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

**Pat and Rick drank a lot of beer in ‘their bar’. They were frustrated by their research into 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 even.” And Rick replied, “I’m not sure about that. Tell me, did these scientists drink any beer? No? Is that 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 attempts to solve the problem of consciousness**

A careful 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 neurophysiological processes in the brain has successfully solved the problem of consciousness. In the current chapter, I will focus mainly on one type of attempt that has been made to decipher the puzzle of CΨ, namely the attempt to develop a neurophysiological theory (TC). Other attempts to solve the problem, such as the electromagnetic field approach, will be discussed later (see Chapter 6).

No TC has provided a satisfactory explanation for consciousness in the same way that scientific theories have explained energy transformations (such as friction and heat, potential and kinetic energy, and electricity and magnetism) or changes to substances (such as the formation of water from chemical bonds of hydrogen and oxygen, and the process of electrolysis by which water is split into these two gases). Therefore, in what follows I address the question: Is it possible to explain CΨ using the concepts applied in the natural sciences? Furthermore, I will break this general question down into three important sub-questions:

1. Can human behavior be explained mechanistically without using the concept of CΨ as an explanatory factor?
2. Has a TC been developed that successfully explains 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 research to explain various types of psychological behavior. A negative answer to all three questions would indicate the need for the development of a new explanatory approach that applies to psychological research addressing the problem of consciousness (see chapters 4 and 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 using models and theories 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-nonnecessity.”

It seems to me that the CΨ-nonncessity approach is based on the following “multi-explanation” argument. It starts with the mathematical observation that for each group of observations (or 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 from the natural sciences (an explanation that is not based on mentalistic concepts such as will, belief, or intention). This means that it is completely unnecessary to appeal to CΨ as an explanatory concept because CΨ is a dubious and unobjective concept, and because we always have at our disposal a mechanistic explanation consistent with the methodological approach of the sciences. Here are several quotes that support CΨ-nonnecessity:

* “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 & Radner, 1989, p. 206)
* “The goal is to formulate an explanation which does not involve any thinking or sentient agent in its premises. The explanations 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)

The following two philosophical theories take a similar approach. The first is the ‘identity theory’, which suggests that mental states (MSs) are identical to neurophysiological states (NSs). The second is ‘functionalism’, which proposes that every MS is functionally defined and can be realized in various ways (see, e.g., Robb & Heil, 2019). These two approaches share a similar conclusion to that of the CΨ-nonnecessity approach: NSs, rather than MSs, are causally linked to human or animal behavior and therefore provide the explanatory factor.

I reject the CΨ-nonnecessity approach for the following reasons. First, it is difficult to understand human or animal behavior without the concept of CΨ. Behaviors are saturated with different attributes of CΨ, such as will, belief, and 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 house cat (Max) and its owners. These episodes would be hard to explain without appealing to the cat’s will, purpose, or intention. It would be difficult to understand the interactions between Max and its owners using only mechanistic explanations consistent with the prevalent scientific methodology, such as those linked to reflexes, instincts, or automatic behavior.

Second, I reject the CΨ-nonnecessity approach because the aforementioned ‘multi-explanation’ is based on a misguided, implicit assumption related to the methodology of psychology. Psychology provides explanations for publicly observable behaviors under certain conditions according to the adopted scientific methodology of the natural sciences (Rakover, 1990). This type of observation strips away any meaning the individual attributes to their behavior and every element of CΨ. Behavioral indices represent behaviors which are analyzed statistically and reported in professional journals, but they do not represent the individual’s goals, meanings, or intentions. Experiments in psychology consider only publicly observable behaviors (e.g., physical movements) performed by study participants. Indices such as correct response rates or reaction time do not take into account the subject’s state of CΨ. These indices represent only motor responses, such as whether a participant presses the button corresponding to the correct or incorrect response, or how much time elapses between the presentation of a stimulus and the response. A sophisticated robot is capable of such responses, although as a machine it is not conscious. For example, sophisticated software can ascertain whether a received facial image appears in a criminal database. In the same way, a human witness can go through police files and identify the face of a criminal. In both cases, the result is the same: Either the suspect is successfully identified or not. However, the critical question is whether the forensics software understands its actions and their consequences in the same way that a human being does. I do not think so. Assuming this is accurate, it can be proposed that because behavioral indices are not imbued with CΨ, there is no need for an explanation based on this concept. Mechanistic explanations (which do not address CΨ at all) may offer quite satisfactory explanations for such objective indices. However, this kind of explanation does not attribute any meaning to behavior; it could equally well explain the behavior of a robot or 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 & Heil, 2019), including the following examples.

Jackson (1982) published an 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 scientist called Mary, who lives in an entirely black-and-white environment. She is, nevertheless, an expert in the neurophysiology of color vision. One day, Mary leaves her black-and-white surroundings and sees, for the first time in her life, the color red. Despite her flawless scientific knowledge, she experiences and learns something new. She learns what it means to see the color red and to consciously experience color perception.

The next example concerns Nisbett and Wilson’s (1977) study on decision-making, in which subjects were asked to evaluate the attributes of an individual shown in videotaped interviews. The researchers reported that the study participants were unaware of the influence of their global evaluations on their ratings of attributes. They took this as evidence of unconscious alteration of judgments. The article drew much criticism (see, e.g., Rakover, 1983). It was suggested that the participants in the Nisbett and Wilson experiment were aware of the representations they formed in their minds; they simply proposed an unsuccessful hypothesis about the connection between the stimulus and their response.

The final example comes from Velmans (1991), who suggested that in many actions, such as speech, the individual is aware of their behavior only after it has been performed. He posited that CΨ is of minimal importance in information processing and that unconscious processes mediate between the appearance of a stimulus and the participant’s responses. This article provoked strong criticism and discussion. In my opinion, Velmans’ argument is incorrect because CΨ *is* the crucial process that mediates between the stimulus and the response (see Rakover, 1996). I suggested the “mental-pool” thought experiment based on the existing psychological knowledge that conscious information is limited, while unconscious information is unlimited. I assert that information received from a stimulus is initially processed unconsciously before reaching the level of CΨ, which is responsible for producing a supervised response.

**Has a TC been developed to explain 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 et al. (2007) concluded that, to date, no neurophysiological explanation of CΨ has been successfully 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 the 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 behavioral explanations, such as response to sensory stimuli, discrimination, focus of attention, organization of information, verbalization of thoughts, and voluntary control of behavior. Such behaviors can be explained by cognitive and neurophysiological 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, e.g., Shear, 1997). I also have strong reservations because I think that most behaviors are saturated with CΨ. Therefore, explaining an easy problem is no simpler than a hard one. As I argued above, cognitive psychology provides mechanistic explanations for the easy problems, which correspond to behaviors with CΨ removed. In other words, explanations for the easy problems apply to the behaviors of zombies or robots and not to human behaviors saturated with CΨ.

Given the above, I will now propose a brief review of several popular and influential TCs that have attempted (without success) to explain CΨ. My review draws on Seth and Bayne’s (2022) “Theories of consciousness”, encompassing over twenty theories (see Table 1, p. 441), and Sattin et al.’s (2021) scoping review of 29 theories of CΨ. Of course, I will not be able to critically review them all. However, I 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; the neurophysiology of the brain; the electromagnetic field of the brain; and the quantum processes of the brain.

This multiplicity of theories highlights the clear lack of agreement among researchers regarding the following questions: Which brain mechanism 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.) It is very difficult to 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 (RPT), 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, research in CΨ is highly biased towards reporting positive results (confirmation bias). Second, most of the findings were interpreted according to the researchers’ preferred theory in a post hoc manner. Thirdly, the choice of research methodology made it possible to determine results that supported the theory in question. These three problems explain why so many TCs are continually being developed. Although the four theories under consideration are based on different neurophysiological processes in the brain, the analysis of the results showed “… 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 have been conducted.

Promet and Bachmann (2022), for example, found that global neuronal workspace theory (GNWT) has received the most empirical support, followed by predictive processing, and finally IIT. (For a description and discussion of these theories, see Promet and Bachmann, 2022.) However, Yaron et al. highlighted serious methodological problems with these theories, which has led me to adopt a cautious approach.

I decided that for the time being it is better to focus less on evaluating theories of CΨ based on empirical findings, which usually tend to support the theory in question. Thus, I will concentrate more on the philosophical-methodological criticisms directed against these theories, and I will also mention strong contradictory findings. I will briefly review three theories 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 problem of consciousness. (Note that due to the elaboration of various issues, the main discussion of electromagnetic theories and HOT can be found in Chapter 6.)

*Integrated information theory (IIT)*

One of the most important TCs in the professional literature is Tononi et al.’s integrated information theory (IIT). This has been severely criticized in recent years (see, e.g., Fallon, 2019, 2020; Tononi, 2015; Tononi et al., 2016). According to IIT, consciousness is integrated information in certain brain areas, and levels of information integration can be measured mathematically by phi (Φ). The greater the integrated information value, the higher the individual’s level of CΨ. This causal connection depends on: (i) a specification of the properties of consciousness (conceived as axioms); (ii) a specification of the properties of the physical substrate (neurophysiology of the brain or any other possible material) necessary for consciousness (conceived as postulates); and (iii) a determination of a maximal causal connection between CΨ and a particular type of physically processed information formulated by Φ, which measures the degree of CΨ.

IIT assumes first that CΨ is an existing phenomenon and has a complex structure. Second, CΨ is directed toward 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 Tononi et al.’s axioms of consciousness, the theory elaborates on the traits that a physical system must have to realize these axioms. For example, the axiom that CΨ is an information-bearing trait suggests the postulate that the physical or neurophysiological system requires a combination of elements within a cause-effect structure aimed at the realization of a specific state of CΨ.

According to IIT, when a person sees a cat on a couch, a “conceptual structure” is created in that person’s mind based on particular concepts and the relationships between them: 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). IIT identifies conscious experience (seeing the cat on the sofa) with the conceptual structure of this experience, which is realized by the neurophysiology of the brain. In short, 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 IIT, a conscious state is the MICS. Therefore, it can be said that a physical system that manifests the above postulates is intrinsically endowed with CΨ, just as mass has the inherent trait of gravity. This system is a cause-and-effect mechanism that organizes information. The degree of complexity of the MICS can be numerically expressed by Φ. The more complex the MICS, the higher the level of CΨ. When MICS is maximal, the size of Φ expresses the maximum degree of CΨ.

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), Φ values are lower than those associated with wakefulness. Also, 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 earlier, IIT has received much criticism (see summary and discussion in Fallon, 2019. See also Blackmore, 2013; van Gulick, 2022; Wu and Morales, 2024). With this in mind, I would like to emphasize the following objections to the theory.

First, if CΨ is founded on the physical substrate (e.g., the neurophysiology of the brain), is it not possible to construct a mechanical system that meets all the requirements of IIT? Thus, one may suggest that this mechanical system has CΨ and that it is possible for the Φ of such a system to express greater CΨ than that of humans. However, this hypothesis is completely contrary to intuition and common sense. Most people know that a machine is just a machine, and, to this day, no device is capable of displaying a hint of CΨ, not even highly sophisticated computers (e.g., Fallon, 2019). Furthermore, I would like to propose an argument here that I call “the live-creatures correlation”, which is based on the following observations. All live creatures (e.g., humans, apes, dogs, cats, fishes, etc.) that have a brain and a nervous system, even those with the most primitive structures, have a certain degree of CΨ (see discussion on the subject in Chapter 1). In contrast, entities that are not alive (such as stones, metals, soil, etc.) or alive but without even a very primitive brain and nervous system (e.g., flowers, trees, etc.) do not possess CΨ. (I do not accept the panpsychist approach to CΨ.) Therefore, it is difficult to propose that a mechanical system constructed from inorganic material can have CΨ, even if this system precisely fulfills the requirements of IIT. Tononi et al.’s (2016) response to such criticism is particularly interesting since they are willing to accept the possibility of such a computer:

“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.” (p. 460; see also p. 458).

Second, it is possible to argue against IIT’s use of the concept of information. Given that the concept of information depends on a person’s CΨ, it follows that IIT’s attempt to understand CΨ through the concept of information is circular. One response to this criticism is that the concept of information, according to IIT, is built into the neurophysiological substrate, which deals with conceptual structures. Thus, Tononi et al. (2016, p. 457) wrote: “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 homunculus within this system, which assesses past, present, and future states, and so the problem of CΨ remains.

Third, Doerig et al. (2019, 2021) proposed 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) which support conscious and unconscious cognitive processes, it can be posited that IIT is either false or unscientific since it resides outside the scientific methodology (for counter-arguments, see Kleiner, 2020; Tsuchiya et al., 2020).

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

If the criticisms of IIT presented in this overview are valid, it seems that this theory has not solved the riddle of CΨ. Interestingly, what the theory has been able to do is locate a particular type of neurophysiological structure in the brain with varying degrees of activation that correlatewith changes in the level of CΨ.

## *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 this based on brain neurophysiological processes (e.g., Dehaene, 2014; Dehaene and Naccache, 2001; Mashour et al. 2020). The fundamental idea behind GWT is this: information becomes conscious when it is made ‘globally available’ to a large number of cognitive processes (such as attention, memory, evaluation, report, and response). In other words, information, which is represented by an MS, becomes conscious when it is broadcast by the cognitive global workspace. A local MS (e.g., local processes such as sight and hearing) becomes conscious when it is treated by the global workspace (e.g., Baars, 1988, 2017). Baars (2017, p. 235) wrote, “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 access function in humans and other animals.” So, one may propose that CΨ broadcasts MSs to many cognitive processes within the cognitive workspace, a process that makes these MSs conscious. The important goal of GWT is to explain the differences between conscious and unconscious MSs and processes. It can be grasped by the analogy of a theater. What becomes conscious is similar to the actor on stage who comes under the spotlight and emerges from 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 that of short-term memory, which is linked to the limited capacity of subjective awareness. It should be noted, however, that Baars (2017) also suggested replacing the theater metaphor with that of the Internet (the World Wide Web) because communication in this domain is multidirectional and better reflects neural processes in the brain.

## The theoretical framework of GWT has enabled the development of various models to explain certain cognitive phenomena, such as backward masking, the attentional blink, and binocular rivalry, using EEG and MRI to measure brain processes.

##  Dehaene and Naccache (2001) proposed that CΨ is required in situations where information must be held for a certain length of time, when new operational combinations are generated, and when goal-directed behavior is needed. They hypothesized that many cerebral networks process information simultaneously and unconsciously. The information becomes conscious when the neural networks representing this information intensify to generate a coherent activity. In this view, 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 related to the transformation of information (from unconscious to conscious) is located in the prefrontal cortex, the anterior cingulate cortex, 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 criticism (e.g., Blackmore, 2013; van Gulick, 2022; Wu and Morales, 2024). I would like to point up the following. Today, most researchers welcome any research approach that attempts to unravel the problem of CΨ. (I do not accept Irvine’s [2013] proposal to eliminate the concept of CΨ.) However, I doubt whether GWT manages to solve our problem. It does not seem to offer a mechanism that describes how CΨ develops from 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). However, this counter-argument seems to be a weak and somewhat 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 seem to be founded on connections that run deeper than 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 observations that CΨ is preserved in humans without the cortex (e.g., Doerig et al., 2021; Merker, 2007). Merker (2007, p. 50) wrote, “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 proposed considering the upper brainstem mechanisms as potential neural processes for CΨ (see Figure 2.1).

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

## The main idea behind HO theories of CΨ is that an unconscious mental state (MS) becomes a conscious one when it is represented by a meta-MS or a higher-order MS (MS\*). For example, the MS of perceiving a cat, which is a first-order MS that can be symbolized by MS(cat), becomes conscious when this MS is represented by a thought about it, that is, MS\*. For example, the thought ‘I see a cat’, which can be symbolized by MS\*[MS(cat)], forms a representation of the stimulus in the world, the first-order MS(cat). Thus, the representation of the cat, MS(cat), is the target of the MS\*, the thought ‘I see a cat’ (e.g., Carruthers, 2017; Rosenthal & Weisberg, 2008; van Gulick, 2022). The findings of neurophysiological studies of the brain indicate the following distinction. While MS(cat) is handled by the visual cortex (located at the back of the brain), the MS\* that represents this MS, the MS\*[MS(cat)], is handled by the prefrontal cortex (at the front of the brain) (e.g., Brown et al., 2019; Lau & Rosenthal, 2011) (see Figure 2.1).

##  There are several interesting variations of this theoretical approach, of which I will mention only two: HOT theory, which I will discuss later in Chapter 6, as mentioned earlier, and Lycan’s (2004) higher-order perception (HOP) theory (also called the ‘inner sense’ theory of CΨ). According to HOP, the inner sense handles MS(cat) perceptually and therefore it becomes conscious. This theory has received severe criticism, 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 like those made by known sensory mechanisms (e.g., Carruthers, 2017; Rosenthal & Weisberg, 2008).

##  The main goal of HO theories is to explain how a non-conscious MS becomes conscious. Therefore, their failure to explain how CΨ is created by neurophysiological processes in the brain does not mean they should be discarded. Nevertheless, I would now like to highlight other criticisms. First, it is not clear how these theories deal with the fact that we consciously feel huge differences between, for example, the thrill of perceiving the beauty of the Mona Lisa and the horror of seeing an execution. 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, HO theories do not explain how a neurophysiological process in the brain can represent a stimulus appearing in the real world, for example, a cat. And how exactly does a particular MS\* represent the MS(cat)? How does this process occur?

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

##  And finally, if neurophysiological findings indicate that it is the prefrontal cortex that treats MS\*s (higher-order MSs), how does this theoretical-neurophysiological approach explain why humans born without a cortex are conscious? Furthermore, how do HO theories stand up to recent findings that CΨ is not affected when the prefrontal cortex suffers serious injuries or electrical stimulation (e.g., Wu & Morales, 2024)?

## *Quantum theories of consciousness*

## There have been several attempts to explain CΨ based on quantum theory (e.g., Atmanspacher, 2017, 2024; Blackmore, 2013; van Gulick, 2022). The fundamental idea behind quantum theories of CΨ is as follows: classical physics, dealing with everyday large bodies, is unsuitable for solving the problem of CΨ. The domain of quantum physics, on the other hand, may lead to a better understanding. This body of scientific laws succeeds in describing the subatomic world (although it is inconsistent with our everyday intuition and with the conceptual infrastructure of classical physics). Here I will very briefly review Hameroff & Penrose’s (1996, 2014) “Orchestrated objective reduction” (Orch OR) theory (Penrose is a 2020 Nobel laureate in physics and Hameroff is an anesthesiologist).

## The starting point was Penrose’s (1989, 1994) suggestion that CΨ cannot be described computationally: CΨ is non-computational and conscious understanding cannot be explained by computer programs. This proposal has been justified in various ways including 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). Penrose (1989, 1994) therefore appealed to quantum theory and proposed that CΨ emerges through the coherent collapse of superposition states (a condition in which a quantum system can exist in multiple states simultaneously). This was described by the Schrӧdinger equation (a [partial differential equation](https://en.wikipedia.org/wiki/Partial_differential_equation) that describes 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 a gravitational process causes a quantum collapse that extends across the brain in the form of “quantum coherence”. (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 by the system.) Hameroff suggested that CΨ is generated by quantum coherence in the microtubules of the brain (these are very small and thin tubes that are part of the body’s cell structure). Quantum coherence in microtubules inside brain neurons can explain certain properties of CΨ relating to incomputability, unity, and certain effects of anesthesia.

##  Unsurprisingly, Penrose and Hameroff’s theory has triggered much criticism (e.g., Atmanspacher, 2017, 2024; Blackmore, 2013; van Gulick, 2022). Here I will briefly comment on two critiques. First, Grush and Churchland (1995) argued that the quantum gravitational hypothesis was implausible. Penrose and Hameroff wrote an article in response. In the first part, Penrose complained that Grush and Churchland had obviously not read his books (especially the one published in 1994) in which he discussed at great length the many problems raised in their article (for example, the argument of non-computability in mathematical thinking based on Gӧdel’s incompleteness theorem). Penrose wrote, “But it would seem from what G&C say that they have not even read, and certainly not understood, these arguments.” (p .98). He added 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 answered Grush and Churchland’s biological criticisms. For example, Grush and Churchland argued that CΨ cannot depend on microtubules because CΨ appears even when administration of the drug colchicine arrests microtubule activity (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 counterargued that even administering a large amount of colchicine directly into the brain has very little effect on microtubules.

##  The second comment relates to Derakhshani et al.’s (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 the Diósi-Penrose theory of gravity-related dynamical wavefunction collapse. After analyzing the most reasonable tubulin superposition scenarios, they concluded that this simple version of the theory was highly implausible. It is worth noting here that the authors emphasized that their results refuted a variation of the Orch OR theory and not the theory itself.

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

## First, no TC is free from criticism. It seems that the scientific community has not accepted a single TC as a paradigmatic theory in the Kuhnian sense.

## Second, some theories have received more scientific attention than others in the professional literature. However, this does not make them more successful in 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 the reduction of *explanatory theories*, a topic discussed extensively in the professional literature, and the reduction of *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 in offering explanations in certain research areas. On the one hand, an explanatory model may treat a phenomenon as a special case of a general law; on the other hand, a model may show how a certain mechanism generates the phenomenon under study. It is therefore difficult to see how one type of instruction can be rationally reduced to another. Consequently, the answer to the present question will not focus on the attempt to show how a *mentalistic explanation model* is based on a *mechanistic explanation model*. Rather, the answer will focus on the possibility of reducing a *mentalistic theory* (based on subjective concepts of CΨ related to the individual’s inner world) toa *mechanistic theory* (grounded in objective concepts related to physics, chemistry, physiology, computer processes, etc.). The rationale behind this reductionist approach is largely that the field of psychology is based on the accepted methodology of the natural sciences and thus favors mechanistic explanations over mentalistic ones.

I will first briefly address the problems associated with psychoneural reduction. Then I will discuss other efforts to grasp mentalistic concepts via mechanistic ones, such as substituting mentalistic explanations (goal-based, teleological explanations) with mechanistic ones (causal explanations based on neurophysiological processes). The discussion will clearly show that there is still no acceptable method to reduce or convert a mentalistic 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 & van Gulick, 2024). First, I will briefly explain what procedure is used when attempting to reduce Theory A to Theory B. To avoid getting tangled up 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 deduce TR from TB, along with the relevant bridge laws linking the concepts of both theories. Bridge laws usually identify the concepts of TR with those of TB. For example, in reducing thermodynamics to mechanical statistics, it has been suggested that the concept of temperature is identical tothat of kinetic energy.

Several arguments have been made against the possibility of psychoneurological reduction between a psychological TR and a neurophysiological TB. A famous example is the multiple realizability argument (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 (by motivating an organism to retreat from a harmful situation). Various neurophysiological brain processes in humans, monkeys, dogs, cats, fishes, etc. can realize MSPain. Even a robot could manifest MSPain through behaviors that represent responses to pain stimuli via an artificial nervous system. 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 neurophysiological reduction is grounded in the requirement for ‘unit equivalency’. According to this requirement, the dimensional units on both sides of an equation must balance (see Rakover, 2002, 2018).

The bridge law cannot meet this requirement because the units of measurement for the relevant neurophysiological processes that appear on one side of the equation are completely different from those that appear on the other, 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. Thus, how can the product of a constant be used to meet the unit equivalence requirement?) It is difficult to find a uniform and common scale for these two types of measurements. In this regard, we are not dealing with a bridge law based on 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 (where an action is undertaken to fulfill a specific purpose) into a causal explanation (where 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 translate this purposive explanation into a causal one: 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. First, it may be seen as natural to transform the goal into the cause of 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 behavior? And vice versa: How does behavior trigger mental processes? To reiterate, there is still no satisfactory solution to the mind-body or CΨ-brain problem. Therefore, it appears that the proposed translation does not provide an 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 know how a mental event generates a physical event.

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

In the example above, the cause is Uri’s desire to see Yaffa in Tel Aviv; the result is Uri’s driving from Haifa to Tel Aviv. In other words, what is going on in Uri’s mind explains what he does: The journey represents both Uri’s goal and his intention because intention is always aimed toward a specific goal. Thus, Uri’s journey is not purposeless; it is meaningful, fueled by his will and intention. In the case of the stone and window, the cause and the effect are two separate events, whereas in the case of Uri’s journey, his goal and the action he takes are intertwined and cannot be separated.

In conclusion, the answer to the basic question of this chapter (whether there is a solution to the question of CΨ, or the body-mind problem in classical terms) is negative. This is not only a personal conclusion but also the conviction of several researchers and philosophers. See, for example, the following quotes:

* “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.” (Schrödinger, 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 answered “probably not” 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 in humans (and animals). It can be described as the inner world that only the individual can feel and observe. No accepted TC has yet been found to explain the relationship between mind and body, between CΨ and the neurophysiology of thebrain. Any TC that has attempted 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 are no changes in MSs without changes in NSs. Each of these theories has been rejected by researchers.

In my opinion, the attempt toexplain CΨ based on neurophysiological processes in the brain is problematic. On the one hand, it is tempting to reduce MSs to NSs since behavior as a whole (including conscious behavior) can then be explained by neurophysiological causal theories already known to science. On the other hand, if behavior can be explained entirely by NSs, what is the value of an explanation based on MSs? Why are MSs necessary? These questions run counter to common sense: Everyday experiences suggest that an individual’s behavior is accounted for by one’s inner world, or CΨ. Here is a simple example: I went to the movies *because* I wanted to see the actress Gal Gadot in the movie ‘Wonder Woman’.

To reiterate, if everything is explained by NSs then MSs have no explanatory value. This theoretical approach would inevitably bring us to the outdated and largely rejected philosophical position of epiphenomenalism that proposes that MSs are explanatorily ineffectual. As Kim (2002) wrote (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 affects behavior. Moreover, in light of the observation that no TC has yet been found to explain CΨ based on the neurophysiology of the brain, 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 am not suggesting that because the problem of CΨ has not been solved it is therefore 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) (for similar arguments, see Carroll, 2016).

Second, I am not claiming that CΨ is independent of brain processes in humans or animals. Rather, I am emphasizing that no theory has yet been found to explain the relationship between the two. I propose that CΨ is an explanatory but unexplained concept. That is, it is a primary explanatory concept that negates epiphenomenalism.

To conclude, I suggest that the fundamental qualities of CΨ which make it an anti-epiphenomenalism concept are:

1. Conscious MSs exist, to varying degrees, in every individual human (and in other living beings).
2. Only individuals themselves are consciously aware of the content of the various representations appearing in their own minds (MSs).
3. Without CΨ humans would function purely on a physiological level like plants.
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).
6. Consciousness is dependent on the normal functioning of the brain.
7. Consciousness enables an individual to experience sensory stimuli, the aliveness-feeling (a sense of being alive), life-meanings, and understanding.

The following chapters will focus on further theoretical development.

**References**

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

Schneider & M. Velmans (Eds.), *The Blackwell companion to consciousness* (2nd ed.). West Sussex: John Wiley & Sons.

Atmanspacher, H. (2024). Quantum approaches to consciousness. In E. N.

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

https://plato.stanford.edu/archives/sum2024/entries/qt-consciousness/

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

University Press.

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

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

*companion to consciousness* (2nd ed.). 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* (3rd ed.). London:

Routledge.

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

& M. Velmans (Eds.), *The Blackwell companion to consciousness* (2nd ed.). West Sussex: 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*, (pp. 9–32). Cambridge, MA: The MIT Press.

Cosmelli, D., Lachaux, J.-P., & Thompson, E. (2007). Neurodynamical approaches to consciousness. In P. Zelazo, M. Moscovich, & E. Thompson (Eds.), *The Cambridge Handbook of Consciousness,* (pp. 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.* New York: Viking Press.

Dehaene, S. & 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., & 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.

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

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

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

argument: Why IIT and other causal structure theories cannot explain 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., & 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):[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*, (pp. 9-24). Cambridge, MA: The MITPress.

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

*Consciousness Studies*, *2*, 10–29.

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

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

Hameroff, S. R., & 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.

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

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

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

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

63–81.

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

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

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

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

Schrӧdinger, 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., & Koch. C. (2016). Integrated information theory: From consciousness to its physical substrate. *Nature Reviews Neuroscience,* *17*, 450-61.

Tsuchiya, N., Andrillon, T., & 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., & Mudrik, L. (2022). The ConTraSt database

for analyzing and comparing 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.* <https://plato.stanford.edu/archives/win2022/entries/consciousness/>

van Riel, R., & van Gulick, R. (2024). Scientific reduction. In E. N. Zalta &

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

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

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

Figure 2.1 schematically depicts several areas of the human brain relevant to the topic of discussion in the current chapter