**Chapter 2: Can consciousness be explained?**

**Pat and Rick drank a lot of beer in ‘our bar’. They were frustrated from their research in the elusive phenomenon of consciousness. Pat said to Rick, "There's an elusive creature in ‘our bar’ that a large expedition of scientists hasn't been able to grasp. That shows ‘our bar’ is very special, spooky." And Rick replied, "Not sure. Tell me, did these scientists drink a few bottles of beer? No! That's what you're saying, so how did they even start their research without beer? No wonder they didn't get it."**

**A brief review of the attempts to solve the consciousness-problem**

The review of the relevant fields of research (e.g., psychology, philosophy, neurophysiology, computer science) has given me the impression that, to date, no theory, based on the neurophysiological processes in the brain, has solved successfully the CΨ-problem. In the current chapter, Chapter 2, I will focus mainly on the attempt to decipher the puzzle of CΨ: the attempt to develop a neurophysiological theory (TC) that try to explain this puzzle. Other attempts, such as the electromagnetic field approach to solving the CΨ-problem, will be discussed later, see especially chapter 6.

There is no TC similar to scientific theories that explain energy transformations, such as friction and heat, potential and kinetic energy, electricity and magnetism; or theories explaining how substances change, such as water being formed from chemical bond of hydrogen and oxygen, and how electrolysis can break down water into these two gases. In order to support this impression, I address the question: is it possible to explain CΨ by using the concepts applied in the natural sciences? Further, I will break this general question into three important sub-questions:

1. Can human behavior be explained mechanistically without using the concept of CΨ as an explanatory factor?
2. Has TC been developed that explains successfully the relationship between neurophysiological processes in the brain and CΨ?
3. Is it possible to reduce mentalisticexplanations grounded in the concept of CΨ to mechanistic explanations (such as neurophysiological explanations)?

An affirmative answer to any one of these three questions would mean that CΨ is not necessary to explain human (or animal) behavior. Standard scientific methodologies can be adapted in order to research and explain the various types of psychological behavior. A negative answer to all three questions indicates the need for serious discussion and the development of a new explanatory approach that is applicable to the areas of psychological research addressing the CΨ-problem (see chapters 4, 5).

**Can human behavior be explained purely mechanistically?**

Many researchers who adhere to approaches such as behaviorism, cognitive psychology, and cognitive-neurophysiological psychology believe that any behavior can be explained mechanistically, by those models and theories that are accepted in the natural sciences. They see no need to invoke the concept of CΨ as an explanatory factor (Rakover 2018). I call this approach the “CΨ-unnecessity.”

It seems to me that the CΨ-unnecessity approach is based on the following “multi-explanation” argument. It starts with the mathematical observation that for each group of observations (results) presented as a set of points in the Cartesian system of Y = f (x), there is an infinite number of appropriate mathematical functions that accurately fit this set of points. It continues with the reasonable assumption that each function may express a mechanistic explanation of the results that occur under certain relevant conditions. And it ends with the probable hypothesis that any psychological phenomenon can be explained by using at least one mechanistic explanation that is acceptable by the natural sciences (an explanation that is not based on mentalistic concepts such as, will, belief, intention). This means that it is completely unnecessary to invoke CΨ as an explanatory concept, because CΨ is a dubious and un-objective concept, and because we always have at our disposal a mechanistic explanation that is consistent with the methodological approach of the sciences. Here are several quotes that argue for the CΨ-unnecessity:

* “When animal consciousness is dismissed as superfluous, we must ask whether the dismissal refers to consciousness as a phenomenon to be explained or as an explanatory device. The most plausible answer is that consciousness is superfluous in the latter role. Anything that can be explained by it can be explained equally well without it.” (Radner and Radner 1989, p. 206)
* “The goal is to formulate an explanation which does not involve any thinking or sentient agent in its premises. The explanans should involve no one who is acting as an intelligent, sentient force, guiding behavior in the right direction.” (Keijzer 2001, p. 26)
* “There is no prediction we can make that if the animal has consciousness it should do X but not conscious it should do Y.” (Dawkins 1995, p. 139)
* “conscious inessentialism” which is “… the view that for any activity *i* performed in any cognitive domain *d*, even if we do *i* consciously, *i* can in principle be done non-consciously.” (Flanagan 1992, p. 129)

A similar approach can be based on the following two philosophical theories. The first is the ‘identity theory’, which suggests that there is an identity between a mental state (MS) and a neurophysiological state (NS). The second is ‘functionalism’, which suggests that a MS is functionally defined, and it can be realized in various ways (see, e.g., Robb and Heil, 2019). These two approaches lead to a conclusion similar to the CΨ-unnecessity approach: the explanatory factor is not a MS but rather a NS, which can be causally linked to human or animal behavior. In other words, an accepted causal model in which the explanatory concept is a NS and not a MS provides the explanation for a behavior.

I reject the CΨ-unnecessity approach for the following reasons. First, without the concept of CΨ, it is difficult to understand human or animal behavior. Behaviors are saturated with different attributes of CΨ, such as will, belief, intention, which cannot be dismissed easily. For example, in my book *To Understand a Cat: Methodology and Philosophy* (Rakover 2007), I describe numerous behavioral episodes related to the relationship between a housecat Max and my wife Aviva or myself. These episodes would be hard to explain without appealing to the cat’s will, purpose, or intention. It would be difficult to understand our interactions with Max by using only mechanistic explanations consistent with the prevalent methodology in the sciences, such as certain explanations of reflexive behavior, instinctive, or automatic.

Second, I reject the CΨ-unnecessity approach because the aforementioned ‘multi-explanation’ is based on a misguided, implicit assumption. This assumption is related to the methodology of psychology. Psychology provides explanations for behaviors under certain conditions that are publicly observable according to the adopted scientific methodology of the natural sciences (Rakover 1990). This type of observation in behavior strips away any meaning attributed to the behavior by the individual, and every element of CΨ. The behavioral indices represent behaviors, which are analyzed statistically and reported in professional journals, do not represent the individuals’ goals, meanings, or intentions. Psychology consider only publicly observable behaviors (e.g., physical movements) performed by the study participants. Indices such as percent of correct responses or reaction time do not take into account the state of CΨ of the experiment participants. These indices represent only their motor responses, such as whether they press the button corresponding to the correct or incorrect response, or how long it took from the time a stimulus appeared until a response was given. These are responses, which a sophisticated robot is capable of performing, although as a machine it is not conscious. For example, sophisticated software can ascertain whether a given face is an image that appears in a database of criminals. In the same way, a human witness can go through police files and identify the face of the criminal he recently saw in a crime scene. In both cases the result is the same: either the suspect was successfully identified or not. However, the critical question is whether the forensics software understands its actions and consequences in the same way that a human being does. I don’t think it does. Assuming this analysis is accurate; it can be proposed that because behavioral indices are not imbued with CΨ, there is no demand for an explanation to be based on this concept. Therefore, mechanistic explanations (which do not address CΨ at all) may offer quite satisfactory explanations for such objective indices. However, it should be noted that this kind of explanation does not attribute any meaning to behavior; it could equally well explain the behavior of a robot or a zombie.

Third, the philosophical literature is replete with suggestions for solving the mind-body or brain-consciousness problem. All of these have been highly criticized (e.g., Kim 1996; Rakover 2018; Robb and Heil 2019), as seen in the following examples.

Jackson (1982) published a well-known article showing that even if we know everything there is to know about a phenomenon, this knowledge still cannot explain conscious experience. Jackson posed an interesting thought experiment about a vision specialist he called Mary. Mary was an expert in physics, chemistry, physiology, and the psychology of color vision. But Mary had lived her entire life in a black-and-white environment. One day, Mary came out of her black-and-white surroundings and saw, for the first time in her life, the color red. She experienced and learned something new, which her flawless scientific knowledge hadn’t made possible. She learned what it meant to see the color red, and to consciously experience the sense of seeing color.

The next example is the article by Nisbett and Wilson (1977) in which they asserted that many experiments in social psychology and decision making have shown that the study participants were unaware of the stimulus, response, effect of the stimulus on the response, or their own relevant cognitiveprocesses. Their argument was based on the finding that the participants’ explanations for their own behavior were incorrect. The article drew many criticisms, including one I wrote (Rakover 1983), which I will briefly describe here. The participants in the Nisbett and Wilson experiments were aware of the representations in their own minds. However, in a way similar to scientists who offered incorrect explanatory hypotheses for the phenomenon under investigation, the participants offered a wrong explanation to their own behavior. Therefore, it is not necessarily true that the participants were unaware of the stimulus and the response, since they simply may propose an unsuccessful hypothesis about the connection between these two.

The final example is that of Velmans (1991), who suggested that in many cases, such as speech, the individual is aware of the behavior after it has been performed. On this basis, Velmans concluded that CΨ is of minimal importance in information processing, and that unconscious processes mediate between the appearance of the stimulus and the participants’ responses. This article inspired extensive criticism and discussion. In my opinion, Velmans’ argument is incorrect, because CΨ as a process that mediates between the stimulus and the response is crucial (see Rakover 1996). I suggested the “mental-pool” thought experiment, based on the existing psychological knowledge that the amount of conscious information is limited, while the amount of unconscious information is unlimited. I assert that information received from a stimulus is initially processed at the unconscious level, then it goes to the level of CΨ, in which a response is emerged and supervised.

**Has TC been developed that explains the relationship between neurophysiological processes in the brain and CΨ?**

Based on a broad, in-depth review of the attempts to construct a brain-consciousness theory, Cosmelli, Lachaux, and Thompson (2007) concluded that to date, no neurophysiological explanation of CΨ has been developed. There has only been success in discovering associations and correlations between neurophysiological and cognitive measures: “…the neurodynamical approach works at the level of correlations, albeit refined ones.” (ibid, 763).

As mentioned previously, Chalmers (1996, 1997) proposed a distinction between easy and hard problems of CΨ: “The easy problems of consciousness are those that seem directly susceptible to the standard methods of cognitive science, whereby a phenomenon is explained in terms of computational or neural mechanisms. The hard problems are those that seem to resist those methods.” (1997, p. 9). In Chalmers’ opinion, the easy problems are related to explanations of behaviors such as response to stimulation, discrimination, focus of attention, organization of information, verbalizing thoughts, and voluntary control of behavior. Such behaviors can be explained via cognitive and neurophysiological mechanisms, that is, via mechanisms that execute the functions of these behaviors. As Chalmers wrote: “To explain access and reportability, for example, we need only specify the mechanism by which information about internal states is retrieved and made available for verbal report,” (Chalmers’ 1997, p. 10).

Chalmers’ approach has been widely challenged by scholars (see for example Shear 1997). I also have severe reservations of his approach, because I think that most behaviors are saturated with CΨ. Therefore, providing an explanation for an easy problem is no simpler than for a hard one. As I argued above, cognitive psychology provides mechanistic explanations for the easy problems, which correspond to behaviors with CΨ removed from them. In other words, the explanations for the easy problems are explanations for the behaviors of zombies or robots, and not of human behaviors that are saturated with CΨ.

Given the above, I will now move on to propose a brief review of several popular and influential TCs that have attempted (without success) to explain CΨ. In Seth and Bayne's (2022) article: “theories of consciousness”, over twenty-two theories are reviewed (see table 1, p. 441). Sattin et al. (2021) analyzed in their scoping review 29 theories of CΨ. This large collection of TCs arouses two important comments. The first is related to the obvious fact that I will not be able to critically review all these theories. This is the reason why I decided to concentrate on reviewing a limited number of important and popular theories. The review will give the reader a broad view of the conceptual infrastructure, from which most of these theories grew. I will concentrate on theories that anchor CΨ in the cognitive-representational approach, in the neurophysiology of the brain, in the electromagnetic field of the brain, and in the quantum processes of the brain.

The second comment regards the fact that this multiplicity of theories is an amazing phenomenon that indicates that there is no agreement among the researchers regarding the following questions: Which brain mechanism is the one that creates CΨ? Which TC is accepted by the scientific community? (Note that these theories are not always aimed at the same phenomenal aspect of CΨ.) Furthermore, it turns out that it is very difficult to decide between a small numbers of theories of CΨ and suggest which theory has received the most empirical support (or disconfirmations). Yaron et al. (2022) tested four popular theories by analyzing 412 experiments: The global neuronal workspace (GNW) theory, the Integrated Information Theory (IIT), the recurrent processing theory (RAT), and the higher-order thought (HOT) theory. A general analysis of all these experiments revealed some rather discouraging conclusions about the state of CΨ research. First, it became clear that research in CΨ is highly biased towards reporting positive results (confirmation bias). Second, most of the findings were interpreted according to the theory favored by the researchers in a post-hoc manner. Thirdly, it became clear that the choice of the research methodology for examining the theory determined the results, which were supporting the theory in question. These three problems explain why so many TCs continue to be developed at the same time. Furthermore, although the four theories under consideration were based on different neurophysiological processes in the brain, the analysis of the results showed the following. "… a remarkable heterogeneity of findings, which by itself is not compatible with the predictions of any of the theories (that is, none of the theories would predict such a vast neural activation as a marker of consciousness). At the anatomical level, a map of all reported findings seems to suggest that almost the entire brain has been implicated in conscious perception.” (p. 598).

As can be expected, additional meta-analysis studies related to the empirical research of theories of CΨ were done, for example, Promet and Bachmann (2022). Despite the serious methodological problems found by Yaron et al. (2022) in the CΨ studies they reviewed, Promet and Bachmann (2022) found large enough differences between them and Yaron et al. that allow them to reach the following conclusion. They found that the empirical support for theories of CΨ is as follows: GNWT (i.e., Global Neuronal Workspace theory) receives the most empirical support, followed by Predictive Processing and finally the IIT. (For a description and discussion of these theories, see Promet and Bachmann, 2022.) Nevertheless, it seems to me that the serious problems found by Yaron et al. lead me to adopt the following cautious approach.

Given the above methodological conclusions of Yaron et al. (2022), I decided that for the time being it is better to minimize the evaluation of the theories of CΨ on the basis of empirical findings, which usually tend to support the theory in question. Thus, I will attempt concentrate more on the philosophical-methodological criticisms directed against these theories, but I will also mention strong theory-contradictory findings. I will briefly review three theories, which were discussed by Yaron et al. (2022): IIT, GWT-GNWT, and HOT. However, as mentioned above, to give a slightly more complete picture of the subject, I will also add a brief discussion of the theories that anchor CΨ to electromagnetic and quantum processes in the brain. This brief review, therefore, will give the reader a broad theoretical view of the CΨ-problem. (Note that due to the elaboration of the various issues of CΨ, the main discussion of electromagnetic theories and HOT will be done later on, in chapter 6.)

*Integrated Information Theory (IIT)*. One of the most important TCs that appears in the professional literature is Giulio Tononi’s and his colleagues’, the Integrated Information Theory (IIT). This theory has attracted much interest in recent years, as well as receiving severe criticisms (see for example Fallon 2019, 2020; Tononi 2015; Tononi, Boly, Massimini, and Koch 2016). Consciousness is produced by maximum integration of information among parts of the brain, and it is indexed mathematically by Φ (Phi). The greater the increase in the integration of the information, the higher is the individual's state of CΨ. IIT is a kind of theory that causally connects between maximal integrated information and CΨ, and is based on the following three ideas: (1) a specification of the CΨ properties (conceived as axioms). (2) a specification of the properties of the physical substrate (neurophysiology of the brain or any other possible material) necessary for the realization of CΨ properties (conceived as postulates). And (3) a determination of a maximal causal connection between CΨ and a particular type of physically-processed information that is formulated by Φ, which measures the degree of CΨ.

IIT takes the following theoretical approach. It assumes first that CΨ is an existing phenomenon and it has a complex structure. Second, CΨ is directed towards certain things. Third, CΨ carries information. Fourth, CΨ is unified (one cannot experience the red color of a tomato separately from its shape). Fifth, CΨ has boundaries because it is aimed at one particular thing and not another.

Based on these agreed-upon CΨ traits (viewed as axioms) the theory elaborates on the traits that a physical system must have in order to realize these CΨ axioms. For example, the axiom of CΨ being an information-bearing trait suggests the postulate that the physical or neurophysiological system must be based on elements that can combine in order to create a structure based on cause and effect, aiming at the realization of a specific state of CΨ.

According to the IIT, when a person sees a cat on a couch, a “conceptual structure” is created in that person’s mind based on a number of particular concepts and their relationships, that is, integrated information. This structure represents what is seen (the cat on the couch) and is treated by the physical system, the “physical substrate” that functions according to the above-stated postulates (in humans, this physical substrate is the neurophysiology of the brain). The IIT is based on the identity between the conscious experience (seeing the cat on the sofa) and the conceptual structure of this experience, which is realized by the neurophysiology of the brain. Thus, the IIT suggests that CΨ is identical to a particular type of integrated information, which is realized by a specific physical system. This system can be divided into subgroups with various cause-and-effect structures. The subgroup with the maximum cause-effect that cannot be reduced to its components expresses the maximally irreducible conceptual structure (MICS).

According to the IIT, the MICS is the state of CΨ. Therefore, it can be said that a physical system, which manifests the above postulates, is intrinsically endowed with CΨ, just as mass has the inherent trait of gravity. This system is a mechanism that works according to cause-and-effect, and thus organizes information. The degree of complexity of the MICS can be represented by Φ. The more complex the MICS, the greater the level of CΨ, that is, the greater the Φ that numerically expresses the MICS. When MICS is maximal, the size of Φ expresses the maximum degree of CΨ.

The IIT has been supported by empirical findings in several studies. For example, it has been shown that in a state of deep, dreamless sleep (with reduced brain activity), indices close to Φ show values smaller than those when awake. Also, in accordance with IIT, it has been found that injury to the cerebellum does not impair CΨ, because the cells in the cerebellum do not interact with each other (e.g., Tononi, 2015).

As mentioned, the IIT has received a great deal of criticism (see summary and discussion in Fallon 2019. See also Blackmore, 2013; Van Gulick, 2022; Wu and Morales, 2024). Here I would like to emphasize the following points. First, since CΨ is founded on physical substrate (e.g., the neurophysiology of the brain) it is possible to construct a mechanical system that meets all the requirements of the IIT. Thus, one may suggest that this mechanical system has CΨ. There may even be a situation where the Φ of this system expresses greater CΨ than that of human. This possibility is completely contrary to intuition and common sense. Most people know that a machine is just a machine, and to this day, no device has been invented, including highly sophisticated computers that display even a hint of CΨ (e.g., Fallon, 2019). Furthermore, I propose here an argument that I call "the live-creature’s correlation", which is based on the following observations. It turns out that all live creatures (e.g., humans, apes, dogs, cats, fishes, etc.) who have a brain and nervous-system, even in the most primitive structure, have a certain degree of CΨ (see discussion on the subject in chapter 1). In contrast, other entities that are not alive (such as, stones, metals, soil etc.) and which are alive but have not even a very primitive brain and nervous system, (e.g., flowers, trees, etc.) have not CΨ. (That is, I do not accept the panpsychism approach to CΨ.) Therefore, it is difficult to propose that a mechanical system, which is constructed by inorganic material, will have CΨ, even if this system fulfills precisely the requirements of IIT. The response to such criticism from Tononi, Boly, Massimini, and Koch (2016), who are willing to accept the possibility of such a computer, is particularly interesting:

“Intriguingly, IIT allows for certain simple systems, such as grid-like architectures, similar to topographically organized areas in the human posterior cortex, to be highly conscious even when not engaging in any intelligent behavior.” (460. See also page 458).

Second, it is possible to argue against the IIT’s use of the concept of information. Given that the concept of information depends on a person’s CΨ, it follows that the IIT’s attempt to understand CΨ via the concept of information is circular. The response to this criticism is that the concept of information, according to the IIT, is built into the neurophysiological substrate, which handles the conceptual structures. Thus Tononi, Boly, Massimini, and Koch (2016, 457) write: “In IIT, information is causal and intrinsic: it is assessed from the intrinsic perspective of a system based on how its mechanism and present state affect the probability of its own past and future states (cause-effect power).” (see also Fallon, 2020). It seems to me that this kind of response, based on the assumption that information is evaluated from an intrinsic perspective nested within the neurophysiological system itself, suggests the existence of a tiny person (homunculus) within this system, which assesses past, present and future states, and so the problem of CΨ remains.

Third, Doerig et al. (2019) and Doering et al. (2021) proposes the “unfolding argument” against IIT. Based on the computational theorem that for the same input-output function there exist two different networks (recurrent and feedforward) when one is conceived as conscious and the other as unconscious, it is shown that either IIT is falsified or unscientific, it resides outside the scientific methodology (however, for counter arguments see Kleiner, 2020; Tsuchiya et al. 2020).

Finally, it is worth noting that in the year 2023, 124 researchers published a declaratory letter accusing IIT of being pseudoscientific theory and not a scientific theory, mainly because in their opinion this theory is not empirically testable. As can be expected, this statement provoked counter-reactions published in the professional literature (e.g., Bayne, 2023; Fleming, et al. 2023; Lenharo, 2023).

If the criticisms of IIT presented in this overview contain indication of truth, it seems that this theory has not solved the riddle of CΨ. What the theory has been able to do is to locate a particular type of neurophysiological structure in the brain that changes in its degree of activation correlatewith changes in level of CΨ. This I do find very interesting.

## *Global workspace theory (GWT) and Global neuronal workspace theory (GNWT) of consciousness*.

## Baars (1988, 2017) proposed a cognitive theory of CΨ and Dehaene and colleagues developed it on the basis of brain neurophysiological processes (e.g., Dehaene, 2014; Dehaene and Naccache, 2001; Mashour et al. 2020). The fundamental idea behind the GWT is this: a mental state (MS) becomes conscious when it is 'globally available' for treatments by a large number of cognitive processes (such as attention, memory, evaluation, report and response). In other words, information, which is represented by a MS, becomes conscious when it is broadcasted by the cognitive global workspace. A local MS (e.g., local processes such as sight, hearing) becomes conscious when it is treated by the global workspace (e.g., Baars, 1988, 2017). Baars (2017, p. 235) writes, “Consciousness seems to be the publicity organ of the brain… It is a capacity for accessing, disseminating, and exchanging information, and for exercising global coordination and control.” … “Global Workspace Theory (GWT) suggests that the brain has a fleeting integrative capacity that enables access between functions that are otherwise separate. … A sizable body of evidence suggests that consciousness is the primary agent of such a global acess function in humans and other animals.” So, one may propose that CΨ is the broadcaster of MSs to many cognitive processes within the cognitive workspace, a process that make these MSs conscious. The GWT’s important goal is to explain the differences between conscious and unconscious MSs and processes. It can be grasped by the analogy of a theater. What becomes CΨ is similar to an actor whom a spotlight brings out of the darkness. One may suggest that attention acts like a spotlight, which brings an unconscious MS (or process) into CΨ on the global workspace. Given this analogy, it is possible to suggest that external and internal stimuli (memories) in the cognitive system compete with each other to win an entrance ticket to CΨ. This concept is similar to the concept of short-term memory, which is the domain of the limited capacity of subjective awareness. It should be noted, however, that Baars (2017) suggests replacing the metaphor of Theater with that of the Internet (the world-wide web) because communication on the WWW is multidirectional and better reflects neural processes in the brain.

## The GWT is a theoretical framework that enables the development of various models that aim to explain certain cognitive phenomena, such as backward-masking, attentional blink, and binocular rivalry, including measurements of brain processes with the help of EEG and fMRI.

##  Dehaene and Naccache (2001) proposed that CΨ is required in situations where information must be held for a certain time, when new operational combinations is generated, and when goal-directed behavior is needed. They hypothesize that a large number of cerebral networks process information simultaneously and unconsciously. The information becomes conscious when the neural networks representing this information are intensified into a coherent activity associated with many neurons distributed through the brain. In their opinion, the global availability of this information in the cognitive-neurophysiological system is what constitutes the subjective feeling of CΨ. The main neurophysiological system in the brain that is related to the transformation of information into conscious one is based on the prefrontal cortex, the anterior cingulate and the brain areas that connect them. (see Figure 2.1). The immediate global neuronal operation ignites certain systems into activity and others into inhibition. This global informational accessibility is conceived of as the originator of CΨ.

##  As can be expected, GWT has received some interesting criticisms (e.g., Blackmore, 2013; Van Gulick, 2022; Wu and Morales, 2024). I would like to point up the followings. Today most researchers welcome any research approach that attempts to unravel the problem of CΨ. (I do not accept Irvine’s, 2013, approach, which proposes to eliminate the concept of CΨ.) However, I doubt whether GWT actually manages to solve our problem. It does not seem that this theory offers a mechanism that describes how CΨ grows out of the neurophysiological processes in the brain. Baars’ (2017) answer is that GWT deals with the distinction between a brain system that supports CΨ (the cortex) and a system that does not support it (the cerebellum). Therefore, this criticism misses its target. This counter-argument seems to me a weak and to some extent evasive answer. Why? Because in light of the above description of GWT, this theory seems to suggest much more than an associative, correlational connection between CΨ and certain brain processes. The many functions that GWT attributes to CΨ and its comprehensive and complicated relationships with neurophysiological processes in the brain, all of these seem to be founded on deeper connections than what is reflected by the concept of correlation.

##  GWT links CΨ to many complex neurophysiological processes that express multidirectional communication in the cortex. In fact, CΨ is involved in all the important functions of the brain. This global broadcasting description of the relationship between brain processes and CΨ is inconsistent with the amazing observations that show that humans without the cortex are in an intact state of CΨ (e.g., Doerig, Schurger & Herzog, 2021; Merker, 2007). Merker (2007, p. 50) writes, “The evidence and functional arguments reviewed in this article are not easily reconciled with an exclusive identification of the cerebral cortex as the medium of consciousness.” He proposes to consider the upper brainstem mechanisms as potential neural processes for CΨ (see Figure 2.1).

## *Higher-Order (HO) theories of consciousness*

## As stated above, a fairly extensive discussion of one of the popular theories of HO theories, the Higher-Order Thought (HOT) theory, will be given later in Chapter 6. Here I will briefly outline the basic principles of the current approach. The main idea of HO theories of CΨ is that an unconscious mental state (MS) becomes a conscious when it becomes the target of, represented by a meta-MS or a higher MS (MS\*). For example, the MS of a cat, MS(cat), a first-order MS, becomes conscious when this MS is represented by MS\*. For example, the thought ‘I see a cat’, which can be symbolized by MS\*[MS(cat)], makes the representation of the stimulus in the world, first-order MS(cat) conscious. In this case, the representation of the cat, MS(cat), is the target of the representation of the MS\*, the thought ‘I see a cat’ (e.g.,Carruthers, 2017; Rosental & Weisberg, 2008; Van Gulick, 2022). Findings related to neurophysiological studies in the brain indicate the following distinction. While the MS(cat) is handled by the visual cortex (in the back of the brain), the MS\* that represents this MS, the MS\*[MS(cat)], is handled by the prefrontal cortex (in the front of the brain) (e.g., Brown, Lau & LeDoux, 2019; Lau & Rosental, 2011) (see Figure 2.1).

##  The HO theoretical approach has several different interesting variations of which I will mention only two: The HOT theory that I will discuss later on, and the Lycan’s (2004) Higher-Order Perception (HOP) theory (called also is the ‘inner sense’ theory of CΨ). According to HOP theory the inner sense handles the MS(cat) perceptually and therefore it becomes conscious. This theory received severe criticisms which I will discuss very briefly. If an inner-sense mechanism is hypothesized, it is very difficult to believe in (a) its existence and (b) the fact that it does not make errors similar to those made by the operation of known sensory mechanisms (e.g., Carruthers, 2017; Rosental & Weisberg, 2008).

##  The main goal of HO theories is to explain how a non-conscious MS becomes conscious. Therefore, the criticism against these theories that they fail to explain how CΨ is created by the neurophysiological processes in the brain is less pressing. Nevertheless, I would like to highlight here the following criticisms. First, it is not clear how these theories deal with the fact that we consciously feel huge differences between, for example, the excitement in the perception of the beauty of the Mona Lisa and the feeling of horror at the sight of execution by slaughter, shooting or hanging. The proposed solution, that these differences are explained by the higher-order MS is not satisfactory: It is not clear how the match is made between MS and MS\*.

##  Second, it is not clear how the neurophysiological process in the brain represents the stimulus appearing in the world, for example, a cat. And it is not clear why exactly a particular MS\* represents the MS(cat) and how this process is done.

##  Third, if the represented-MS by the representing-MS\* makes the represented-MS conscious, then why wouldn't the thought of a cup of coffee cause the cup of coffee to become conscious? The answer that a cup of coffee is not MS is correct, but brings us back to the previous problem and everything that related to it: how does the brain represent stimuli and what becomes representing-MS\* and what represented-Ms? (e.g., Van Gulick, 2022).

##  And finally, if the neurophysiological findings in the brain indicates that the prefrontal cortex is the one that treats MS\* (higher-order MS), the following is not clear. How this theoretical-neurophysiological approach copes with the findings described above, according to which humans are in a state of CΨ even though they have no cortex (e.g., they were born without a cortex). And how the present approach cope with the findings that CΨ was not affected when the prefrontal cortex was suffered serious injuries or electrical stimulation (e.g., Wu and Morales, 2024)?

## *Quantum theories of consciousness*

## As you might expect, there are several attempts to explain CΨ based on quantum theory (e.g., Atmanspacher, 2017, 2024; Blackmore, 2013; Van Gulick, 2022). Here I will review very briefly the development of the theory of Hameroff & Penrose (1996, 2014), called "Orchested objective reduction” (Orch Or) theory (Penrose is a 2020 Nobel laureate in physics and Hameroff is an anesthesiologist). The fundamental idea behind quantum theories of CΨ is as follows: classical physics that deals with everyday large bodies is not suitable to solve the problem of CΨ. What may offer an understanding of this problem is the quantum physics that succeeds in describing the subatomic world (although it is inconsistent neither with our everyday intuition nor with the conceptual infrastructure of classical physics).

## The starting point of Hameroff & Penrose’s theory is Penrose’s (1989, 1994) suggestion that CΨ cannot be described computationally, that is, that CΨ is non-computational and conscious understanding cannot be explained by computer programs. This proposal is justified in various ways including the Gӧdel’s incompleteness theorem (very briefly, the theorem shows that for a formal system, e.g., natural numbers, there are true statements that cannot be proven or disproven within that system). Given this, Penrose (1989, 1994) appealed to Quantum theory and proposed that CΨ is emerged by a coherent collapse of the superposition states (a condition in which a quantum system can exist in multiple states simultaneously) that is described by Schrӧdinger equation (a [partial differential equation](https://en.wikipedia.org/wiki/Partial_differential_equation) that describe the evolution over time of the [wave function](https://en.wikipedia.org/wiki/Wave_function) of a quantum-mechanical system). Penrose proposed a new theory “objective reduction”, according to which gravitational process causes this quantum collapse that extends over the brain in a “quantum coherence” form. (The general assumption is that the quantum system goes from a state of superposition (multiple states) to a single state as a result of an observation or measurement of the system.) Hameroff suggested that CΨ is generated by the quantum coherence in the microtubules of the brain (these are very small and thin tubes that are part of the body's cell structure). The quantum coherence in microtubules can explain certain properties of CΨ such as being incomputable, unity, and certain effects of anesthesia.

##  As you may surmise, the theory of Penrose and Hameroff has provoked many criticisms (e.g., Atmanspacher, 2017, 2024; Blackmore, 2013; Van Gulick, 2022). Here I will briefly comment on the following critiques. First comment concerns Penrose and Hameroff (1995) respond to the criticisms leveled at their theory by Grush and Churchland (1995). Their article is divided into two parts. In the first part, Penrose's response appears, and in the second part, Hameroff's. In the first part, Penrose complains that Grush and Churchland haven't read his books (especially the one from 1994) because if they had, they would have immediately seen that he discussed at great length the many problems that Grush and Churchland raise in their article. (For example, the argument of non-computability in mathematical thinking based on Gӧdel’s incompleteness theorem.) And so Penrose writes, “But it would seem from what G&C say that they have not even read, and certainly not understood, these arguments.” (p .98). He adds in footnote 1, “It is extremely frustrating, considering the efforts involved in writing a book with particularly detailed arguments, when these arguments are simply treated as though they did not exist.” (p. 98). In the second part of the article, Hameroff technically answers Grush and Churchland biological criticisms. For example, Grush and Churchland believe that CΨ does not depend on microtubules because CΨ appears even without their activity caused by the administration of the drug colchicine (colchicine is a [medication](https://en.wikipedia.org/wiki/Medication) used to treat such diseases as [gout](https://en.wikipedia.org/wiki/Gout), a kind of inflammatory arthritis). Hameroff counter argues that even giving a large amount of colchicine directly into the brain has very little effect on the microtubules in the brain.

##  The second comment relate to Derakhshani et al. (2022) test of Penrose and Hameroff’s Orch OR quantum theory of CΨ. They tested a variant of this theory, which is based on a simple version of Diósi-Penrose theory of gravity-related dynamical wavefunction collapse. Given this, they analyzed most reasonable tubulin superposition scenarios, and reached the conclusion that this simple version of theory is highly implausible. It is worth noting here that the authors emphasize that these results refer to the refutation of a variation of the Orch OR theory and not to the theory itself.

##  Well, as a general summary of the current review of the various theories for explaining CΨ, the following two conclusions can be proposed.

## First, there is no TC that is free from severe criticisms. It seems that there is no TC that the scientific community accepts as the paradigmatic theory (in the kuhnian paradigm sense).

## Secondly, there are theories that have received more scientific attention than others have. These theories have been discussed in the professional literature more than others have. However, the impression is that this fact does not make them more successful than other theories with regard to answering the fundamental question: How do brain processes create CΨ?

**Is it Possible to Reduce Cognitive Explanations Grounded in the Concept of Consciousness to Mechanistic Explanations (Such as Neurophysiological Explanations)?**

The answer to this question is complicated, due to the difficulty of drawing a parallel between a process of reduction between *explanatory theories*, a topic that has been discussed extensively in professional literature, and a process of reduction between *models of explanation*. Why? Because *explanatory models* are not scientific theories, such as the theories of perception, learning, and recognition. They are essentially procedures that guide the researcher regarding how to offer explanations in certain areas of research. For example, in one case, the explanatory model deals with a phenomenon as a special case of a general law, but in the other case, the model shows how the current phenomenon is generated by a certain mechanism. It is therefore difficult to see how one type of instructions can be rationally reduced to another type of instructions. The result is that the answer to the present question will not focus on the attempt to show how a *mentalistic explanation model* (such as a goal-based, teleological explanation) is based on a *mechanistic explanation model* (such as a causal explanation). Rather, the answer will focus on the possibility of a reduction of a *mentalistic theory* (based on subjective concepts of CΨ related to the individual’s inner world, will, beliefs, intentions, feelings, emotions, etc.) toa *mechanistic theory* (grounded in objective concepts related to physics, chemistry, physiology, computer processes, etc.). The rationale behind this reductionist effort is, *inter alia*, that the field of psychology has based itself on the accepted methodology of the natural sciences, and thus favors mechanistic explanations over mentalistic explanations.

I will first briefly address the problems associated with psycho-neural reduction. Then I will address other efforts to grasp mentalistic concepts via mechanistic concepts, such as substituting mentalistic explanations (goal-based, teleological explanations) with mechanistic explanations (causal explanations based on neurophysiological processes). The discussion of these cases will clearly show that there is still no acceptable method to reduce or convert a mentalist theory to a mechanistic one.

I will start with the question: Can a psychological theory based on concepts related to the individual’s inner world (desire, belief, intention, purpose, emotion) be reduced to a neurophysiological theory? (see, for example, Kim 1998; Rakover 1990; van Riel and Van Gulick, 2024). First, I will briefly explain what procedure is used when attempting to reduce Theory A to Theory B. In order not to get tangled in this complicated subject, I will describe only Nagel’s (1961) classical approach. Accordingly, theory A, called the reduced theory (TR), can be reduced to theory B, called the basic reducing theory (TB) when it is possible to derive, deduce TR from TB, along with the relevant bridge laws linking the concepts of these two theories. Bridge laws areusually seen as identities between the concepts of TR and TB. For example, in reducing thermodynamics to mechanical statistics, it has been suggested that the concept of temperature is identical tothe concept of kinetic energy.

Several arguments have been made against the possibility of conducting a psycho-neurological reduction between a psychological TR and a neurophysiological TB. I will briefly discuss the famous argument called multiple realization (e.g., Fodor 1974, 1998). Consider for example the mental state called “pain” (MSPain). This MS is functionally defined: to prevent or reduce injury to the individual. Various neurophysiological brain processes in humans, monkeys, dogs, cats, fish, etc can realize MSPain. Furthermore, MSPain representing behaviors undertaken in response to pain stimuli could be manifest among robots, via different materials otherthan those that exist in humans. Using this argument, it would be impossible to find a bridge law between the concepts of TB and TR, because the concepts of TR can be implemented in different ways and via different processes. Therefore, if a psychophysiological bridge law cannot be found, the aim of reducing a psychological theory to a neurophysiological theory cannot be fulfilled.

Another argument against the possibility of conducting a neurophysiological reduction is grounded in the requirement for ‘unit equivalency’. According to this requirement, the combination of measurement units on one side of the equation (expressing a law or theory) must be the same as the combination of measurement units on the other side of the equation (see Rakover 2002, 2018).

The bridge law cannot meet this requirement. Why? Because the units of measurement for the relevant neurophysiological processes that appear on one side of the equation are completely different from the units of measurement that appear on the other side of the equation, which are associated with MSs and mental processes. (In fact, no one knows yet how to measure MSs directly. They can only be indirectly interpreted from observations of behavior and verbal reporting. For this reason, it is difficult to see how the product of a constant can be used to meet the unit equivalence requirement.) It is difficult to find a uniform and common scale for these two types of measurements. Therefore, in this respect we do not discuss the bridge law based on the identity between concepts of two different theories. At best, we are addressing associations or correlations (this is like seeking a correlation between the size of tomatoes and the height of giraffes).

To end this section of the discussion, I will show that it is difficult if not impossible to translate a goal-oriented explanation (an action is undertaken to fulfill a specific purpose) into a causal explanation (an action is caused by a specific factor). Consider the following example of a goal-oriented explanation: Uri drove his car from Haifa to Tel Aviv *because* he wanted to meet his girlfriend, Yaffa. It is possible to offer a translation of this purposive explanation into a causal explanation: The desire to meet Yaffa in Tel Aviv *caused* Uri to drive his car from Haifa to Tel Aviv. This translation is based on the simple idea of transforming the *goal* into the *cause* of action. However, as will become clear, this translation raises major problems that undermine the very idea of providing a causal explanation for a goal-based explanation. First, it may be seen as natural to transform the goal into the cause of the action by identifying the person’s CΨ of the goal as being responsible for the action (driving the car). However, this raises the mind-body problem: how does a mental process (thought) lead to a behavior? And vice versa: how does behavior trigger mental processes? To date, there has been no satisfactory solution to the mind-body or the CΨ-brain problem. Therefore, it appears that the proposed translation does not solve the problem of explanation, but merely re-introduces an old problem. Just as we cannot understand how a future event can explain a present or past event, we do not understand how a mental event generates a physical event.

Secondly, an attempt to translate a goal-oriented explanation into a causal explanation will encounter extremely difficult methodological problems. Cause and effect are viewed as different and separate events. For example, a stone that fell from the sky (a cause) broke the window in the house (an effect). This distinction is impossible in the case of a purposive, goal-directedexplanation, because there is a dependence of the concepts that appear in the explanation: the individual’s will, beliefs, and action.

In the example given above, the cause is Uri’s desire to see Yaffa in Tel Aviv; the result is Uri’s action of driving from Haifa to Tel Aviv. The trip is the realization of Uri’s belief that this trip will enable him to achieve his goal of meeting Yaffa in Tel Aviv. In other words, the mental reason for the travel represents both Uri’s goal and his intention, because intention is always aimed towards a specific goal. In the same way, it can be asserted that what is done to realizethe goal (Uri’s travel) is not a purposeless act with no intention; it is a meaningful act fueled by Uri’s will and intention. In the case of the ‘stone & window’, the cause and the effect are two separate events, whereas in the case of Uri’s travel, his reason and the action are intertwined and cannot be separated completely.

In conclusion, then, the answer to the basic question of the chapter (is there a solution to the CΨ**-**question or to its classical formulation, the body/mind problem) is negative. To date, a convincing solution to this problem has not yet been proposed. This is not only a personal conclusion of the writer of the present book, but also a conclusion of several researchers and philosophers. See for example the following quotes, which I collected while reading the relevant literature:

* “Whatever our mental functioning may be, there seems to be no serious reason to believe that it is explainable by our physics and chemistry.” (Putnam, 1975, p. 297)
* “We have been trying for a long time to solve the mind-body problem. It has stubbornly resisted our best efforts. The mystery persists. I think the time has come to admit candidly that we cannot resolve the mystery.” (McGinn, 1989, p. 349)
* “The sensation of color cannot be accounted for by the physicist’s objec­tive picture of light-waves. Could the physiologist account for it, if he had fuller knowledge than he has of the processes in the retina and the nervous processes set up by them in the optical nerve bundles and the brain? I do not think so.” (Scrödinger, 1992, p. 154)
* “To be brutally honest, scientists do not yet have even the remotest idea of how visual experiences—or indeed any other kinds of experiences—arise from physical events in the brain.” (Palmer, 1999, p. 618)
* “The reason the mind-body problem does not go away, despite our being clear about the options in responding to it, is because of the constant battle between common sense, which favors the view that the mental is a basic feature of reality, and the pull to see it as an authoritative deliverance of science that this is not so. We find ourselves constantly pulled between these two poles, unable to see our minds as nothing over and above the physical, unwilling to see the universe as containing anything not expli­cable in terms of its basic, apparently non-mental, constituents.” (Ludwig, 2003, pp. 29–31)
* “Even if we accept the familiar idea that minds are somehow dependent on brains, we have no clear idea of the nature of this dependence. The mental-physical relation appears utterly mysterious.” (Heil, 2003, p. 217)
* “The problem of consciousness is completely intractable. We will never understand consciousness in the deeply satisfying way we’ve come to expect from our sciences.” (Dietrich & Hardcastle, 2005, p. 1 [the open­ing sentence])
* “Are neuroreductionist explanations of cognition possible?” (This is the title of Uttal’s [2014] paper, to which he proposed a “probably not” answer for methodological, conceptual, and empirical reasons [p. 37].)

**A tentative proposal: Consciousness as an explanatory but unexplained Concept**

Consciousness refers to a mental-behavioral phenomenon occurring in humans (and animals). It can be described as the inner world that only the individual is able to feel and observe. So far, there seems to be no satisfactory explanation for this phenomenon orthe relationship between CΨ and the brain. That is, no accepted TC has yet been found to explain the relationship between the body and mind, between the neurophysiology of thebrain and CΨ. Any TC that attempts to describe the mind in terms of the neurophysiology of the brain has not borne fruit. All proposed theories are problematic, including theories based on causality, identity, functionalism, multiple realization, or supervenience (which proposes that there is no change in the MS [(mental state] without a change in the NS [neurophysiological state]). In each case, serious problems were raised against the proposed theory and it was rejected by many researchers.

In my opinion, the attempt toexplain CΨ via neurophysiological processes in the brain raises a significant problem. On the one hand, the attempt to reduce the MS to the NS is tempting, since then behavior as a whole (including conscious behavior) can be explained by relying on the neurophysiological causal theories already known to science. On the other hand, if the explanation for behavior is covered entirely by the NS, what is the value and importance of an explanation based on the MS? Why is the MS necessary? These questions run counter to common sense. That is, everyday experiences, which suggest that an individuals’ behavior is accounted for by referring to one’s inner world, to CΨ. Here is a very simple example: I went to the movies *because* I wanted to see the actress Gal Gadot in the movie ‘Wonder Woman’.

In other words, if everything is explained by neurophysiological processes, the NS, then the MS has no explanatory value. It seems completely unnecessary. This theoretical approach inevitably brings us to the outdated and largely rejected philosophical approach of epiphenomenalism that proposes that the MS is explanatorily ineffectual. As Kim (2002) writes in the précis to his previous book (Kim 1998, p. 643):

“To summarize, then, the problem of mental causation is solvable for cognitive/intentional mental properties. But it is not solvable for the qualitative or phenomenal characters of conscious experience. We are therefore left without an explanation of how qualia can be causally efficacious; perhaps, we must learn to live with qualia epiphenomenalism.”

I do not accept Kim’s (2002) view that conscious experiences are epiphenomena. I propose an anti-epiphenomenalism approach, namely that CΨ has an effect on behavior, it is effectual and is not an epiphenomenon. Moreover, CΨ should be regarded as a primary explanatory concept, precisely because a satisfactory explanation for it has yet to be discovered. That is, in light of the current chapter that no TC has yet been found that explains CΨ on the basis of the neurophysiology of the brain, the following suggestion is warranted: Let us consider CΨ as a primary, theoretical explanatory concept that cannot be explained by more basic concepts. This proposal requires the following clarifications.

First, I do not suggest here that because there has been no explanation for the problem of CΨ that therefore it is reasonable to assume that CΨ in animals, especially humans, may be considered a novel force in nature. Such an assumption would create enormous confusion in the conventional infrastructure of mechanistic explanations (e.g., energy conservation laws may be broken) (see for similar arguments in Carroll 2016).

Second, I do not claim that CΨ is independent of brain processes in humans or animals. Rather, I emphasize that no theory has yet been found that explains the relationship between these two. I only propose that CΨ is an explanatory but unexplained concept. That is, it is a primary explanatory concept and can be characterized as referring to the state of anti-epiphenomenalism.

Meanwhile, then, as a conclusion of the above, I suggest that the fundamental qualities of CΨ as anti-epiphenomenalism concept are (for further theoretical development see the following chapters):

1. Consciousness MSs exists, to varying degrees, in every individual human (and also in other living beings).
2. Only the individual himself/herselfis consciously aware of the content of the various representations appearing in his or her own mind (the MSs).
3. Without CΨ, humans would function purely on a physiological level and would be considered to be in a state similar to that of a plant.
4. Consciousness can affect one’s physical functioning.
5. Consciousness is influenced by physical phenomena (e.g., sensory stimuli such as light and sound elicit in the individual conscious feelings typically to these stimuli).
6. Consciousness is dependent on the normal functioning of the brain.
7. Consciousness enables the experiences of sensory stimuli, the aliveness-feel (a sense of being alive), life-meanings, and understanding.

 **References**

Atmanspacher, H. (2017). Quantum approaches to brain and mind. In S.

Schneider and M. Velmans (Eds.), *The Blackwell companion of consciousness (2ed)*. John Wiley & Sons.

Atmanspacher, H. (2024). Quantum Approaches to Consciousness. In E. N.

Zalta & U. Nodelman (eds.), *The Stanford Encyclopedia of Philosophy*,

URL = <https://plato.stanford.edu/archives/sum2024/entries/qt-consciousness/>.

Baars, B. J. (1988). *A cognitive theory of consciousness*. Ney York: Cambridge

University Press.

Baars, B. J. (2017). *The global workspace theory of consciousness: Predictions*

*and results*. In S. Schneider and M. Velmans (Eds.), *The Blackwell*

*companion to consciousness (2ed)*. West Sussex: John Wiley & Sons.

Bayne, T. (2023). Where does consciousness start? Debate is heating up over

some of the leading theories. *Science* (alert).

Blackmore, S. (2013). *Consciousness: An introduction (3ed)*. London:

Routlsdge.

Brown, R., Lau, H. & LeDoux, J. E. (2019). Understanding the Higher-Order

approach to consciousness. *Trends in Cognitive Sciences*, 23, 754-768.

Carroll, S. M. (2016). *The Big Picture: On the Origins of Life, Meaning, and the Universe Itself*. London: Oneworld Publications.

Carruthers, P. (2017). Higher-Order theories of consciousness. In S. Schneider

and M. Velmans (Eds.), *The Blackwell companion of consciousness (2ed)*. John Wiley & Sons.

Chalmers, D. J. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press.

Chalmers, D. J. (1997). Facing Up to the Problem of Consciousness. In J. Shear (Ed.), *Explaining Consciousness: The Hard Problem*, (9-32). Cambridge, MA: The MIT Press.

Cosmelli, D., Lachaux, J.-P. and Thompson, E. (2007). Neurodynamical Approaches to Consciousness. In P. Zelazo, M. Moscovich, and E. Thompson (Eds.), *The Cambridge Handbook of Consciousness,* (731-774). Cambridge and New York: Cambridge University Press.

Dawkins, M., S. (1995). *Unravelling Animal Behavior (2nd ed.)*. Essex: Longman Scientific & Technical.

Dehaene, S. (2014). *Consciousness and the brain: Deciphering how the brain*

*codes our thoughts.* Viking.

Dehaene S. and Naccache L. (2001). Towards a cognitive neuroscience of

 consciousness: Basic evidence and a workspace framework. *Cognition*. 79,

1–37.

Derakhshani, M., Diósi, L., Laubenstein, M. Piscicchia, K. and Curceanu, C.

(2022). At the crossroad of the search for spontaneous radiation and the Orch OR consciousness theory. *Physics of Life Reviews*, 42, 8-14.

Dietrich, E. & Hardcastle, V. G. (2005). *Sisyphus’s boulder: Consciousness and*

*the limits of the knowable.* Amsterdam/Philadelphia: John Benjamins.

Doering, A., Schuster, A. & Herzog, M. H. (2021). Hard criteria for empirical

theories of consciousness. *Cognitive Neuroscience*, 12, 41-62.

Doering, A., Schuster, A., Hess, K. & Herzog, M. H. (2019). The unfolding

argument: Why IIT and other causal structure theories cannot exolain consciousness. *Consciousness and Cognition*,72, 49-59.

# Fallon, F. (2024). Integrated Information Theory of Consciousness. *The Internet*

# *Encyclopedia of Philosophy.*

Fallon, F. (2020). Integrated Information Theory, Searle, and the Arbitrariness

Question. *Review of Philosophy and Psychology*. 11, 629–645.

Flanagan, O. J. (1992). *Consciousness Reconsidered.* Cambridge, MA: The MIT Press.

Fleming, S. M., Frith, C., Goodale, M., Lau, H., LeDoux, J. E. and Lee, A. L. F.

(2023). [The Integrated Information Theory of Consciousness as Pseudoscience](https://psyarxiv.com/zsr78/). *PsyArXiv (Preprint)*. [doi](https://en.wikipedia.org/wiki/Doi_%28identifier%29%22%20%5Co%20%22Doi%20%28identifier%29):[10.31234/osf.io/zsr78](https://doi.org/10.31234/osf.io/zsr78)

Fodor, J. A. (1974). Special Sciences, or the Disunity of Science as a Working Hypothesis. *Synthese* 28, 97-115.

Fodor, J. A. (1998). Special Sciences: Still Autonomous After All These Years. In J. Fodor (Ed.), *In Critical Condition: Polemical Essays on Cognitive Science and Philosophy of Mind*, (9-24). Cambridge, MA: The MITPress.

Grush, R. and Churchland, P. S. (1995). Gaps in Penrose’s toilings’. *Journal of*

*Consciousness Studies*, 2, 10-29.

Hameroff, S. R. and Penrose, R. (1996). Conscious events as orchestrated

 spacetime selections. *Journal of Consciousness Studies*, 3(1): 36–53.

Hameroff, S. R. and Penrose, R. (2014). Consciousness in the universe: A

review of the Orch OR theory (with commentaries and replies). *Physics*

*of Life Reviews*, 11, 39–112.

Heil, J. (2003). Mental causation. In S. P. Stich & T.

A. Warfield (Eds.), *The Blackwell guide to philosophy of mind (Pp. 214-234)*. Malden, MA: Blackwell.

# Irvine, E. (2013). *Consciousness as a scientific concept: A philosophy of*

# *Science perspective*. Dordrecht: Springer.

Kim, J. (1996). *Philosophy of mind.* Boulder, CO.: Westview Press.

Kim, J. (1998). *Mind in a Physical World: An Essay ON THE Mind-Body Problem and Mental Causation.* Cambridge, MA: The MIT press.

Kim, J. (2002). Précis of Mind in a Physical World. *Philosophy and Phenomenological Research* 65, 640-43.

Keijzer, F. (2001). *Representation and Behavior*. Cambridge, MA: The MIT Press.

Kleiner, J. (2020). Brain states matter. A reply to the unfolding argument.

*Consciousness and Cognition*, 85, October 2020, 102981

Lau, H. & Rosenthal, D. (2011). Empirical support for higher-order theories of

conscious awareness. *Trends in Cognitive Sciences*, 15, 365-373.

Lenharo, M. (2023). Consciousness theory slammed as ‘pseudoscience’ –

sparking uproar. *Nature* (news).

Ludwig, K. (2003). The mind-body problem: An overview. In S. P. Stich & T.

A. Warfield (Eds.), *The Blackwell guide to philosophy of mind (Pp. 1-46)*. Malden, MA: Blackwell.

Lycan, W. G. (2004). The superiority of HOP to HOT. In R. J.

Gennaro (Ed.), *Higher-order theories of consciousness: An anthology* (pp. 93-113). John Benjamins.

Merker, B. (2007) Consciousness without a cerebral cortex: A challenge

 for neuroscience and medicine. *Brain and Behavioral Sciences*, 30,

63–81

Mashour, G. A., Roelfsema, P., Changeux, J.-P. and Dehaene, S. (2020).

Conscious processing and the global neuronal workspace hypothesis. *Neuron*, 105, 776-798.

McGinn, C. (1989). Can we solve the mind-body problem? *Mind*, 98, 349-366.

Nagel, E. (1961). *The Structure of Science: Problems in the Logic of Scientific Explanation*. New York: Harcourt-Brace & World Inc.

Palmer, S. E. (1999). *Vision science: Photons to phenomenology*. Cambridge,

MA: the MIT Press.

Penrose, R. (1989) *The Emperor’s New Mind*. Oxford: Oxford University Press.

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

Penrose, R. and Hameroff, S. (1995). What ‘gaps’? Reply to Grush and

Churchland. *Journal of Consciousness Studies*, 2, 98-111.

Promet, L. and Bachman, T. (2022). A comparative analysis of empirical

theories of consciousness. *Psychology of Consciousness: Theory, research and practice.* http://doi.org/10.1037/cns0000341

Putnam, H. (1975). *Mind, Language and reality: Philosophical papers. V.2.*

Cambridge: Cambridge University Press.

Radner, D. and Radner, M. (1989). *Animal Consciousness.* New York: Prometheus Books.

Rakover, S. S. (1983). Hypothesizing from Introspections: A Model for the Role of Mental Entities in Psychological Explanation. *Journal for the Theory of Social Behavior* 13: 211-30.

Rakover, S. S. (1990). *Metapsychology: Missing Links in Behavior, Mind and Science*. New York: Paragon/Solomon.

Rakover, S. S. (1996). The Place of Consciousness in the Information Processing Approach: The Mental-Pool Thought Experiment. *Behavioral and Brain Sciences* 19: 535-36.

Rakover, S. S. (2002). Scientific Rules of the Game and the Mind/Body: A Critique Based on the Theory of Measurement. *Journal of Consciousness Studies* 9: 52-58.

Rakover, S. S. (2007). *To Understand a Cat: Methodology and Philosophy.* Amsterdam/Philadelphia: John Benjamins.

 Rakover, S. S. (2018). *How to Explain Behavior: A Critical Review and New Approach.* Lanham: Lexington Books.

Robb, D. and Heil, J. (2019). Mental Causation. In by E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*, <https://plato.stanford.edu/archives/sum2019/entries/mental-causation/>

Rosental, D. & Weisberg, J. (2008). Higher-order theories of consciousness.

*Scholarpedia,* 3(5):4407

Sattin, D., Magnani, F. G., Bartesaghi, L., Caputo, M., Fittipaldo, V.,

Cacciatore, M., Picozzi, M. and Leonardi, M. (2021). Theoretical models of consciousness: A scoping review. *Brain Sciences*, 11, 535.

Scrӧdinger, E. (1992). *What is life? The physical aspects of the living cell.*

Cambridge: Cambridge University Press.

Seth, A. K. & Bayne, T. (2022). Theories of consciousness. *Nature*

*Reviews/Neurosciences*, 23, 439-452.

Shear, J. (Ed.). (1997). *Explaining Consciousness: The Hard Problem*. Cambridge, MA: The MIT Press.

Tononi, G. (2015). Integrated information theory*. Scholarpedia*, 10(1):4164.

Tononi, G., Boly, M., Massimini, M. and Koch. C. (2016). Integrated Information Theory: From Consciousness to its Physical Substrate. *Nature Reviews Neuroscience* 17, 450-61.

Tsuchiya, N., Andrillon, T. and Haum, A. (2020). A reply to “the unfolding

argument”: Beyond functionalism/behaviorism and toward a science of causal structure theories of consciousness. *Consciousness and Cognition*, 79 (2020) 102877 102981.

Yaron, I., Mellon, L., Pitts, M. and Mudrik, L. (2022). The ConTraSt database

for analyzing and comaring empirical studies of consciousness theories. *Nature Human Behaviour*, 6, 593-604.

Uttal, W. R. (2014). Are neuroreductionist explanations of cognition possible?

*Behavior and Philosophy*, 41, 37-64.

Van Gulick, R. (2022). Consciousness. In E. N. Zalta & U. Nodelman (eds.),

*The Stanford Encyclopedia of Philosophy* URL = <https://plato.stanford.edu/archives/win2022/entries/consciousness/>

van Riel, R. and Van Gulick, R. (2024). Scientific Reduction. In E. N. Zalta &

U. Nodelman (eds.), *The Stanford Encyclopedia of Philosophy*, URL = <https://plato.stanford.edu/archives/spr2024/entries/scientific-reduction/>.

Wu, W. and Morales, J. (2024). The neuroscience of consciousness. In E. N.

Zalta & U. Nodelman (eds.), *The Stanford Encyclopedia of Philosophy*,URL = <https://plato.stanford.edu/archives/sum2024/entries/consciousness-neuroscience/>.

**I'll try to find a better drawing over time.**

Figure 2.1 depicts schematically several areas of the human brain relevant to the topic of discussion of the current chapter (see text) 