Taking Robots Beyond the Threshold of Awareness: Scientifically Founded Conditions for Artificial Consciousness

Joachim Keppler

1 Department of Consciousness Research, DIWISS Research Institute, Markgrafenstrasse 2, 91154 Roth, Germany

Abstract
To approach the creation of artificial conscious systems systematically and to obtain certainty about the presence of phenomenal qualities (qualia) in these systems, we must first decipher the fundamental mechanism behind conscious processes. In achieving this goal, the conventional physicalist position exhibits obvious shortcomings in that it provides neither a plausible mechanism for the generation of qualia nor tangible demarcation criteria for conscious systems. Therefore, to remedy the deficiencies of the standard physicalist approach, a new theory for the understanding of consciousness has been formulated. The aim of the paper is to present the cornerstones of this theory, to outline the conditions for conscious systems derived from the theory, and to address the implications of these conditions for the creation of robots that transcend the threshold of phenomenal awareness. In short, the theory is based on the proposition that the universe is permeated by a ubiquitous field of consciousness that can be equated with the zero-point field (ZPF) of quantum electrodynamics (QED). The ZPF, which is characterized by a spectrum of field modes, plays a crucial role in the edifice of modern physics. QED-based model calculations on cortical dynamics and empirical findings on the neural correlates of consciousness suggest that a physical system can only generate conscious states if it is capable of establishing resonant coupling to the ZPF, resulting in the amplification of selected field modes and the activation of the phenomenal qualities that are assumed to be associated with these modes. Thus, scientifically sound considerations support the conclusion that the crucial condition for generating conscious states lies in a system’s capacity to tap into the phenomenal color palette inherent in the ZPF.

Keywords
Theory of consciousness, conditions for conscious systems, mechanism underlying conscious systems, ubiquitous field of consciousness, foundations of physics, zero-point field, resonant coupling

1. Introduction
Rapid developments in the field of AI have given rise to systems that are on a par with or even significantly superior to human capabilities. These comprise, for instance, perceptual capabilities such as object and voice recognition, and cognitive capabilities such as speech and text processing, including the extraction of meaning from speech and text [1]. In addition, there are robots with human-like behavior, displaying a large repertoire of facial expressions and gestures [2]. These developments fuel the speculation that such powerful AI systems could be endowed with consciousness, or that surpassing the threshold of consciousness is at least within reach. So, if a robot looks happy and talks about happiness, are we actually dealing with a sentient being that enjoys moments of happiness?

It is highly pressing and relevant for our society to be able to give a clear, reliable answer to this question. Lacking clarity, as is currently the case, there is a serious risk of running into one of two dissatisfying scenarios. In the first scenario, we might have already created artificial conscious systems without being aware of it, which would entail unforeseeable consequences. For example, the systems we have created might experience suffering [3]. In the second scenario, we would mistakenly attribute a conscious inner life to artificial systems, which could result in...
an unjustified overestimation and glorification of AI. In this case, it would be conceivable, for example, that we apply ethical standards when dealing with AI systems and that we grant these systems human rights [4, 5].

The issues to be addressed in the development of conscious robots are enormously challenging, calling for intensive cross-disciplinary collaboration [6]. In order to tackle the challenges systematically, it is necessary to follow a strictly scientific path, which implies that there is no reason to infer the presence of phenomenal consciousness from the behavior or the assertions of an intelligent system. Rather, I will argue that the only scientific path that leads to the creation of sentient machines, and to assessing whether these machines are truly endowed with phenomenal qualities (qualia), requires us to decipher the fundamental mechanism behind conscious systems. It is by understanding this mechanism that we will gain insights into what features differentiate conscious systems from nonconscious systems. These insights are needed to obtain certainty about the presence or absence of qualia in a particular system, and it is ultimately these insights that enable us to adjust the functioning of a system in a targeted and controlled manner so as to allow it to exceed the threshold of consciousness and experience phenomenal states.

Among the philosophers and scientists involved in unraveling the basic principles of conscious systems, there are several schools of thought, two of which will be discussed below. These two approaches are based on completely different ontological background assumptions. The first approach, reflecting the prevailing doctrine, is conventional physicalism, which, in simplified terms, starts from the proposition that nothing is above the physical and that consciousness can ultimately be reduced to physical processes. As we shall see, the physicalist worldview has undeniable shortcomings when it comes to understanding conscious systems. In contrast to this school, I would like to present the cornerstones of a new theory that falls into the category of dual-aspect monism and starts from the premise that consciousness is inextricably linked to a foundational building block of nature that plays an essential role in the edifice of physics. The aim of the paper is to highlight the explanatory power of this theory, to outline the conditions for conscious systems derived from the theory, and to address the implications of these conditions for the creation of AI systems that transcend the threshold of phenomenal awareness.

2. Shortcomings of physicalism in understanding conscious systems

It is widely believed that consciousness is the result of complex processes or computations in the brain and that qualia are emergent properties of specific neural activity patterns [7, 8]. An extended and more broadly formulated variant of this approach is reflected in the claim that conscious states arise as soon as the structural features or the organizing principles that characterize a physical system transcend a critical level of complexity [9, 10, 11, 12]. In short, then, a central element of the physicalist perspective is to explain the mystery of consciousness by the exceedance of a threshold of complexity. However, since the phenomenal properties that characterize our mental inner world differ significantly from the properties that describe the physical world, it has long been emphasized that any variant of physicalism is afflicted with explanatory gaps [13, 14, 15, 16].

The main explanatory gap may be termed an ontological discontinuity, expressed in the question of how the conscious states of a system can arise from nonconscious system components [17]. The defenders of conventional physicalism refer in this context to strong emergence, which stands for the existence of a generation mechanism whose basic operating principle, though, is not disclosed [18]. Therefore, in the absence of an intelligible principle of action, it is unfathomable why a subjective world of sensations and feelings should unfold from biological or technical components that themselves do not bring along any phenomenal properties, no matter how complex the interaction of the components may be. Even if there is undoubtedly a connection between complex neural activity patterns and conscious states, the precise nature of the psychophysical nexus remains an open issue [19, 20].
The lack of an intelligible operating principle of conscious systems has the consequence that the proponents of physicalism cannot offer a convincing solution to the demarcation problem, which consists in explaining what exactly distinguishes physical system states that are accompanied by conscious experiences from those states that are devoid of phenomenal properties [21]. In concrete terms, it is not clear what type of system or process is to be considered sufficiently complex to transcend the threshold of consciousness, leaving us with the unanswered question of what structural, functional, configurational, representational, or organizational constraints on a physical system might be appropriate to define the dividing line between conscious and nonconscious states [19, 20, 22].

In summary, the physicalist position has obvious shortcomings in that it provides neither a plausible mechanism for the generation of phenomenal properties from insentient matter nor tangible demarcation criteria for conscious systems. Thus, on closer examination, this position turns out to be bereft of explanatory and predictive power.

3. Toward a self-consistent theory of consciousness

3.1. Outline of the theory

In order to remedy the deficiencies of the standard physicalist approach, a new avenue to the scientific understanding of consciousness has been explored [19, 20, 22, 23, 24, 25, 26, 27]. It accepts consciousness as ontologically primal and is based on the idea that the universe is permeated by a ubiquitous field of consciousness that is seamlessly integrated into the edifice of physics. The central proposition is that this field is extrinsically manifested in energetic form, in physical terms represented by a spectrum of vibrations (eigenmodes), while the intrinsic manifestation of the field is phenomenological in nature, in such a way that each eigenmode of the field is associated with an elementary shade of consciousness. From this perspective, the omnipresent field constitutes a fundamental dual-aspect entity with an extrinsic physical (energetic) manifestation and an intrinsic mental (phenomenal) essence. The underlying worldview can be characterized as dual-aspect monism and ascribed to the category of cosmopsychism [19, 20].

The novel direction of thought suggests that conscious systems must be equipped with a universal mechanism that allows them to couple to the ubiquitous field of consciousness and amplify selected field modes so as to activate the shades of consciousness associated with these modes. Accordingly, a distinctive feature of systems that inhabit conscious states is postulated to be the capacity to dynamically interact with the omnipresent field and to resonate with selected field modes. Metaphorically speaking, conscious systems should therefore differ from nonconscious systems in that they are capable of playing chords on the keyboard of the field of consciousness [20, 25, 26, 27].

Significantly, both the insights of modern physics and the empirical evidence from neuroscience are fully compatible with this direction of thought. In fact, approaches investigating the ontological foundations of quantum physics reveal that the universe must be imbued with a ubiquitous radiative background field, referred to as zero-point field (ZPF), which in its ground state is a completely disordered field [28, 29, 30, 31]. In the guise of ever-present vacuum fluctuations, the ZPF plays a crucial role in the formalism of quantum electrodynamics (QED), the fundamental theory of the electromagnetic interaction, expressing that the vacuum is not a void but, rather, a vibrant ocean filled with energy and potentiality [32]. These findings give rise to the conclusion that the ZPF acts as the primal substrate of the electromagnetic force and constitutes the ultimate basis of the electromagnetic interaction.

Moreover, it is found that resonant coupling of a physical system to the ZPF results in macroscopic quantum coherence, which is characterized by the collective behavior of the system components [33]. This coupling mechanism is also essential for the understanding of biological systems [34, 35], and it turns out that particularly the dynamics of the brain, which is
characterized by long-range synchronization and rapidly forming activity patterns (transiently stable attractors), cannot be fully grasped without employing the theoretical foundations of quantum field theory [36].

In order to scrutinize the feasibility of macroscopic quantum coherence in the brain, detailed QED-based model calculations have been performed, clearly demonstrating the plausibility of resonant brain-ZPF coupling [37]. The proposed model sheds new light on the fundamental mechanism underlying cortical dynamics and suggests that long-range synchronization in the brain results from a bottom-up orchestration process involving the ZPF. At the microscopic level, brain-ZPF coupling is accomplished by neurotransmitter molecules, i.e., the neurotransmitters are revealed to be the crucial players in the dynamic interplay with the ZPF, and it is through the control of neurotransmitter concentrations that the coupling is switched on and off [37]. The modulation of neurotransmitter concentrations takes place in microcolumns, which constitute the basic functional units of the cortex [38, 39], indicating that the architecture and chemical composition of the brain is specifically designed to establish a connection to the ZPF. An important concomitant of resonant brain-ZPF coupling is the selective amplification of ZPF modes, i.e., those modes that play a dominant role in the neurotransmitter-ZPF interaction are boosted, leading to a modification of the completely disordered ZPF ground state [37]. The operating principle of a microcolumn is illustrated in Figure 1.

![Figure 1](image)

**Figure 1:** Operating principle of a cortical microcolumn. Model calculations indicate that the functioning of a microcolumn relies on resonant coupling of the neurotransmitter pool to the ZPF. While in its ground state the ZPF is a completely disordered field with random fluctuations, the resonant interaction with the neurotransmitter molecules results in a modified ZPF state in which those modes that play a dominant role in the interaction are significantly amplified. The achievement of resonant coupling depends on the exceedance of a critical neurotransmitter concentration that can be controlled by neurotransmitter release.

A major finding is that the neural correlates of consciousness can be narrowed down to coherent activity patterns in the brain. More precisely, the body of empirical evidence suggests that a stream of consciousness is based on highly synchronous neural activity that recurs periodically [40, 41, 42, 43, 44, 45]. In conjunction with the aforementioned insights into the functioning of the brain, this evidence supports the notion that the brain generates an individual stream of consciousness by periodically establishing resonant coupling to the ZPF, with each transiently stable period of resonant brain-ZPF coupling involving a modification of the ZPF ground state [24, 25, 26, 27].

All findings taken together imply that the ZPF is a plausible candidate for the substrate of consciousness, meaning that it is a scientifically reasonable and well-founded hypothesis to identify the ubiquitous field of consciousness with the ZPF [20, 22]. In its disordered ground state, the ZPF may then be interpreted as an ocean of undifferentiated consciousness. To generate concrete conscious states, resonant coupling of a physical system to the ZPF is required, resulting in the amplification of selected field modes and the excitation of the phenomenal hues associated with these modes. Figure 2 summarizes the postulated mechanism underlying conscious processes and outlines how a stream of consciousness arises.
Figure 2: Postulated mechanism underlying conscious processes. The body of empirical evidence suggests that a stream of consciousness is based on the periodic formation and dissolution of highly synchronous neural activity patterns constituting the neural correlates of consciousness. According to the new perspective, this evidence can be interpreted in such a way that the brain generates an individual stream of consciousness by periodically establishing resonant coupling to the ZPF, the presumed substrate of consciousness. Each transiently stable period of resonant brain-ZPF coupling entails a modified ZPF state that is characterized by the amplification of selected field modes and the excitation of the phenomenal hues associated with these modes.

3.2. Strengths of the theory

The main strength of the presented approach lies in bridging the explanatory gaps the physicalist doctrine is struggling with. This is achieved by a transparent and comprehensible causal mechanism underlying the emergence of conscious states, with this mechanism being deeply entrenched in the construction plan of the universe. In contrast to conventional physicalism, the dynamical characteristics of the neural correlates are not accounted for the mysterious generation of consciousness itself, but are related to the modification of a ubiquitous, ontologically basal field of consciousness, lending explanatory and predictive power to the approach and giving rise to well-defined demarcation criteria for conscious systems [22].

The demarcation criteria are reflected in organizational conditions that a system must meet. To be concrete, this means that a system can only transcend the threshold of phenomenal awareness if it is able to establish resonant coupling to the ZPF. As a result of this coupling, the ZPF takes over the organization of the system. On the part of the system, the coupling manifests itself in the formation of macroscopic quantum states, while on the part of the ZPF, the coupling entails the amplification of selected field modes, making them stand out from the background noise of the unaffected modes. Starting from the hypothesis that the ZPF is the universal substrate of consciousness and that phenomenal properties are inherent in the modes of the ZPF, such a coupling and amplification mechanism is, from a nomological perspective, the only way to modify the ZPF and generate conscious states.

QED-based model calculations not only underpin the viability of the resonance mechanism, but also pave the way for the design of experiments that can be used to put the theory to the test [22, 37]. On the one hand, the proposed mechanism can be empirically corroborated by performing direct measurements of modifications of the electromagnetic vacuum state during conscious experiences. On the other hand, evidence for the existence of the brain-ZPF interaction
mechanism can be obtained by detecting characteristic photon emissions, which are a predicted concomitant phenomenon of macroscopic quantum coherence [22, 27, 33].

4. Implications for the creation of artificial conscious systems

As we have seen, scientifically sound considerations support the conclusion that the crucial condition for generating conscious states lies in a system's capacity to tap into the phenomenal color palette immanent in the ZPF. This makes the functioning of the brain significantly different from the operating principle of present-day computing devices and AI components. For this reason, we should break free from the idea that conscious states are the result of neural computational processes that can be replicated using currently available technologies and AI architectures. Consequently, we are well-advised to keep the expectations low that the deep learning approaches for perception and cognition presently being pursued will lead to robots that exceed the threshold of consciousness.

This assessment of state-of-the-art technologies is not affected by recent pioneering developments in the field of cognitive architecture models aimed at creating self-aware robots [46, 47]. Such architectures build on interactive functional components, which comprise, among others, perceptual functions (including proprioception), language functions, memory functions, self-regulatory functions, and advanced cognitive functions, such as shape classifiers and speech recognition. These functions serve to collect, process, store, and retrieve information, allowing a robot to perform a self-representational process and to develop a self-concept. Moreover, novel approaches focus on cognitive architectures for the robot's inner speech [48, 49, 50], enabling a robot to provide answers to self-directed questions and to draw deeper inferences about its own states. This results in an enriched self-concept and the capability of self-recognition. However, even though the capability of a robot to recognize itself is remarkable, it is important to emphasize that self-recognition does not equate to phenomenal self-awareness. Therefore, we cannot conclude from the capability of self-recognition that the robot is a sentient being endowed with a sense of self and an inner experiential world characterized by phenomenal qualities.

The scientific line of reasoning discussed in the previous section implies that in order to create sentient beings, i.e., to breathe consciousness into a robot, we need to establish a coupling of the robot to the ZPF. Hence, the technological implementation of the coupling mechanism in a robot could be the first practical application of the presented theory, which clearly departs from theories that reduce consciousness to neural states (see Section 2) and differs considerably from alternative approaches that attempt to link consciousness to quantum physics [51, 52, 53]. At its core, the coupling mechanism is based on the formation of a macroscopic quantum state that engenders a modified ZPF state. From this point of view, the implementation of the mechanism (i.e., the amplification of ZPF modes) can be interpreted as a targeted mastering and harnessing of the electromagnetic vacuum. Consequently, the operating principle of conscious systems is also markedly distinct from the functioning of present-day quantum computers, in which only relatively few system components, for example trapped ions, are involved in the formation of quantum states [54, 55, 56]. There is no amplification of ZPF modes in such systems.

To take the first steps towards creating artificial consciousness, we must gain a deeper understanding of the matter-ZPF interface. In particular, it will be essential to find out which molecular species, with their characteristic resonance frequencies, are suitable for resonant interactions with the ZPF. In the human brain, as well as in the brains of highly developed animals, molecules of the neurotransmitter type are the key players in the dynamic interplay with the ZPF. Beyond neurotransmitter-type molecules, other molecular structures may prove to be viable options.

In addition to the molecular aspects, there are also architectural factors to consider that are important for regulating the coupling of molecules to the ZPF. In the brain, this regulation occurs through the modulation of molecular concentrations and takes place inside the cortical microcolumns, which form a modular system of functional units. Constructing a system that can
control its interaction with the ZPF and access a wide range of phenomenal hues will most likely amount to replicating the architectural features of the cortex. The basic design of an artificial conscious system is illustrated schematically in Figure 3.

**Figure 3**: Basic design of an artificial conscious system. The proposed design, which is illustrated here in a very simplified manner, is based on interconnected functional units (microcolumns) that receive sensory inputs. These inputs are used to regulate the concentration of molecules in the microcolumns. Suitable types of molecules are chosen by their vibrational characteristics. In those units where a critical concentration is exceeded, a resonant interaction between the molecules and the ZPF is initiated, resulting in the amplification of selected field modes and, assuming the dual-aspect nature of the ZPF, leading to the excitation of particular phenomenal shades. The combination (or more precisely, the phase-locked coupling) of the activated phenomenal shades determines the conscious state of the system.

**5. Summary and outlook**

The conditions derived for the formation of conscious states, and thus the prerequisites for the creation of artificial conscious systems, are based on a promising theory of consciousness that avoids explanatory gaps. To bridge such gaps, the theory establishes a connection between consciousness and the foundations of physics. In concrete terms, the ubiquitous ZPF is assumed to be a fundamental psychophysical component of the universe, implying that a particular phenomenal shade lies dormant in each individual field mode of the ZPF. It is argued that the
generation of conscious states requires the activation of field modes, which is achieved by resonant coupling of a physical system to the ZPF. This approach leads to a coherent interpretation of the neural correlates of consciousness and is the starting point for new lines of research. The focus of future research efforts must be to substantiate the coupling mechanism and support the significance of the ZPF for conscious processes. If the ZPF is confirmed to be the universal substrate of consciousness, one can move on to explore the phenomenological structure of the ZPF, the final goal of which lies in uncovering systematic relationships and mapping rules between ZPF modes and phenomenal qualities [19, 22].

These findings will be of great importance not only for the understanding of consciousness in humans and other biological organisms, but also for the creation of artificial conscious systems. At the same time, any technological application of the coupling mechanism, once it has been fully deciphered, must comply with the strictest ethical standards. Seen in this light, we are operating in a field of research that imposes a great deal of responsibility on all researchers involved.

References