Journal of Human Cognition
2024, VOL. 8, NO. 2, 39-49
DOI: 10.47297/wspjhcWSP2515-469905.20240802

Computational Explanation of Consciousness: A Predictive Processing-based Understanding of Consciousness

Zhichao Gong

School of Philosophy, Shanxi University, Taiyuan, China

ABSTRACT
In the domain of cognitive science, understanding consciousness through the investigation of neural correlates has been the primary research approach. The exploration of neural correlates of consciousness is focused on identifying these correlates and reducing consciousness to a physical phenomenon, embodying a form of reductionist physicalism. This inevitably leads to challenges in explaining consciousness itself. The computational interpretation of consciousness takes a functionalist view, grounded in physicalism, and models conscious experience as a cognitive function, elucidated through computational means. This paper posits that predictive processing offers a fresh paradigm for comprehending human cognition and serves as a neural foundation for the computational elucidation of consciousness. Additionally, free energy theory, as a mathematical construct explaining the predictive processing model, furnishes a theoretical scaffold for understanding consciousness computationally.

1 Introduction

In 1990, F. Crick and C. Koch introduced the concept of “Neural Correlation of Consciousness (NCC)”[1], marking the beginning of a mainstream paradigm in consciousness research, anchored in neuroscience. They argued against abstract, logic-based discourses, asserting that the investigation of consciousness must be entwined with scientific insights into brain neurons. Thus, the exploration and definition of neural correlates, molecular mechanisms, biophysics, and neurophysiological properties became vital to demystifying consciousness[2]. However, the subsequent introduction of D. Chalmers' “easy problem” and “hard problem” categorization of consciousness[Chalmers distinguishes between problems that can be addressed by conventional cognitive science methods, and those that concern the qualities of subjective experience, or qualia] revealed inherent complexities. NCC research stumbled upon the interpretative challenge: how and why do neurophysiological activities yield “conscious experience”? As opposed to delving into qualia or subjective experience, this paper aligns with A. Seth's assertion that “The real issue is not to explain how consciousness emerges in the physical world, but how brain activity gives rise to conscious experience”[3]. With the evolving landscape of cognitive science, the focus of consciousness research has shifted from neural correlation to computational interpretation. This approach abstracts conscious experience as a cognitive function, explained through computational means. It highlights predictive processing as a novel model for understanding human cognition, suggesting that it can provide the neural foundation for a computational explanation of consciousness. In this context, conscious perception is interpreted as the result of unconscious perceptual reasoning and a unique way of maintaining perceptual reasoning under predictive processing. Conscious experience is portrayed as an emergent result of multi-level neural representations. The free energy theory, which mathematically explicates the minimization of prediction errors, offers theoretical backing for this computational understanding of consciousness.

2 Basic Properties of Conscious Experience

The fundamental question we must first address is, "What is conscious experience? " Noted philosopher T. Nagel employs the sensory experience of bats to draw an analogy with human consciousness experience. He posits that the reason for the bats' sensory experience lies in their direct perception of what it feels like to be alive [4]. Consequently, human consciousness arises from this same type of conscious experience. F. Dretske supports this view, asserting that understanding the nature of conscious experience will simultaneously reveal the nature of consciousness itself [5]. We argue that an examination of the basic properties of conscious experience is an essential path to explore the issue of consciousness. Therefore, we must identify what content or specific properties of conscious experience constitute its defining features.

A. Seth offers a perspective on consciousness by exploring it from four facets [6]: integration, differentiation, metastability, and dynamic complexity. Conscious experience is integrated in the sense that it is perceived as a unified whole, and differentiated in that it consists of numerous parts, comprising a myriad of possible experiences. This complex interplay between integration and differentiation emphasizes the highly informational complexity of consciousness. Furthermore, conscious experience is characterized by attributes such as subjectivity, emotionality, and volitional experience, all of which contribute to the multifaceted nature of consciousness [7].

According to C. Pennartz, the essential attributes of healthy human conscious experience are multimodal richness, situationality, unified integration, dynamic stability, and intentionality [8]. These can be described as follows:

First, conscious experience is qualitative, characterized by sensations in multiple distinct modalities such as vision, audition, somatosensation, smell, taste, and vestibular sense. These modalities encompass various submodalities like color, texture, shape, and motion, which are experienced in a spatial arrangement. This multimodal richness is considered a primary characteristic of conscious experience and is philosophically connected to the concept of qualia.

Second, conscious experience possesses the property of being situated, meaning we find ourselves immersed within a spatial and temporal context filled with objects and stimuli. Our own body's sensations contribute to this situatedness, creating a seamless connection between the subject and its environment.

Third, conscious experience is marked by unity, where sensory input streams from different modalities are unified into a single, situational experience. The integration of these sensations occurs across various neural mechanisms, resulting in an experience that is perceived as unified or unitary.

Fourth, the nature of conscious experience is defined by a balance between dynamics and stability. Dynamics refers to continuous updates in our perception, while stability represents the consistent positioning of objects as we perceive them, despite body movements or changes in the external environment.

Fifth, intentionality is a fundamental hallmark of conscious experience, encompassing the brain's capacity to interpret neural activity patterns. This includes multi-interpretability, the non-veridical nature of hallucinations, and the sudden recognition of objects in Gestalt compositions.

In summary, the attributes of conscious experience can be dissected into multiple dimensions, each shedding light on the intricate nature of consciousness. The exploration of these characteristics fosters a comprehensive understanding, paving the way for a complete theory of consciousness. By examining these dimensions in depth, we can further our knowledge and appreciation of the complex phenomenon of conscious experience.

3 The Neural Basis of Conscious Experience: A Representational Model Based on Predictive Processing

Conscious experience, in terms of its basic properties, can be described as a multimodal, contextually rich phenomenon marked by situational characteristics. The connection between this experience and brain activities prompts an intriguing inquiry. The theory of 'Direct Realism' posits that conscious experience does not depend on intermediaries such as perception, sensory impressions, or representations, but instead directly perceives physical objects. Building on this, activism contends that conscious experience emerges from the interactions between our bodies and the external world, with the brain serving as a mere physical conduit for receiving sensory input and initiating behavioral output.

J. K. O'Regan critiques representation theory, arguing that it fails to elucidate how internal representations can give rise to conscious experience, positing instead that the external world serves as its own representation in generating conscious experience [10]. However, both direct realism and activism face significant challenges in explaining phenomena such as the persistence of conscious experience without overt motor activity, non-motor occurrences of conscious experience, and the interaction with unreal objects, including rapid eye movement (REM) states and hallucinations.

Consequently, it becomes apparent that conscious experience is inextricably tied to representation, and the body's perception of the outside world is a manifestation of the intentional activity of the brain's internal representation. Though mental representations are intangible, they are often implied through concepts such as mental states or belief states, reflecting their constant presence in our thought processes, actions, and explorations. This underscores that representational activities are unique to humans and their intellectual derivatives, artificial intelligence, as representation inherently contains meaning, and meaning entails understanding and interpreting behavior [11]. The pressing question that remains is how the brain carries out these representational activities.

3.1 The Predictive Processing Representation Model of the Brain

Predictive processing has emerged as a novel model for understanding human cognition and representational activities in the brain, providing a neural basis for the generation of conscious experience. This theory was conceived to tackle the problem of perception, specifically, the manner in which the brain constructs perceptions about the external world based on sensory input.

Though its theoretical roots can be traced back to the concept of "unconscious reasoning" proposed by physicist H. Helmholtz in the 19th century, the formation of predictive processing's theoretical core has been enriched by Bayesian inference and recent advancements in neuroscience and brain science. Over the past decade, efforts have been made to evolve this theory into a unified cognitive research paradigm, explaining perception, action, and attention [12].

Brown and Friston et al. amalgamated the predictive generative model with the hierarchical predictive coding method, characterizing the "top-down" multi-level connection as prediction-driven sensory signals, where only "prediction errors" transmit information within the system [13]. The predictive processing model integrates a "top-down" probabilistic generative model in a multi-level bidirectional cascade with a predictive coding strategy, aiming to minimize predictive processing through the symbiotic interaction of perceptual and active reasoning.

Initially applied to the realm of perception, this approach was extended to the field of action by Friston et al., supplying a consolidated research framework for the brain's representational activities. The underlying premise of predictive processing posits the brain as a predictive apparatus, continuously aligning "top-down" prior predictions with "bottom-up" real perceptual inputs, deducing the cause of current perception, and acting to minimize the discrepancy between predicted and actual states.

A significant facet of this assertion is not the brain's direct reception of perceptual information but its proactive prediction and neural construction of perception's source. This entails two distinct forms of predictive processing: "top-down" and "bottom-up." The former refers to predictions formed independent of cognitive content, while the latter depends on current perceptual input. Together, these predictive processes forge a complex "hierarchical forecasting" model.

Prediction error, which arises when sensory evidence clashes with the brain's top-down prior predictions, prompts lower-level neural circuits to communicate the error to higher-level circuits. This interaction leads to the correction and updating of the original prediction, culminating in an optimal prediction. Perception's content, thus, is independent of external stimuli and is rather a product of the brain's multi-level neural circuits' reciprocal stimulation.

J.Hohwy succinctly delineates the predictive processing model as[14]：

(1) prediction error = input – prediction

This basic notion of predictive coding provides an efficient message passing scheme because

prediction errors carry information about the quality of the prediction and are used to update the model, leading to new predictions:

(2) new prediction = old prediction + prediction error

On its own, the scheme in Eq. (2) is quite liberal: it weights the prediction error and the

learning that has already occurred equally, putting relatively little trust in the history of prior

learning. To counter this, the updating can conform to Bayes′ rule, such that the system engages in step-wise optimal perceptual inference of hidden causes. This requires a weight to be put on the prediction error. The weight is based on the precisions of the probability distributions (or, in the continuous case, probability density functions), where precision is the inverse of variance such that a highly precise distribution has very little variance (it is convenient to express these constructs using sufficient statistics, the numerical values for the means and precisions of, here, normal probability distributions). Thus, where πP is the prior precision (concerning how much is already known) and πL is the likelihood precision (concerning how much is learnt from the current sensory input):

(3) new prediction = old prediction + $((\pi L / (\pi P + \pi L)) \times$ prediction error)

The weight $\pi L / (\pi P + \pi L)$ expresses the learning rate, which increases with the likelihood precision and decreases with the prior precision. This allows inference to trust the learning that has already occurred, and largely ignore imprecise new evidence. The account of perceptual inference can be enriched with a notion of precision optimisation. Here, precisions are themselves inferred such that the model can generate descending predictions of precisions, which are used in the denominator of Eq.3 to create a variable learning rate that is sensitive to contextual information (cf. empirical Bayes, where priors are learnt over time and extracted from higher levels of the hierarchy). The internal model here needs to be hierarchical, such that the weight on prediction errors of any given level can be modulated by learnt regularities across several spatiotemporal ranges. For example, the weight on visual prediction error is decreased when I navigate a room without my glasses on, and decreased further if it is a room I am very familiar with.

This enables reasoning to trust learning that has already occurred and largely ignore imprecise new information. The weight of prediction errors at any given level can be modulated by learned regularities, and through repeated perceptual inference steps across levels, the internal model generalizes the dynamic causal structure of the system's surrounding environment. In hierarchical inference, each level n predicts activities at level n-1, receives weighted prediction errors for that level, and passes the weighted prediction errors to level n+1 in response to predictions made there.

We believe that predictive processing in the brain proceeds in two basic ways. One is perception, and predictive processing can adjust the prediction error to correct the model through perceptual inference, which will reduce the error of different time scales and different levels of model hierarchy, and improve the model fit. On short timescales, the forecast error is minimized to find the best posterior probability. On longer time scales, generative models adjust parameters through perceptual learning to perform better in perceptual inference; the second is action. Predictive models generated by perceptual inference may result in a temporary increase in prediction error, but the model can be held fixed and used to generate predictions that are subsequently minimized by action or selective sampling of sensory input. In the framework of predictive processing, action is also active inference, which assumes that an organism must admit a "minimum-uncertainty" causal model of the probabilistic relationship between relevant events. In order to survive and reproduce when faced with a constantly changing external world, organisms must receive sensory inputs and the possible environmental causes of those inputs. Moreover, organisms must also simulate the causes of the external world and perform possible behaviors that are physiologically permitted[15].

In conclusion, predictive processing operates in two fundamental manners in the brain: perception and action. Perception, through perceptual inference, adjusts prediction error to correct the model and improve its fit, while action, through predictive models, may increase prediction error temporarily but utilizes predictions for minimization through action or selective sensory input sampling. The synergy of perceptual and active reasoning actualizes the predictive processing of the brain. The question of consciousness, a core topic in cognitive research, prompts an inquiry into how conscious experience is engendered within predictive processing models.

3.2 The Generation Mechanism of Conscious Experience: An Emergent Mechanism Based on Predictive Processing

Hohwy posits evidence for a dissociation between attention and conscious perception. However, the delineation between these functions is far from straightforward. A common understanding would imply a direct correlation between consciousness and attention; one attends to things they are conscious of, and they are conscious of things they attend to. This perspective suggests that a comprehensive account of attention must recognize both a close connection with conscious perception and their distinct functional roles. There is reason to consider that our understanding of unconscious perceptual inference may illuminate attention and accommodate its various aspects. Throughout the discussion, the notion of expected precisions has repeatedly emerged as an essential component of perceptual inference (and active inference), analogous to the necessity of considering variance when comparing means statistically [16].

Predictive processing theorizes that conscious perception arises from unconscious perceptual reasoning, with conscious perception and attention serving distinct functions within this framework. Conscious perception is identified as a process of generating previous predictions, not involving conscious Bayesian updates. Conscious perception is defined by the assumptions about the world at that moment, with empirical probability representing the best-performing prediction at that time. According to K. Friston, attention serves as the precision optimization within predictive processing to minimize hierarchical prediction errors [17]. In the complex, noisy world we inhabit, perceptual inference depends on both the magnitude and precision of the prediction error, with conscious perception and attention collectively maintaining the process of perceptual reasoning.

A fundamental question that persists is why neural activity based on predictive processing produces conscious experience. Specifically, why do numerical representations in multilevel regions of predictive processing models enable entire neural networks to generate experiences with non-numeric properties? One proposed answer is that calculations for a specific modality must be processed by a higher-level brain region to become conscious. However, this hypothesis encounters challenges, as no single structure within the human brain is capable of translating functions into conscious perception, given that the brain consists of interconnected neurons without the representational power to encode complex phenomena such as conscious perception [18].

Consciousness, in our view, emerges through multi-level neural representations within the brain's hierarchical prediction model (minimizing prediction errors). We regard consciousness as an emergent property, characterized by novelty, irreducibility, and unpredictability. This concept of emergence refers to properties of the whole generated by component interactions but not by the components alone. Historically, emergent thought has roots in Aristotle's proposition that "the whole is not equal to the sum of its parts." Cognitive Emergentism offers a new perspective on cognition or consciousness, considering it as the result of a complex system's interactions, including the mind, body, and environment. The process of emergence requires multidisciplinary analysis, involving philosophy, biology, cognitive neuroscience, cognitive science, and artificial intelligence [19].

Kim's work on emergent theory highlights potential problems with dual causation, leading to over-determination and violating the causal closure of physical domains [20]. Our understanding of emergence as an objective conscious experience avoids any top-down psychological causation. To articulate the concept of multi-level representation, Pennartz suggests avoiding misleading metaphors, such as the hardware-software distinction. Instead, Godel's proof of his first incompleteness theorem serves as a more accurate metaphor. The multi-level understanding of consciousness provides a valuable approach to addressing the Hard Problem of consciousness, even though it does not provide an immediate and tangible solution. Neurorepresentationalism highlights the inaccessibility of transitioning from neurons to conscious experience, as intermediate levels of representation are not directly accessible to subjective experience. Critiques of this approach may stem from the challenge of reconciling our perceptions with the complex underlying neuronal mechanisms.

4 Computational Explanation of Consciousness

Predictive processing, as a brain representation model with robust explanatory and predictive power, has opened a pathway to understanding the mechanism underlying conscious experience. In this context, free energy theory, through its mathematical exposition of the predictive processing model, establishes a theoretical foundation for the computational explanation of consciousness. This shift from mere neural correlation to computational explanation has allowed for new insights in consciousness research.

4.1 From Neural Correlates of Consciousness to Computational Explanations

Chalmers provided a canonical definition of the neural correlates of consciousness (NCC), identifying NCC as a minimal neural system (N) where there is a correlation between the states of N and states of consciousness[21]. This definition extends to include various cases of neural correlates, which I will omit here for brevity. The nature of "conditions C," or the specific circumstances under which NCC applies, is still open to debate and further investigation.

Wiese et al. brought forth challenges to Chalmers' definition of NCC from three aspects[22]:

Challenge 1: Global dynamics. The definition of an NCC needs to distinguish neural activity associated with consciousness from neural activity that is not. Computational correlates, being neutral with respect to neural structures, present a promising approach to addressing this challenge by considering global dynamics associated with consciousness.

Challenge 2: Non-arbitrary mappings. This challenge pertains to understanding why a particular neural structure N is an NCC of consciousness. Though Chalmers' definition allows for seemingly arbitrary mappings, a more profound understanding of consciousness necessitates that these mappings be non-arbitrary, general, and lawful.

Challenge 3: Unusual conditions. The definition of NCC is tested by unusual circumstances, such as when brain function is affected by lesions. This has led to a complex discourse around the sufficiency and necessity of neural activity in consciousness.

Consequently, the neural correlates of consciousness face problems such as non-arbitrary mapping, the absence of a one-to-one mapping relationship, and the correlation challenge[23]. These limitations spurred Cleeremans to propose the Computational Correlation of Consciousness (CCC), emphasizing computational differences rather than neural activity[24]. CCC explores computational properties tied to conscious experience, determined by computational models, thus elucidating the unity of conscious experience. Moreover, it represents non-arbitrary mapping and describes computational functions necessary for the generation of conscious experience.

However, both neural and computational correlates of consciousness are correlation-based studies lacking causal explanatory power. To move from correlation to causation requires a nuanced understanding of conscious experience properties. This understanding can be illuminated by the free energy theory rooted in predictive processing, reflecting the abstract specifications of causal relationships in human cognition. The remainder of this text appears to be truncated, and further examination would be required for a complete analysis and interpretation.

4.2 Philosophical Basis of Computational Interpretation

The possibility of a computational explanation of conscious experience begs a classic metaphysical question, namely, the relationship between the physical processes of the brain and its psychological attributes. From a functionalist perspective grounded in physicalism, mental properties are regarded as emerging from neural representational processes. These physical and mental processes are conceived as dual properties manifested by the same entity, abstracted into a function. Functionalists interpret this function in light of underlying legitimate mechanisms, like Miller's law pertaining to short-term memory, which illustrates that humans can store specific information units in working memory[25]. Analogous to memory function, neural representation and conscious experience are different facets of consciousness, with the latter emerging from the neural representation process. M. Solms contends that abstract mathematical algorithms can delineate the functional attributes of consciousness[26]. Thus, the question arises: is a computational explanation founded on mathematical principles adequate?

Chalmers identifies two basic theses that articulate the foundational role of computation. First is the thesis of computational sufficiency, which underpins the belief in artificial intelligence by asserting that an appropriate computational structure suffices for possessing a mind and various mental properties. Second is the thesis of computational explanation, proclaiming that computation furnishes a universal framework for explicating cognitive processes and behavior[27]. Stability in this foundation necessitates a clear understanding of computation. While the abstract mathematical theory of computation is well established, a bridge connecting physical systems with abstract computational theory is requisite. Specifically, a theory of implementation is needed, elucidating the relationship between an abstract computational object and a physical system. Answering the key question regarding the relationship between computation and cognition further clarifies this foundation.

Chalmers, in minimal computationalism, emphasized two fundamental propositions. First is computational sufficiency, positing that the correct computational structure suffices for the constitution of various psychological properties. Second is computational explanation, postulating that computation can furnish a general framework for cognitive processes and behaviors. The implementation of a given computation in a physical system is contingent on the causal structure reflecting the computation's formal structure. Essentially, this relationship between computational and physical systems is an isomorphic one, where computation offers an abstract specification of the system's causal organization. If cognitive systems derive their mental properties from their causal organization, specified computationally, the argument for computational sufficiency holds. Similarly, the proposition of computational explanation holds if the system's causal organization primarily pertains to the explanation of behavior.

In summary, free energy theory, framed as an abstract mathematical law, can furnish a plausible computational rationale for conscious experience. The revised section helps unify the themes and provide clear subsections to guide the reader through complex ideas, aligning it more closely with the expectations of academic writing.

4.3 Interpretation of Calculations Based on Free Energy Theory

Free energy theory originates from the notion of "free energy" in information theory and statistical thermodynamics. Friston leveraged this theory to explicate the behavior and perception of organisms within the framework of the predictive processing model. Essentially, the process of minimizing prediction error corresponds to the minimization of variational free energy. Within the context of free energy theory, conscious experience is engendered through a steady-state mechanism, defined as a relatively stable state attained by the self-organizing system via the minimization of free energy. This minimization serves as a fundamental function enabling organisms to sustain homeostasis, actualized through predictive processing. Consequently, the functional attributes of conscious experience emanate from the brain's predictive processing mechanism. The physiological aspect reveals that the functional characteristics of homeostasis and conscious experience are localized in identical regions of the brain. Therefore, free energy minimization emerges as the principal force underlying the explanation of consciousness. A self-organizing system can instantiate a steady-state mechanism through three formulations[28]:

First Formulation: Free energy is depicted as energy minus entropy. This description holds importance for three primary reasons: a. It bridges the free energy concept in information theory with those in statistical thermodynamics. b. It elucidates that free energy can be assessed by an agent, as energy represents the surprise regarding the joint occurrence of sensations and perceived causes, whereas entropy pertains to the agent's own recognition density. c. It demonstrates that free energy relies on a generative model of the world, articulated in terms of the probability of simultaneous occurrence of a sensation and its causes. This implies that an agent requires an implicit generative model to comprehend how causes contribute to producing sensory data, defining both the agent's nature and the quality of the free-energy constraint on surprise.

Second Formulation: Free energy is articulated as surprise plus a divergence term. Here, the (perceptual) divergence simply refers to the disparity between the recognition density and the conditional or posterior density of sensation's causes, given the sensory signals. As this difference is consistently non-negative, free energy serves as a cap on surprise. Thus, the minimization of free energy, achieved by modifying the recognition density without altering sensory data, curtails perceptual divergence. Consequently, the recognition density converges to the conditional density, and free energy translates into surprise.

Third Formulation: Free energy is portrayed as complexity minus accuracy, employing terminology from model comparison literature. Complexity, also known as Bayesian surprise, is the distinction between the recognition density and the prior density on causes. Accuracy pertains to the surprise concerning expected sensations under the recognition density. This construct illustrates that the minimization of free energy, by modifying sensory data without shifting recognition density, enhances an agent's predictive accuracy, a concept referred to as active inference. A tangible instance of this procedure, when elevated into consciousness, might be akin to navigating in darkness by confirming tactile anticipations.

In conclusion, free energy is predicated on a specific model detailing the generation of sensory data, as well as on a recognition density related to the model's parameters, or sensory causes. The reduction of free energy can be achieved solely by modifying the recognition density to alter conditional expectations about the samples taken or by adjusting the sensory samples to ensure that they align with these expectations. Subsequent sections will explore these considerations, examining their relationship to prevailing theories regarding brain function.

It is important to clarify that free energy theory, in and of itself, does not constitute a theory of consciousness, nor does it offer a direct computational explanation for consciousness. Instead, the essence of the computational explanation of consciousness lies in its functional elucidation, specifically in detailing the cognitive abilities associated with consciousness. This examination includes both subjective and objective characteristics, describing the functional properties of consciousness through the experience of consciousness itself. Objectively speaking, conscious experience emerges from the brain's predictive processing model, for which the free energy theory furnishes a coherent mathematical framework. Thus, from the vantage point of physicalism's functionalism, free energy theory can be construed as an abstract construct suitable for explicating conscious experience.

5 Epilogue

In conclusion, conscious perception is synthesized through predictive processing layered prediction error minimization. This perspective articulates that consciousness emerges in the representation process of the brain's hierarchical prediction model, which is realized through multi-level neural representations. Ontologically, the neural representation process and conscious experience can be seen as different manifestations of consciousness as a cognitive function. The free energy theory serves as an abstract framework explaining these functional attributes. Methodologically, computational approaches offer a comprehensive explanatory framework for cognitive processes and behaviors. The structural attributes of computational systems reflect the causal intricacies of physical systems, and thus, free energy theory grounded in predictive processing models can furnish computational elucidations for conscious experience.

Both predictive processing models and free energy theory can be classified as adaptive representations. The term adaptive representation denotes that the cognitive system is capable of autonomously representing the target object within a specific environment or context. This capability can be modified and enhanced autonomously in response to changes in the environment or context. It is vital to emphasize that the generation process of adaptive representation is synonymous with adaptation, and it does not necessarily imply optimization. The notion of adaptive representation integrates aspects of biological evolution theory (adaptation to the environment), context theory (historical association), and generative cognitive science (structural coupling). It accentuates the environmental adaptability, historical relevance, and meaningful coupling reality of representation.

Should the concept of "adaptive representation" reflect an essential quality of cognitive systems, cognitive theories addressing this representational capacity should refer to this capability as well. The critical point here is not merely to illustrate how adaptive representations are generated but to focus on demonstrating how adaptive representation becomes a categorizing ability of the cognitive system and whether it reflects the essential attributes of the cognitive system. Additionally, it is essential to explore whether adaptive representation is a shared characteristic among various cognitive science theories.

Summarizing, the central hypothesis presented here is that cognitive systems, be they natural (as in the human brain) or artificial (as in computers), function as adaptive representational systems. This implies that cognition, inclusive of the mind, is intrinsically adaptive. Such a perspective provides a novel cognitive view distinct from existing theories and research paradigms. Adaptability in this context embodies autonomy and regulation, while representation encapsulates intentionality and mediation. These attributes are indispensable for an intelligent subject, as without them, intelligent behavior cannot manifest. One may infer from this that all knowledge, be it in natural sciences, humanities, social sciences, based on reason or experience, is the product of adaptive representation.

From the standpoint of adaptability, the term signifies the subject's capacity to make corresponding adjustments in response to environmental alterations to harmonize the interaction between the subject and its surroundings. In this context, free energy theory elucidates the prediction error minimization process of the predictive processing model. From the perspective of representation, the predictive processing model operates as a mental representation founded on structuralism, with free energy theory serving as a symbolic representation predicated on mathematical theory. This concept of adaptive representation amalgamates the subject's ability to perceive and symbolize external objects. From the viewpoint of life-mind continuity, evolution implies adaptation, and virtually all human abilities, inclusive of mental representation, are adaptive. With mental representation ability comes cognitive ability, and with cognitive ability comes mind and intelligence. Therefore, the calculation of consciousness elucidates its essence as the adaptive representation process of organisms.

Funding

"Philosophical Study on the Challenge of Artificial Cognition to Natural Cognition" (21&ZD061), a major project of the National Social Science Foundation of China.

References

[1] Crick F., & Koch C. (1990). Towards a neurobiological theory of consciousness. In Seminars in the Neurosciences (Vol. 2, pp. 263-275). Saunders Scientific Publications.

[2] Liu M., & Zhao X. (2016). Neural correlates of consciousness and intrinsic mental view. Science and Technology Philosophy Research, 33(05), 12-17.

[3] Seth A. (2021). Illuminating consciousness. New Scientist, 251(3350), 44-48.

[4] Nagel T. (1974). What is it like to be a bat? The Philosophical Review, 83(4), 435-450.

[5] Dretske F. (1993). Conscious experience. Mind, 102(406), 263-283.

[6] Seth A. (2009). Explanatory correlates of consciousness: theoretical and computational challenges. Cognitive Computation, 1(1), 50-63.

[7] Wegner D. M. (2002). The illusion of conscious will. Cambridge: A Bradford Book.

[8] Pennartz C. M. A. (2022). What is neurorepresentationalism? From neural activity and predictive processing to multi-level representations and consciousness. Behavioural Brain Research, 432, 113969.

[9] 魏屹东. 论具身人工智能的可能性和必要性.2021,298(02), 1-10.

[10] O'regan J. K., & Noë A. (2001). A sensorimotor account of vision and visual consciousness. Behavioral and Brain Sciences, 24(5), 939-973.

[11] 魏屹东.人工智能会超越人类智能吗?人文杂志,2022,314(06): 88-98.

[12] Helmholtz, H. (1860/1962). Handbuch der physiologischen optik: Vol. 3. New York: Dover.

[13] Brown H., Friston K., & Bestmann S. (2011). Active inference, attention, and motor preparation. Frontiers in Psychology, 2, 218.

[14] Hohwy J. (2020). New directions in predictive processing. Mind & Language, 35(2), 209-223.

[15] Constant A., Clark A., Kirchhoff M., et al. (2022). Extended Active Inference: Constructing Predictive Cognition beyond Skulls. Mind & Language, 37(3), 373-394.

[16] Hohwy J. (2013). The predictive mind. OUP Oxford.

[17] Friston K. (2009). The free-energy principle: a rough guide to the brain? Trends in Cognitive Sciences, 13(7), 293-301.

[18] Pennartz C. M. A. (2015). The brain's representational power: on consciousness and the integration of modalities. MIT Press.

[19] Kim J. (1999). Making sense of emergence. Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition, 95(1/2), 3-36.

[20] 魏屹东,苏圆娟.认知涌现论:一种新的意识整体论,科学技术哲学研究, 2021,38(2), 1-6.

[21] Chalmers D. J. (2000). What is a neural correlate of consciousness? In Neural Correlates of Consciousness: Empirical and Conceptual Questions (p. 17-39).

[22] Wiese W., & Friston K. J. (2021). The neural correlates of consciousness under the free energy principle: From computational correlates to computational explanation. Philosophy and the Mind Sciences, 2.

[23] Dehaene S., & Changeux J. P. (2011). Experimental and theoretical approaches to conscious processing. Neuron, 70 (2), 200-227.

[24] Cleeremans A. (2005). Computational correlates of consciousness. Progress in Brain Research, 150, 81-98.

[25] Mongillo G., Barak O., & Tsodyks M. (2008). Synaptic theory of working memory. Science, 319(5869), 1543-1546.

[26] Solms M. (2019). The hard problem of consciousness and the free energy principle. Frontiers in Psychology, 9(2714), 1-16.

[27] Chalmers D. J. (2011). A computational foundation for the study of cognition. Journal of Cognitive Science, 12(4), 325-359.

[28] Friston K. (2010). The free-energy principle: a unified brain theory? Nature Reviews Neuroscience, 11(2), 127-138.

[29] 魏屹东.适应性表征:架构自然认知与人工认知的统一范畴.哲学研究,2019(9):117.