where $H$ is the Hamiltonian of the system and $U_t$ is its unitary time-evolution operator.

In [82, Chapter VI], von Neumann discussed the conceptual distinction between observed and observing systems. In this context, he applied (4.1) and (4.2) to the general situation of a measured object system (I), a measuring instrument (II), and (the brain of) a human observer (III). His conclusion was that it makes no difference for the result of measurements on (I) whether the boundary between observed and observing systems is posited between I and (II & III) or between (I & II) and III. As a consequence, it is inessential whether a detector or the human brain is ultimately referred to as the
“observer.” (Notably, von Neumann’s chain of observing systems stays always in the material domain. When he refers to subjective (mental) experiences, he presupposes some psychophysical parallelism allowing him to treat these experiences as brain processes.)

In contrast to von Neumann’s fairly cautious stance, London and Bauer [46] went much further and proposed that it is indeed human consciousness which completes quantum measurement (see [37] for a detailed account). In this way, they attributed a crucial role to consciousness in understanding quantum measurement—a truly radical position. In the 1960s, Wigner followed up on this proposal in [86] (in his later writings, he repudiated this point of view about the role of consciousness, stating that “it is outside the realm of quantum mechanics” [87]), coining the example of his now proverbial “Wigner’s friend.” In order to describe measurement as a real dynamical process generating irreversible facts, Wigner called for some nonlinear modification of (4.2), replacing von Neumann’s projection (4.1). A critical discussion of the problems of Wigner’s approach toward measurement together with the presentation of alternatives can be found in [66].

Since the 1980s, Stapp has continued the tradition of von Neumann and Wigner in exploring mind-matter issues from the viewpoint of quantum theory. Stapp does not suggest any specific modifications to present-day quantum theory plus the projection postulate. He bases his approach to addressing the “mind–brain interface” [71] on interpretational rather than formal changes of the theory. Inspired by von Neumann, Stapp uses the freedom to place the interface between observed and observing systems and locates it in the observer’s brain. Specifically, he argues that quantum uncertainties at the synaptic level can have effects large enough to generate superpositions of macroscopic patterns of brain activity at the level of neuronal assemblies. The neural correlate of conscious events is assumed to be the collapse of such a superposition into an actualized (activated) neuronal assembly.

In his earlier work [71], Stapp starts with Heisenberg’s distinction between the possible and the actual [33] to endow the operational Copenhagen interpretation of quantum mechanics with an ontology. Heisenberg’s notion of the actual is related to a measured event in the sense of the Copenhagen interpretation. However, Heisenberg’s notion of the possible, of a tendency, relates to the situation before measurement, which reflects the idea of an “objective” reality in an ontological sense. (In another terminology, this is expressed as the distinction of epistemic and ontic descriptions [10]. Margenau’s notion of latent observables [48] and d’Espagnat’s notion of an independent reality [23] represent other ways to achieve an ontological interpretation of quantum theory.)

Immediately after its actualization, each actual event holds the tendency for the impending actualization of another subsequent actual event. Therefore, actual events are by definition ambiguous. With respect to their actualized aspect, Stapp’s essential move is to (see [71, page 149])

attach to each Heisenberg actual event an experiential aspect. The latter is called the feel of this event, and it can be considered to be the aspect of the actual event that gives it its status as an intrinsic actuality.

With respect to their tendency aspect, it is tempting to understand actual events in terms of scheme (2.2). Stapp himself does not state this explicitly, but some of his formulations
suggest a mode of description at which mind and matter are in fact unseparated. This is expressed, for instance, by his notion of a “hybrid ontology” with “both idea-like and matter-like qualities” and “two complementary modes of evolution” (see [72, page 159]).

Another significant aspect of Stapp’s approach is the existence not only of mind-matter correlations but genuine causal interactions. More precisely, he argues that his approach entails the possibility (see [72, page 153]) that “conscious intentions of a human being can influence the activities of his brain.” Concerning the neurophysiological implementation of this idea, conscious mental states are assumed to correspond to quantum collapses of superposition states at the level of macroscopic brain activity. (In detail, Stapp argues that the probabilities $|\langle b_j | \rho \rangle|^2$ for eigenstates $b_j$ after collapse can be mentally influenced. Specifically, the frequency of such intentional acts can protract the lifetime of the state $b_j$ due to quantum Zeno type of effect.) Additional commentary by Stapp concerning the causal efficacy of conscious states on brain states in relation to James’ idea of a holistic stream of consciousness [36] is contained in [72].

After all, a future elaborated version of Stapp’s largely informal, “pragmatic” approach with a satisfactory ontological basis might turn out to be a dual aspect approach (2.2) (cf. Section 4.6) with information rather than matter or mind as the fundamental ontological feature (see [72, page 160]). At a level at which conscious mental states and material brain states are distinguished, each conscious experience, according to Stapp (see [72, page 153]), has as its physical counterpart a quantum state reduction actualizing “the pattern of activity that is sometimes called the “neural correlate” of that conscious experience.”

For further progress, it will be mandatory to develop a coherent formal framework for this approach and elaborate on concrete details. For instance, it is not worked out how precisely quantum superpositions and their collapses are supposed to occur in neural correlates of conscious events. In addition, it is not sufficient to state that quantum theory in its present form can be interpreted in such a way that both mental and material systems as well as causal interactions among them are included. Although Stapp’s conception has interesting and inspiring elements, this basic assumption remains to be backed up with more details. For instance, given the operational importance of statistical states in standard quantum physics, it is hard to imagine how mental states could be evaluated statistically.

Finally, it should be noted that the status of the proposal discussed in this section depends on the acceptance of von Neumann’s projection postulate (4.1). Since all results of quantum physical experiments can be described with the weaker expectation value postulate, (4.1) is not an essential ingredient of quantum theory [66].

4.3. From Umezawa to Vitiello: quantum field theory of mind/matter states. In the 1960s, Ricciardi and Umezawa [67] suggested to utilize the formalism of quantum field theory to describe brain states, with particular emphasis on memory. The basic idea is to conceive memory states, analogously to states of many-particle systems, as inequivalent representations of vacuum states of quantum fields (related proposals with a similar basic idea are due to Fröhlich [27] and Pribram [60]). This proposal went through several refinements (e.g., [38, 75, 76]). Major recent progress was achieved by including effects
of dissipation, chaos, and quantum noise [58, 79]. Readable nontechnical account of the approach in its present form is [80] (and the references therein) and [81].

It should be noted that most discussions of this approach do not consistently distinguish between mental and material states. This suggests reducibility of mental activity to brain activity, within scenario (2.1), as an underlying assumption. In this sense, the proposal addresses the brain as a many-particle system, where the “particles” are more or less neurons. In the language of Section 3.1, this clearly refers to the level of neuronal assemblies, with the major benefit that this is the level which directly correlates with mental activity.

Quantum field theory has been extremely successful in statistical physics, thermodynamics, and condensed matter physics. A crucial distinction between standard quantum mechanics and quantum field theory is that quantum field theory is not restricted to the validity domain of the uniqueness theorem by Stone and von Neumann. This theorem defines the framework for quantum mechanics with finitely many degrees of freedom. It states that all representations of the canonical commutation relations in a noncommutative algebra of observables over a compact state space are unitarily equivalent.

Leaving the domain of validity of this theorem leads to the possibility that infinitely many degrees of freedom (such as the modes of a radiation field) can be included, interactions of open systems with their radiative environment can be treated, irreversible (nonunitary) dynamical evolution can be incorporated, and superselection rules can be formulated to derive classical observables from a noncommutative (irreducible) algebra of observables. Quantum field theory yields infinitely many representations of the commutation relations, which are inequivalent to the Schrödinger representation of standard quantum mechanics.

Such inequivalent representations can be generated by spontaneous symmetry breaking, occurring when the ground state (or the vacuum state) of a system is not invariant under the full group of transformations providing the conservation laws for the system. If the symmetry breaks down, collective modes are generated (so-called Nambu-Goldstone boson modes) which propagate over the system and introduce long-range correlations in it. These correlations are responsible for ordered patterns. Roughly speaking, they represent the quantum field theoretical equivalent of coherent activity in neuronal assemblies.

The activation of a neuronal assembly is necessary to make the encoded content consciously accessible. This activation is considered to be initiated by external stimuli. Unless the assembly is activated, its content remains unconscious, unaccessed memory. Coherent neuronal assemblies correlated to such memory states are vacuum states; their activation leads to excited states with a finite lifetime and enables a conscious recollection of the content encoded in the vacuum (ground) state. The stability of such states and the role of external stimuli were investigated in detail in [75, 76].

In addition to these desirable features of the model, there are problems. One of them is the so-called overprinting problem: in order to store new information, existing vacuum states must be overprinted. Furthermore, the lifetime of vacuum states is in principle infinite, implying that their memory content can never be forgotten. A decisive step removing these problems and providing additional insight was achieved by taking dissipation into account.
Dissipation is possible when the interaction of a system with its environment is considered. In [79], Vitiello described how the system-environment interaction causes a doubling of the collective modes, which yields infinitely many differently coded vacuum states, offering the possibility of many memory contents without overprinting. Moreover, dissipation leads to finite lifetimes of the vacuum states, thus representing temporally limited rather than unlimited memory [1, 2]. Finally, dissipation generates a genuine arrow of time for the system, and its interaction with the environment induces entanglement. In a recent contribution [58], additional effects of chaos and quantum noise are addressed.

Two major merits of the quantum field theory approach are that (i) it avoids the restrictions of standard quantum mechanics in a formally sound way and (ii) refers to a neurophysiological level at which neural correlates of mental representations are in fact to be expected. In these regards, the approach appears to be quite convincing. Conceptually, however, it contains ambiguities demanding clarification, for example, concerning the continuous confusion of mental and material states (and their properties). If mental states are the primary objects of reference, the quantum field theoretical treatment shows, in a very subtle and refined manner, metaphorical features. If quantum field theory is supposed to literally apply to material brain states, it remains largely unclear how this is backed up by the results of contemporary neurobiology.

A related issue is the abstract treatment of the environment in the dissipative model, leaving open what is concretely described. A provocative topic introduced in [80, Section 7.7] is the significance of forward and backward arrows of time due to the doubling of collective modes. In the final chapters of [80], Vitiello argues that this doubling leads to a “time-reversed copy” of brain behavior which might be interpreted as its mental counterpart. In his words, “consciousness seems thus to emerge as a manifestation of the dissipative dynamics of the brain.” Such a scenario ascribes comparable significance to material and mental properties and is evidently nonreductive.

4.4. Beck and Eccles: quantum mechanics at the synaptic cleft. Probably the most concrete and detailed suggestion of how quantum mechanics can play a role in brain processes is due to Beck and Eccles [14], later refined in [13]. It refers to particular mechanisms of information transfer at the synaptic cleft. However, ways in which these quantum processes might be relevant for mental activity, and in which their interactions with mental states are conceived, have hardly been elaborated so far.

As presented in Section 3.2, the information flow between neurons in chemical synapses is initiated by the release of transmitters in the presynaptic terminal. This process is called exocytosis, and it is triggered by an arriving nerve impulse with a probability (much) less than one. In order to describe the trigger mechanism in a formal statistical way, thermodynamics or quantum mechanics can be invoked. A look at the corresponding energy regimes shows [14] that quantum processes are distinguishable from thermal processes for energies higher than $10^{-2} \text{ eV}$ (at room temperature). Assuming a typical length scale for biological microsites of the order of several nanometers, an effective mass below 10 electron masses is sufficient to ensure that quantum processes prevail over thermal processes.

The upper limit of the time scale of such processes in the quantum regime is of the order of $10^{-12}$ second. This is significantly shorter than the time scale of cellular processes,
which is $10^{-9}$ second and longer. The remarkable difference between the two time scales makes it possible to treat the corresponding processes as decoupled from one another.

The detailed trigger mechanism proposed in [14] is based on the quantum concept of quasi-particles, reflecting the particle aspect of a collective mode. Skipping the details of the picture, the proposed trigger mechanism refers to tunneling processes of two-state quasi-particles, resulting in state collapses. It yields a probability of exocytosis in the range between 0 and 0.7, in agreement with empirical observations. Using a theoretical framework developed earlier [39, 47], the quantum trigger can be concretely understood in terms of electron transfer between biomolecules.

As indicated above, the proposal outlined so far is the most empirically concrete and theoretically detailed approach to treating brain processes from a quantum theoretical point of view. However, the question remains how the quantum trigger for exocytosis may be relevant for conscious mental states. There are two aspects to this question.

The first one refers to Eccles’ intention to utilize quantum processes in the brain as an entry point for mental causation. The idea, indicated in Section 1, was that the fundamentally indeterministic nature of individual quantum state collapses offers room for the influence of mental powers on brain states. In the present picture, this is conceived in such a way that “mental intention (volition) becomes neurally effective by momentarily increasing the probability of exocytosis” [14]. Further justification of this assumption is not given.

The second aspect refers to the problem that processes at single synapses cannot be simply correlated to mental activity, whose neural correlates are entire assemblies of neurons. Most plausibly, prima facie uncorrelated, random processes at individual synapses would result in a stochastic network of neurons [34]. Although Beck has indicated possibilities (such as quantum stochastic resonance) to achieve ordered patterns at the level of assemblies from basically random synaptic processes [13], this remains an unsolved problem.

With the exception of Eccles’ idea of mental causation, the approach by Beck and Eccles essentially focuses on brain states and brain dynamics. In his more recent account [13], Beck states explicitly that “science cannot, by its very nature, present any answer to [...] questions related to the mind.” It may open the door to controlled speculation about mind-matter relations, but more cannot be achieved in an exclusively biophysical approach.

### 4.5. Penrose and Hameroff: quantum gravity and microtubuli

In the scenario developed by Penrose and neurophysiologically augmented by Hameroff, quantum theory is claimed to be effective for consciousness, but this happens in an extremely sophisticated way. It is argued that elementary acts of consciousness are nonalgorithmic, that is, non-computable, and they are neurophysiologically realized as gravitation-induced reductions of coherent superposition states in microtubuli.

Other than the approaches in Sections 4.2, 4.3, and 4.4, which are essentially based on (different features of) status quo quantum theory, the physical part of the scenario, proposed by Penrose, refers to future developments of quantum theory. It is assumed that the physical process underlying quantum state reduction, called objective reduction, will remain incomprehensible unless effects of gravitation are included. The grander picture
would be that a full-blown theory of quantum gravity is required to ultimately understand quantum measurement.

This is a far-reaching assumption, and Penrose does not offer a concrete solution to this problem. However, he gives a number of plausibility arguments which clarify his own motivations and inspired others to take his ideas seriously. Penrose’s rationale for referring to state reduction is not that the corresponding randomness offers room for mental causation to become efficacious (although this is not excluded). His conceptual starting point, at length developed in [55, 56], is that elementary conscious acts must be nonalgorithmic. Phrased differently, the emergence of a conscious thought is a process which cannot be described algorithmically, hence cannot be computed. His background in this respect has a lot to do with the nature of creativity, mathematical insight, Gödel’s incompleteness theorem, and the idea of a Platonic reality beyond mind and matter.

In contrast to the unitary time evolution of quantum processes as in (4.2), Penrose suggests that a valid formulation of quantum state reduction replacing (4.1) must faithfully describe an objective physical process. Since present-day quantum theory does not contain such a picture, he argues that effects not currently covered by quantum theory should play a role in state reduction. Ideal candidates for him are gravitational effects since gravitation is the only fundamental interaction which is not integrated into quantum theory so far. Rather than modifying elements of the theory of gravitation (i.e., general relativity) to achieve such an integration, Penrose discusses the reverse: novel features have to be incorporated in quantum theory for this purpose. In this way, he arrives at the proposal of gravitation-induced objective state reduction.

Why is such a version of state reduction noncomputable? Initially, one might think of an objective version of state reduction in terms of a stochastic process. This would certainly be indeterministic, but probabilistic and stochastic processes can be standardly implemented on a computer, hence they are definitely computable. Penrose sketches some ideas concerning genuinely noncomputable, not only random, features of quantum gravity in [56, Sections 7.8 and 7.10]. In order for them to become viable candidates for explaining the noncomputability of gravitation-induced state reduction, a long way still has to be gone.

With respect to the neurophysiological implementation of Penrose’s proposal, his collaboration with Hameroff was decisive. From his background as an anaesthesiologist, Hameroff suggested to consider microtubules as an option where reductions of quantum states can take place in an effective way; see, for example, [32]. The respective quantum states are assumed to be coherent superpositions of tubulin states, and their gravitation-induced collapses are interpreted as individual elementary acts of consciousness. Even if all other problems were solved, the question remains how tubulin-related conscious acts lead to conscious mental representations (correlated to neural assemblies). In this context, it might be a good idea to distinguish ordinary consciousness from a hypothetical, tubulin-related “proto-mentality.”

The idea of focusing on microtubuli was partly motivated by the argument that special locations are required to ensure that quantum states can live long enough to become reduced by gravitational influence rather than by interactions with the warm and wet environment within the brain. Speculative remarks about how the noncomputable aspects
of the expected new physics mentioned above could be significant in this scenario can be found in [56, Section 7.7]. (Interestingly, a recent study showed that gravitation seems to be necessary for the development of an ordered network of microtubuli [53]. But there is no obvious relation between this result and objective state reduction.)

Influential criticism of the possibility that quantum states can in fact survive long enough in the thermal environment of the brain was raised by Tegmark [77]. He estimated the decoherence time of tubulin superpositions due to interactions in the brain to be less than $10^{-12}$ second. Comparing to typical time scales of microtubule processes of the order of milliseconds and more, he concluded that the lifetime of tubulin superpositions is much too short to be significant for neurophysiological processes in the microtubuli. In a response to this criticism [31], it was argued that a revised version of Tegmark’s model provides decoherence times up to the neurophysiologically relevant range of 10 to 100 milliseconds.

However, decoherence is just a tiny piece in the debate about the overall picture proposed by Penrose and Hameroff. From a philosophical perspective, their proposal has occasionally received outspoken rejection; see, for example, [29, 57]. Indeed, their approach collects several top level mysteries, among them the relation between mind and matter itself, the ultimate unification of all physical interactions, the origin of mathematical truth, and the understanding of brain dynamics across hierarchical levels. Combining such deep issues is as ambitious as it is provocative, and it is certainly fascinating.

By and large, the scenario by Penrose and Hameroff represents a highly speculative approach with conceptual problems and without plausible ideas for concrete empirical confirmation. Nevertheless, it is worthwhile to remember Bohr’s bonmot that the question may not be whether a theory is too crazy but whether it is crazy enough. One should not reject this scenario too quickly: it may contain elements that are crazy enough to be true.

4.6. Mind and matter as dual aspects. Dual-aspect approaches consider mental and material domains of reality as aspects, or manifestations, of one underlying reality in which mind and matter are unseparated. In such a framework, the distinction between mind and matter is often considered as a, or even as the, basic tool to achieve epistemic access, that is, gather knowledge. (In a pronounced way, Spencer-Brown proposed such a procedure as the basis of all cognitive activity in [18]: “we take as given the idea of distinction and the idea of indication, and that we cannot make an indication without drawing a distinction.”) Consequently, the status of the underlying, psychophysically neutral domain is considered as ontic relative to the mind-matter distinction.

As mentioned in Section 2, dual-aspect approaches have a long history. As regards quantum theoretically inspired variations on this theme, interesting versions have been proposed by Pauli and Jung [9, 11, 40, 50] and by Bohm and Hiley [16, 17, 35]. In the latter approach, the notions of implicate and explicate orders mirror the distinction between ontic and epistemic domains. At the level of the implicate order, the term active information expresses that this level is capable of “informing” the epistemically distinguished, explicate domains of mind and matter.

At this point, it should be emphasized that the usual notion of information is clearly an epistemic term. Nevertheless, there are quite a number of dual-aspect approaches
addressing something like information at the ontic, psychophysically neutral level. (The pioneering quantum conception in this regard is von Weizsäcker’s ur-theory [83], most prominent is Wheeler’s “it from bit” [84], most recent are proposals by Brukner and Zeilinger [19], Fuchs [28], or Clifton et al. [22]. See also [21] for a discussion of information-based dual aspects.) Using an information-like concept in a nonepistemic manner is inconsistent if the common significance of Shannon-type information is intended, which requires distinctions in order to construct partitions, providing alternatives, in the set of given events. Most information-based dual-aspect approaches do not sufficiently clarify their notion of information so that misunderstandings are almost unavoidable.

While the proposal by Bohm and Hiley is essentially the sketch of a conceptual framework without further details, particularly concerning the mental domain, the suggestions by Pauli and Jung offer some more material to discuss. An intuitively appealing way to represent their approach considers the distinction between epistemic and ontic domains of material reality due to quantum theory in parallel with the distinction between epistemic and ontic mental domains.

On the physical side, the epistemic/ontic distinction refers to the distinction between a “local realism” of empirical facts obtained from classical measuring instruments and a “holistic realism” of entangled systems [10]. Basically, these domains are connected by the process of measurement, so far conceived as independent of conscious observers. The corresponding picture on the mental side refers to a distinction between the conscious and the unconscious. (It is obvious that the term “mental” is here used with a connotation more general than consciousness. Preconscious, subconscious, and unconscious domains, personal and collective, are included as well.) In Jung’s depth psychological conceptions, these two are related by the emergence of conscious mental states from the unconscious, analogously to physical measurement.

In Jung’s depth psychology, it is crucial that the unconscious has a collective component, unseparated between individuals and concerning its contents, the so-called archetypes. They are regarded as constituting the psychophysically neutral level covering both the collective unconscious and the holistic reality of quantum theory. At the same time, they operate as “ordering factors,” being responsible for the arrangement of their psychical and physical manifestations in the epistemically distinguished domains of mind and matter. Much more detailed illustrations of this picture can be found in [9, 11, 40, 50].

This scheme is clearly related to scenario (2.2), combining an epistemically dualistic with an ontically monistic approach. There is a causal relationship (in the sense of formal rather than efficient causation) between the psychophysically neutral, monistic level and the epistemically distinguished mental and material domains. In terms of Pauli and Jung, this kind of causation is expressed by the ordering operation of archetypes in the collective unconscious. This may correspond to the operation of active information at the level of implicate order in the conception of Bohm and Hiley.

A remarkable feature of scenario (2.2) is the possibility that the mental and material manifestations may inherit mutual correlations due to the fact that they are commonly
caused by the psychophysically neutral level. One might say that they are remnants reflecting the lost holism at this level. In this way, they are not the result of any direct causal interaction between mental and material domains. Thus, they are not suitable for an explanation of direct mental causation in the usual sense. Psychologically speaking, some unconscious activity would be mandatory for their existence, and effects appearing as mental causation would have their origin there. Independently of quantum theory, a related move was suggested by Velmans [78]. But even without mental causation, scenario (2.2) is relevant to the ubiquitous correlations between conscious mental states and brain states.

In the proposal by Pauli and Jung, these correlations were called *synchronistic* (see also [64]), and have been applied to psychosomatic relationships as well [49]. An essential condition required for synchronistic correlations is that they are *meaningful* for those who experience them. It is tempting to interpret the use of meaning as an attempt to introduce semantic information as an alternative to syntactic information as addressed above. Although this entails all kinds of problems concerning a clear-cut definition and operationalization, something like meaning, both explicitly and implicitly, might be a relevant informational currency for mind-matter relations [7].

Very recently, Primas [65] has proposed a dual-aspect approach where the distinction of mental and material domains originates from the distinction between two different modes of time: tensed (mental) time, including nowness, on the one hand, and tenseless (physical) time, viewed as an external parameter, on the other. These two concepts of time emerge due to a symmetry breaking of a timeless level of reality that is psychophysically neutral (another proposal [8], based on a similar idea, implements temporal features differently). Nowness and the directedness of time originate in the mental domain, whose tensed time is quantum-correlated with the parameter time of physics via time entanglement. It must be admitted that this is still a tentative scheme without concrete indications of how to confirm or reject it empirically. Nevertheless, it is highlighted here since, for the first time, it offers a formally elaborated and conceptually consistent dual-aspect quantum framework for basic aspects of the mind-matter problem.

As indicated above, the approaches by Stapp and Vitiello contain elements of dual-aspect thinking as well, although these are not much emphasized by the authors. The dual-aspect quantum approaches discussed in the present section tend to focus on the issue of entanglement more than on state reduction. The primary purpose here is to understand correlations between mental and material domains rather than direct interactions between them. In this respect, it is worthwhile to refer to an attempt at generalizing the axiomatic basis of standard quantum theory in such a way that the concept of entanglement becomes applicable even beyond physical examples [12].

Another critical issue of dual-aspect approaches in general refers to the problem of panpsychism (see the recent review in [68]). In the limit of a *universal* symmetry breaking at the psychophysically neutral level, every system has both a mental and a material aspect. In such a situation, it is important to understand “mentality” much more generally than “consciousness.” Unconscious or proto-mental acts as opposed to conscious mental acts are notions sometimes used to underline this difference. The special case of human consciousness within the mental domain as a whole might be regarded as special as its material correlate, the brain, within the material domain as a whole.
5. Conclusions

The historical motivation for exploring quantum theory in trying to understand consciousness derived from the realization that collapse-type quantum events introduce an element of randomness, which is primary (ontic) rather than merely due to ignorance or missing information (epistemic). Approaches such as those of Wigner, of Stapp, and of Beck and Eccles emphasize this (in different ways) insofar as the ontic randomness of quantum events is regarded to provide room for mental causation, that is, the possibility that conscious mental acts can influence brain behavior. The approach by Penrose and Hameroff also focuses on state collapse, but with a significant move from mental causation to the noncomputability of (particular) conscious acts.

Any discussion of state collapse or state reduction refers, at least implicitly, to entangled states since those are the states that are reduced. In this sense, entanglement is always coaddressed when state reduction is discussed. By contrast, dual-aspect quantum approaches utilize the topic of entanglement differently, and independent of state reduction in the first place. Inspired by the entanglement-induced nonlocal correlations of quantum physics, mind-matter entanglement is conceived as the hypothetical origin of mind-matter correlations. This reflects the highly speculative picture of a basically holistic, psychophysically neutral level of reality from which correlated mental and material domains emerge.

Each of the examples presented in this paper has both promising and problematic aspects. The approach by Beck and Eccles is most detailed and concrete with respect to the application of standard quantum mechanics for exocytosis processes. However, it does not solve the problem of how the activity of single synapses enters the dynamics of neural assemblies, and it leaves mental causation of quantum processes as a mere claim. Stapp’s approach suggests a modified interpretation of status-quo quantum theory plus the projection postulate, comprising brain and mind without extending the theory formally. Although invoking inspiring philosophical, and some psychological background, it is largely hypothetical. The proposal by Penrose and Hameroff exceeds the domain of present-day quantum theory by far and is the most speculative example among those discussed. It is not easy to see how the picture as a whole can be formally worked out and put to empirical test.

The approach initiated by Umezawa is embedded in the framework of quantum field theory, more broadly applicable and formally more sophisticated than standard quantum mechanics. Together with Stapp’s proposal, it refers explicitly to the activity of neuronal assemblies as the neural correlates of mental representations. A clear conceptual distinction between brain states and mental states is most often missing, although the approach is not intended to be reductionistic. Vitiello’s more recent accounts offer some clarifying hints in that direction, which point to an understanding in terms of a dual-aspect approach. Other such approaches, like those of Pauli and Jung and of Bohm and Hiley, are conceptually more transparent in this respect. On the other hand, they are essentially unsatisfactory in respect to a sound formal basis and concrete empirical scenarios. A novel dual-aspect quantum proposal by Primas, based on the distinction between tensed mental time and tenseless physical time, marks a significant step forward, particularly concerning a consistent formal framework.
Yet another possible point of view might be relevant for studying relations between quantum theory and consciousness: the idea to represent the foundations of quantum theory in a basically stochastic fashion, as proposed by Prigogine and colleagues for several decades. In their later work, they elaborated this within the theory of distributions in rigged Hilbert spaces (or Gel’fand triplets) rather than state functions in ordinary Hilbert space. This generalized formalism has also been applied to mixing dynamical systems and their properties [6], in particular their temporal features in terms of time operators [4, 5, 51]. The brain is a paradigmatic example of a complex system establishing dissipative structures far from thermal equilibrium. Given the tight relation between these issues and the mind-matter problem, Prigogine’s visions will most likely prove to be fruitful for future research on consciousness.

Acknowledgments

Inspiring discussions on numerous topics treated in this paper are gratefully appreciated with Ioannis Antoniou, Guido Bacciagaluppi, Hans Flohr, Karl Gustafson, Hans Primas, Baidyanath Misra, Stefan Rotter, Giuseppe Vitiello, and Max Velmans. Useful comments by Guido Bacciagaluppi, Friedrich Beck, Thomas Filk, Hans Flohr, Hans Primas, and Henry Stapp helped to improve an earlier version of the paper.

References


Quantum theory and consciousness


[72] ———, *Attention, intention, and will in quantum physics*, Journal of Consciousness Studies 6 (1999), no. 8–9, 143–164.


Harald Atmanspacher: Department of Theory and Data Analysis, Institute for Frontier Areas of Psychology and Mental Health, 79098 Freiburg, Germany; Center for Interdisciplinary Plasma Science, Max Planck Institute for Extraterrestrial Physics, 85740 Garching, Germany

E-mail address: haa@igpp.de