Open Peer Review on Qeios

The Evolution of Consciousness Theories

Ashkan Farhadi[1]

[1] University of California, Irvine

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.

Abstract

Consciousness is usually perceived as a state of being aware of one's environment as well as self. Despite its omnipresence in our life, understanding this concept is challenging. This has given rise to several theories attempting to explain the nature of consciousness, as well as hard and soft problems of consciousness. In fact, the boundaries of consciousness defined by these theories are a topic of continued discussion, particularly in light of the recent advances in artificial intelligence (AI). Some of these theories consider consciousness as a simple integration of information while others purport the need for an agency in the process of integration for an entity to be considered conscious. Some theories consider consciousness as a graded entity and some equate consciousness with content of awareness. In this work, major theories of consciousness are reviewed and compared, focusing on awareness, attention, and sense of self. These findings are interpreted in relation to AI in order to ascertain what makes AI distinct from natural intelligence.

Ashkan Farhadi MD, MS, FACP, FACG[1,2,*]

[1] *Director, Digestive Disease center, MemorialCare Foundation.*

[2] *Associate Professor, University of California, Irvine, Department of Medicine.*

[*]Correspondence: farhadiashkan@gmail.com

Keywords: consciousness, awareness, attention, free will, self-awareness.

Current Perspectives on Consciousness

Even though consciousness has been extensively studied, it remains the most intriguing subject in the field of cognitive science and philosophy, as evident from the myriad of theories attempting to explain this concept. Thus, the most pertinent theories of consciousness are discussed and compared below.

Global Workspace Theory

Global workspace (GW) theory (Baars, 1988) is one of the earliest attempts to explain the juxtaposition of conscious existence and the capacity of human mind to carry abundance of unconscious information. According to this theory, GW is a specialized mental module likened to a stage lit by a spotlight of attention. Accordingly, only integrated information included in the GW will have the opportunity to reach conscious awareness, while the rest will reside in the unconscious mind. In that sense, GW theory shares some key features with the spotlight theory, as discussed later in the manuscript, given that it portrays our mind as a dichotomous entity. Even though it posits that information can transition from the unconscious to the conscious mind, it does not explain how this process occurs and if its outcome is permanent.

Neuronal Global Workspace Theory

Neuronal version of GW theory attempts to overcome these shortcomings by postulating that the information that reaches the GW will result in consciousness only if it can be globally accessible across multiple systems, including long-term memory and motor, evaluational, attentional, and perceptual systems (Dehaene et al., 1998). Thus, it explains why we may process the information differently after awareness. However, as it is founded on the GW theory, it inherits its pitfalls, including the inability to explain why some information is never processed in the neuronal GW.

Consciousness as episodic Memory

An offshoot of GW theory presented by Budson (2022) who argues that consciousness is nothing but an episodic memory and a mechanism for storing memories by binding multisensory details. His theory resembles GW theory, but he claims that he adds a purpose to consciousness, i.e. to store prior experiences in the format of an episodic memory. His theory also follows conscious/unconscious systems proposed by Kahneman (2011).

Many researchers before Budson claimed consciousness is a form of memory (Dafni-Merom & Aray, 2021, Tulving 1985, Moscovitch, 1995). They categorize it into anoetic (non-knowing), noetic (knowing) and autonoetic (self-knowing) consciousness. Some others argued that consciousness is an evolutionary process allowing us to understand the world around us and to act according to the situations (Schacter et al., 2007; Suddendorf & Corballis, 2007). Thus, according to their claim, consciousness is combining episodic memories, using them for understanding the present moment, making meaningful predictions about future events and finally acting properly according to the situations. In this perspective, when we learn from memories, we feel consciousness and able to anticipate, plan and execute an intentional action. On the same note, Cleeremans (2011) added some purpose to remembering process in his “Radical Plasticity Thesis.” He presented consciousness as a learning process by the brain continuous attempts in predicting the consequences of actions of self and environment using memories. In his perspective our mind creates a meta-representation that is interpreted as a conscious experience.

Based on all these theories, consciousness has evolved over years from furnishing episodic memories to problem solving, abstract reasoning, and language by engaging conscious/unconscious systems (Budson 2022).

Integrated Information Theory

Tononi et al.'s (2016) integrated information theory (IIT) is possibly one of the most attractive theories of consciousness, as it is seen as the ideological foundation of panpsychism, given that it assigns consciousness to any entity that processes integrated information. On one hand, IIT is based on the assumption that awareness and consciousness are largely interchangeable, but on the other hand, it purports that consciousness is a subsystem of awareness. Based on this perspective, mind is conscious when any part of it integrates any information, but a conscious mind would allow awareness of select information. In other words, the mind could be conscious but not aware. This property of IIT opposes the conscious/unconscious dichotomy of mind proposed by the GW theory. In this view, mind is an entity that could switch in its entirety between conscious and unconscious state. Besides, the theory does not elaborate whether an integration in a small part of the mind in enough to make the mind conscious, or the integration needs to reach a certain level of dissemination to involve special subsystems in the mind as neuronal GW theory claims. To remedy some of those shortcomings, IIT proposes a level of consciousness for an entity that corresponds to the amount of integrated information. Moreover, similar to the GW theory, IIT remains silent on how the selection of information for processing is accomplished.

Still, given that, according to IIT, the amount of integrated information an entity is capable of processing determines its level of consciousness, natural intelligence (NI) and artificial intelligence (AI) could be considered equally conscious as long as information processing occurs at the same level.

Recurrent Processing Theory

Recurrent processing theory (Lamme, 2006), as an extension of the GW theory, and neuronal GW theory of consciousness in particular, defines consciousness as a result of recurring activity in cerebral sensory areas with highly interconnected feed-forward and feedback connections. In that sense, this theory is the bridge between the GW theory and IIT (Tononi et al., 2016), where the integration of information occurs in a special sensory area of the brain. Nonetheless, this theory does not overcome the shortcomings of GW theory and IIT, as it does not provide an explanation for the process by which information that needs to be processed for awareness is selected.

Higher-order Theories of Consciousness

By introducing the concept of higher-order thought processes, Rosenthal (2002) addressed the dilemma of consciousness by defining it as the cognition of cognition or thinking of thinking process. Consequently, sensation will only turn into perception when it is represented by higher-order theory of consciousness. In other words, only through introjecting oneself as the subject (first-order state) of the sensory experience we will become conscious of that experience. This theory resonates with IIT since it conceives consciousness as integration of information. However, it departs from this theory by contemplating self as a part of conscious experience and recognizing the role of agency/intention in this process. Nonetheless, the previously noted shortcomings of the GW theory and IIT still apply, precluding the understanding of how awareness of select information is achieved.

Attention Schema Theory

Attention schema theory (Webb & Graziano, 2015) is based on the evolutionary information processing, and thus purports that our ability to consciously dedicate our attention to a particular subject has developed as a survival mechanism. Moreover, according to this view, we can manage our attention more efficiently by making a schema of attention, which allows our brain to create a subjective experience of events in the form of awareness. Unfortunately, attention schema theory does not adequately differentiate among attention, awareness, and consciousness, and does not explain how we focus our attention on a particular subject while neglecting others.

Psychoanalytic Theory of Personality

Psychoanalytic theory of personality proposed by Freud (1924) may not be a true theory of consciousness since it presupposes the existence of a hierarchical architecture of human mind. Nevertheless, it is beneficial for the understanding of consciousness, human behavior, and psychology, as it was one of the first attempts to separate the mind into the conscious and the unconscious mind. The metaphor of tip of the iceberg for conscious mind—the part of the mind of which we are aware—is coined after Freud. Based on this dichotomy (Freud, 1915), conscious mind consists of mental functions that are accessible in the form of awareness. However, Freud never explained the process by which the distinction between conscious and unconscious is made, nor he elucidated the role of agency in this designation.

Trilogy Theory of Mind

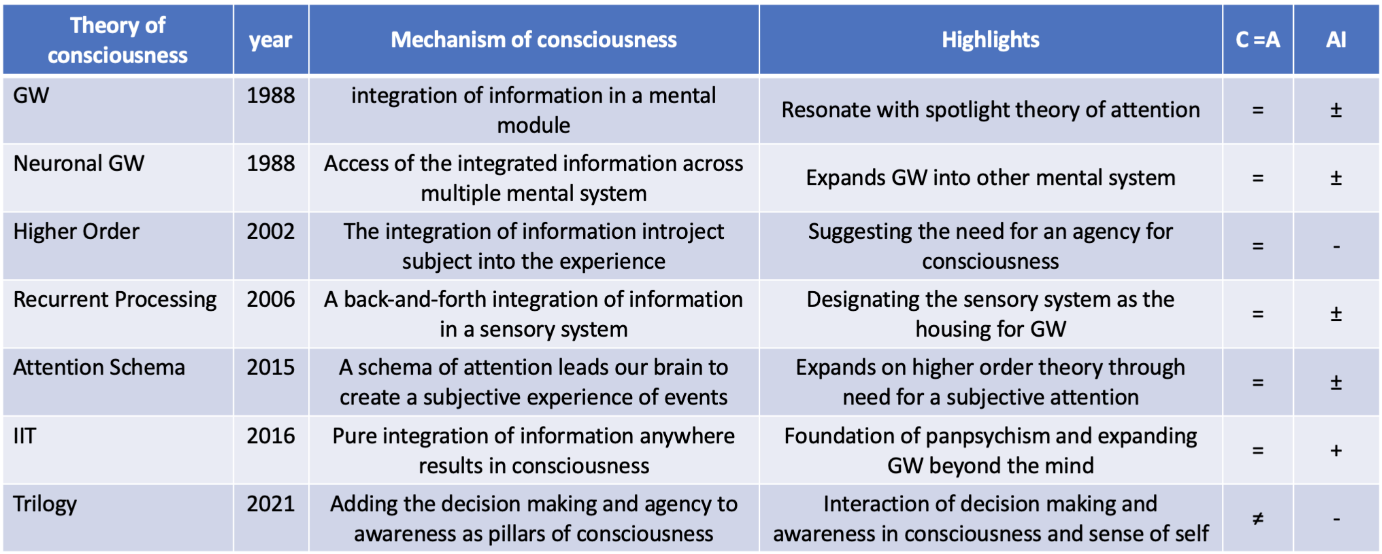
Trilogy theory of mind (Farhadi, 2021, 2023)—also known as “trilogy”—is a relatively recent theory of consciousness that makes a clear distinction between consciousness and awareness, as it purports that awareness is necessary but not sufficient for consciousness, which also necessitates a unique interaction of awareness and decision making. In this model a new mental function is amended to the awareness called awareness-based choice selection (ABCS) that posits that the decision-making process requires awareness as input resulting in emergence of true free will in our decision-making process. According to this perspective, ABCS stands in contrast to making a decision based on an algorithm that is the base of decision-making in AI. Moreover, by amending the process of decision making by newly proposed mental function of discretionary selection of information for awareness (DSIA), intentional attention arises. The intertwined actions of these two mental functions—ABCS and DSIA—comprise a new entity called “I” which is the faculty of our consciousness and separates NI from AI. Consequently, rather than segregating mind into conscious and unconscious domains, this theory considers mind as an unconscious entity that executes all mental functions except ABCS and DSIA. Therefore, as shown in Figure 1, trilogy depicts human beings as a union of “I,” our minds, and our bodies.

In trilogy, both awareness and the decision-making process consist of preselection, selection, and postselection stages. In the preselection stage of decision making, our minds synthesize and analyze a blend of informational inputs as well as emotional intelligence in a process called “reasoning.” This stage is similar to the naturalistic decision model proposed by Drummond (1991) and the decision-making model proposed by Dijksterhuis (2004). However, it departs from these two models by introducing the concept of counter-reasoning that runs parallel to our reasoning process and is an argument that challenges the result of the reasoning process or our most logical choice.

Consequently, counter-reasoning allows us to consider alternatives in the selection stage of the decision-making process, whereby ABCS is applied to the entire matrix of information used for reasoning and counter-reasoning, producing a final choice. However, as due to the function of DSIA not all elements comprising the matrix of information reach our awareness at the same time and consequently our selection may not be the most rational or logical one. This limitation aligns the selection stage of decision making in trilogy with the concept of bounded rationality proposed by Simon (1956) that explains why we may select a choice that is neither most rational nor most closely aligned with our goals and interests as it was purported by the naturalistic or Dijksterhuis model of decision-making. In contrast, as AI relies on an algorithm (SCBA) when making decisions, it produces a choice that closely aligns with the naturalistic or Dijksterhuis model of decision-making—most rational or best aligned with the entity goals.

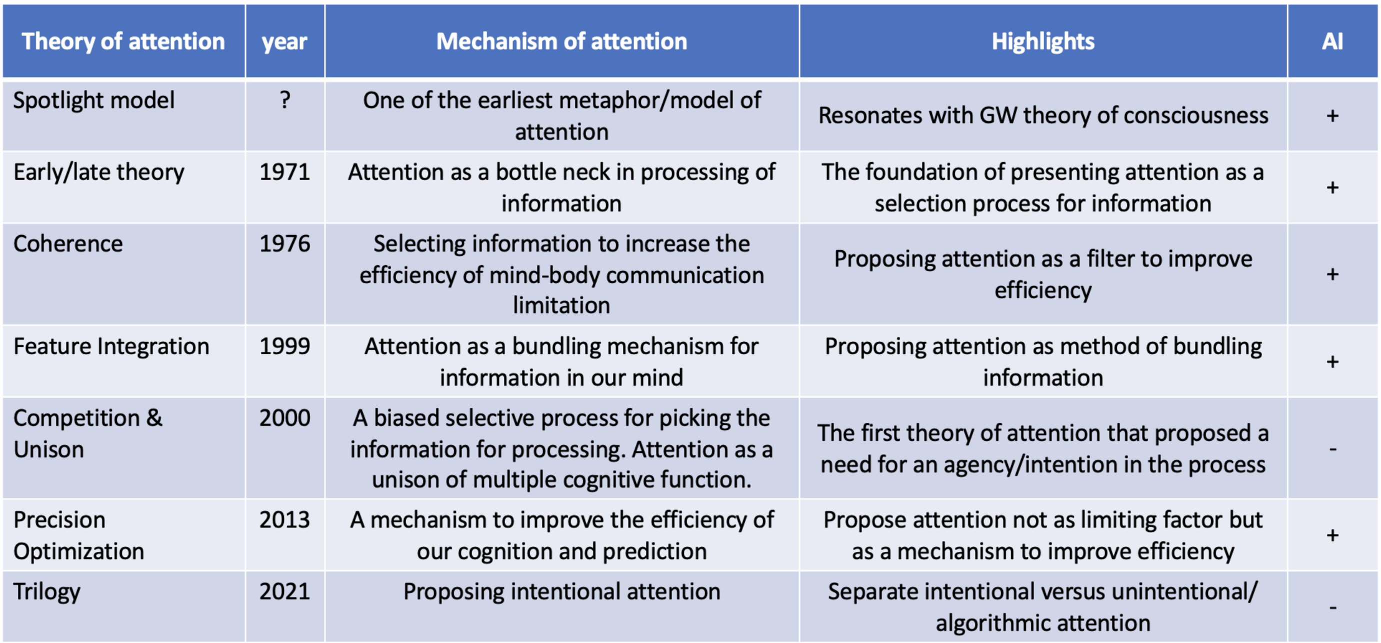
Comparing Different Theories of Consciousness

As can be seen from the brief overview presented above and summarized in Table 1, theories of consciousness have evolved over time to reflect the advances in different scientific domains. As a result, their focus has shifted from integration of information in a specific module of our mind (as is the case in the GW theory) to expansion of the information integration process to other subsystems of brain (as is done in the neuronal GW theory or recurrent processing theory) and eventually to introjecting subject into the conscious experience in the higher-order and attention schema theories. Further advancements can be seen in IIT, where the definition of consciousness has expanded to any entity that is capable of integrating any form of information and finally trilogy introduced intentional attention and added decision-making process to the compound of consciousness. Therefore, it is not surprising that the inclusive designation of consciousness in IIT can be easily expanded to encompass AI while the exclusive designation of consciousness in trilogy reserves this privilege solely for NI. Although other theories of consciousness primarily designed to model consciousness in NI, their definition of consciousness could be adapted to apply to AI as it will be elaborated later in the manuscript. Moreover, in all theories except for trilogy, awareness and consciousness are deemed synonymous, where awareness is a necessary but not sufficient condition for consciousness.



Theory of consciousness	year	Mechanism of consciousness	Highlights	C =A	AI
GW	1988	integration of information in a mental module	Resonate with spotlight theory of attention	=	±
Neuronal GW	1988	Access of the integrated information across multiple mental system	Expands GW into other mental system	=	±
Higher Order	2002	The integration of information introject subject into the experience	Suggesting the need for an agency for consciousness	=	-
Recurrent Processing	2006	A back-and-forth integration of information in a sensory system	Designating the sensory system as the housing for GW	=	±
Attention Schema	2015	A schema of attention leads our brain to create a subjective experience of events	Expands on higher order theory through need for a subjective attention	=	±
IIT	2016	Pure integration of information anywhere results in consciousness	Foundation of panpsychism and expanding GW beyond the mind	=	+
Trilogy	2021	Adding the decision making and agency to awareness as pillars of consciousness	Interaction of decision making and awareness in consciousness and sense of self	≠	-

Table 1. The highlight of the select theories of consciousness in succession showing how theories have been evolved to tease out awareness (A) from consciousness (C) and adding agency as part of the process. Also, the point of view of these theories on AI as a conscious entity.



Theory of attention	year	Mechanism of attention	Highlights	AI
Spotlight model	?	One of the earliest metaphor/model of attention	Resonates with GW theory of consciousness	+
Early/late theory	1971	Attention as a bottle neck in processing of information	The foundation of presenting attention as a selection process for information	+
Coherence	1976	Selecting information to increase the efficiency of mind-body communication limitation	Proposing attention as a filter to improve efficiency	+
Feature Integration	1999	Attention as a bundling mechanism for information in our mind	Proposing attention as method of bundling information	+
Competition & Unison	2000	A biased selective process for picking the information for processing. Attention as a unison of multiple cognitive function.	The first theory of attention that proposed a need for an agency/intention in the process	-
Precision Optimization	2013	A mechanism to improve the efficiency of our cognition and prediction	Propose attention not as limiting factor but as a mechanism to improve efficiency	+
Trilogy	2021	Proposing intentional attention	Separate intentional versus unintentional/ algorithmic attention	-

Table 2. The select theories of attention in succession, showing how theories have been evolved

The other point of distinction is conscious/unconscious dichotomy of mind. Almost all theories except for IIT and trilogy purport that mind has conscious and unconscious parts. While, IIT considers mind as being either conscious or unconscious (or subconscious depending on the level of consciousness) in its entirety, trilogy considers mind as an unconscious entity that requires specific mental functions presented as “I” to provide human beings with the consciousness.

Yet other distinction among theories of consciousness is the degree of consciousness. For example, human could be considered more conscious than a bee, considering the large difference in the amount of integration of information. IIT explicitly claims that there is level in consciousness while trilogy takes the complete opposite stance, claiming that consciousness is an “all or none” phenomenon. There are several other theories of consciousness that were not elaborated in this brief review, and all are proponents of graded consciousness (Jonkisz, Wierzchoń, & Binder, 2017; Doerig, Schurger, & Herzog, 2021) while on the other hand there are other scholars that purport consciousness has no grade or level. Among the latter, some believe that the degree of consciousness is incoherent concept (Bayne, Hohwy, & Owen, 2016; Carruthers, 2019) and some argue that there is no way that we can prove that one NI is more conscious than others (Birch, Schnell, & Clayton, 2020; McKilliam, 2020). There are many interpretations of the level of consciousness. For example, dimensions of consciousness presented by Jonkisz et. Al. (2017), such as phenomenal quality, semantic abstraction, physiological complexity, and functional usefulness. Lee (2022) argues that all theories of consciousness—including the ones elaborated in our review— ought to believe in graded consciousness whether or not, they explicitly acknowledge it unless they present consciousness as a property of soul. The way trilogy posits “I” as the venue for consciousness, enable us to define consciousness without any need for a level or resorting to metaphysical property for mind—soul and purports that the graded consciousness is confusing the the complexity of the content of awareness. Based on trilogy, the state of consciousness is immeasurable and the complexity of the content of awareness has nothing to do with presence or lack thereof of awareness nor consciousness.

Another aspect of the theories of consciousness discussed in the preceding section is the selection of information for awareness. This aspect is the most neglected part of these theories. The selection of information for awareness is either omitted or it is assumed to be performed on an automatic basis. Only Trilogy purports an intentional attention as the requirement of awareness.

Finally, the role of agency is not a topic of discussion in most theories. Agency is implicitly assumed to be needed for consciousness in higher-order and attention schema theories. However, its role is explicitly recognized only in trilogy, where the agency is responsible in the selection of information for awareness.

Reciprocal Role of Consciousness and Sense of Self

One of the main aspects of consciousness is self-consciousness. Its importance was first highlighted by Alan Turing who claimed that a computer could never be the subject of its own thought, as it lacks self-awareness or self-identity. Literally, “I” is defined as any means that we use for referring to self and comprises of our body, mind, soul, or their combination. Prior to Cartesian renaissance, “I” was understood as a metaphysical or religious description of soul or psyche, whereby Berkeley claimed that our spirit is constantly observing us (Downing, 2020). Later, “I” started to be viewed as an entity that is interchangeable with mind, but also as an observer in the Cartesian theatre (Dennett & Kinsbourne, 1992).

Further advancements in explicating “I” were made by John Locke who interpreted self as a continuity of conscious memory that makes us who we are in any moment and over time. David Hume later expanded on this idea by purporting that the sense of self is nothing but a bundle of different perceptions. William James, on the other hand, argued that the sense of self is the core stream of consciousness that carries our innermost thoughts. Most recently, Antonio Damasio proposed the existence of two types of self—the “protoself” and the “autobiographical self”—respectively pertaining to our current self-awareness and our memories (Araujo et al., 2015).

Among current theories of consciousness, the first-order theory and attention schema theory of consciousness discuss the importance of the role of agency in consciousness but do not delve deep into self-consciousness, whereas trilogy ties awareness directly to the sense of self. In particular, this theory approaches self-awareness from three perspectives. From one perspective, self-awareness is a literal translation of awareness of self. This form of self-awareness resembles the autobiographical self as envisaged by Damasio. According to this view, when we tune our attention to our memories and sense of self, we can achieve a form of self-awareness that is called self-image in trilogy. However, a true sense of self-awareness extends beyond intentional awareness of self in the form of self-image. This was first proposed by Avicenna in his “floating man” thought experiment (Goodman, 2013). Avicenna argued that there would be no need for bodily senses for a floating man to have a sense of self. Similarly, Aristotelian form of self-awareness eliminates any need for bodily or mental awareness of self (Cory, 2013). This form of self-awareness that is completely distinct from self-image is called self-consciousness in trilogy, and explains why “I” acts as a gateway for this particular form of self-awareness, which is only possible due to the unique interaction with ABCS and DSIA that give us the sense of agency that is both able and aware. A similar argument was presented by Bermudez and colleagues (1995), according to whom sense of agency is an integral part of self-awareness or self-consciousness.

On the other note, Cartesian cogito “I think, therefore I am” equated thinking to the existence of self. Later, Bertrand Russell (1945) teased out the sense of self from the thinking process and modified the Cartesian cogito to “I think, therefore, there exist thoughts.” In so doing, he argued that a thought presupposes the existence of awareness of that thought, which automatically places self as the subject of the thought (Shoemaker, 1986). This view also resonates with the notion of self-consciousness in trilogy. First, when we consider “I think” we inevitably presuppose: 1- an intention to thinking exists. Without a decision to thinking, there would be no thinking. Thus, we need to make decision to think for thinking process to starts 2- an intentional attention to the thinking process exists. There is a massive stream of mental functions in our mind that would never come to our awareness. For us to be specifically aware of our thinking process, we need to intentionally focus our attention and select this process over other mental functions and thoughts 3- an awareness of our thinking process exists. Now, with these three presuppositions in place, we have a proper setting for consciousness and as a byproduct of consciousness we feel the sense of agency or self and hence, “Therefore, I am” comes through. In this way, trilogy renews the assertion of Cogito with a twist.

The third form of self-awareness presented in trilogy is called mindful awareness and is a type of subjective experience of oneself that could only be experienced in special circumstances such as transcendental meditation. Not everyone has a first-hand experience of this state of mind and in general this form of self-awareness can only be achieved through special training and practice. Nonetheless, the result would be intentional focus of attention on bodily senses without interruption of thoughts. This stands in contrast with the notion that thoughts are essential for having the sense of self. This special form of self-awareness has been previously presented (Lutz et al., 2016; Vago, 2014) and its spectrum spans from attention to self and self-interest to complete selflessness (Hanley et al., 2017). Based on the trilogy postulates, mindful awareness can be understood as the capacity for directing the intentional attention to bodily/environmental sensations without involving the brokering effects of our mind and its thoughts.

It needs to be emphasized that “I” in trilogy is not representative of self and is in fact a selfless mental function that allows all forms of self-awareness to emerge from the interaction of “I” with body and mind.

Attention and its Role in Consciousness

If awareness is the pillar of our consciousness, attention is the keystone. Therefore, it is not surprising that attention is an essential step for improving the information processing by either AI or NI. As all theories of consciousness have touched on this subject in one way or another, a brief review of the theories of attention is presented below in order to draw parallels with the theories of consciousness. One of the first definitions of attention was provided by John Locke who described it as an essential “mode of thought” (Mole, 2009). According to other definitions, attention is the state of mind that is ready for impression—a state that builds anticipation for sensory reception (Mole, 2021).

Early and Late Selection Theories of Attention

One of the first theories of attention defined it as a bottleneck for information processing rather than a state of readiness for reception of information as was previously presented (Broadbent, 1971). According to this perspective, due to bottleneck selection, information may never enter the mind, or can be discarded during processing (Deutsch & Deutsch, 1963; Norman, 1968; Prinz, 2012). However, most authors concur that filtering can be applied at several stages of information processing (Allport, 1993; Johnston & McCann, 2006; O’Connor et al., 2002).

Feature Integration Theory

Feature integration theory describes attention from a completely different perspective. In this theory, attention is a mechanism for bundling information in our mind (Treisman, 1999). Consequently, we become aware of a particular piece of information by binding it with other information. However, critics of this theory argue that the binding process is neither essential nor useful for our awareness, giving rise to a pseudo-problem (Bennett & Hacker, 2003; O’Regan & Noe, 2001). Its further shortcoming stems from the lack of explanation for how and where this binding takes place and why certain piece of information binds and reaches our awareness while other piece of information does not.

Coherence Theory of Attention

In this theory, attention is viewed as an inherent limiting factor in the mind–body interaction (Hirst et al., 1980). As one of the proponents of this theory, Neisser (1976) believed that the vast capabilities of human mind can easily overwhelm the limited behavioral capabilities of the body. Accordingly, attention allows the information needed for mind–body coordination to be selected, and is thus nothing more than a selection process for action (Neumann, 1987) aimed at preventing distraction and maintaining coherence of our agency (Watzl, 2017; Wu, 2011).

Precision Optimization Theories

In this group of theories, rather than serving as a limiting step, attention is an optimization factor that improves the efficiency of our cognition and prediction (Clark, 2013; Hohwy, 2013). This model has been implemented in practice for AI to improve its efficiency, and a similar model has been proposed as the basis of attention through a series of adaptations and predictions to optimize the mental function. The main drawback of this theory is in the sequencing of the processes comprising the so-called attention. Even though this theory presents attention as an optimization process, there is no escape from selection of information prior to optimization since the mind will be easily overwhelmed by optimizing information without a prior selection process. Therefore, a form of selection has to be integrated into the optimization process, which begs the question how we select a certain piece of information for optimization.

Competition and Unison Theories of Attention

This group of theories marks the first attempt to elaborate on the selection process of attention, positing that our mind acts through a top-down biased selection, which presupposes existence of agency (Desimone & Duncan, 1995; Reynolds & Desimone, 2000).

More recently, Mole (2011) raised the issue of relevance of attention to the cognitive function and proposed cognitive unison theory that conceives attention as a unison of many cognitive functions creating a harmonic sync among cognitive processes in the brain. Through offering attention, a metaphysical property, it creates a symphony—unison—from collective members of an orchestra—cognitive processes in the brain. This theory conveniently derelicts its obligation to explain the core function of the attention, i.e., why attention turns into unison on one subject but not the others.

Spotlight Theory of Attention

The spotlight theory of attention is closely connected to the GW theory of consciousness as well as its neuronal counterpart. This theory is more of a metaphor rather than a true theory but has nonetheless gained popularity due to its simple common-sense view of attention. Still, most scholars are of view that it oversimplifies a complicated mental function while placing excessive emphasis on the need for agency (Fernandez-Duque & Johnson, 2002; Henry, 2017).

Trilogy Theory of Mind

Trilogy theory of mind is the only theory of consciousness that models attention as its integral part. Trilogy categorizes attention into intentional attention or DSIA and algorithmic (unintentional) attention or SIBA, both of which are forms of information selection for processing. Based on this perspective, NI employs both SIBA and DSIA for selecting a particular piece of information for awareness (Figure 1), while SIBA is the only venue for selection of information for alertness in AI (Figure 2).

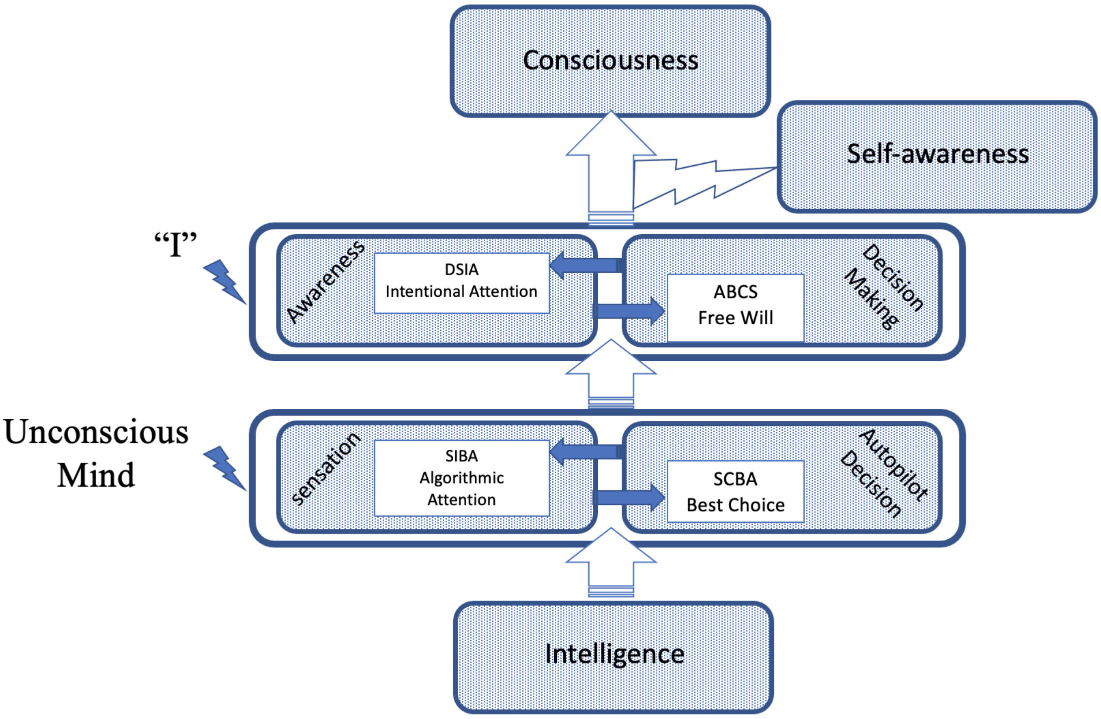


Figure 1. Consciousness and self awareness are the result of two mental function of ABCS and DSIA in NI

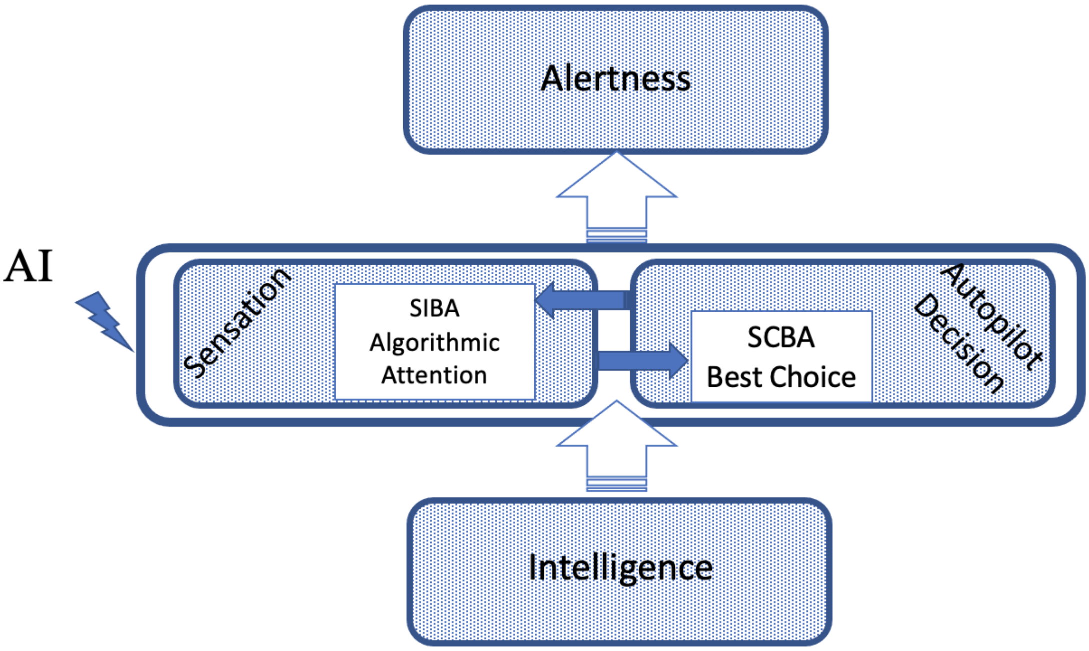


Figure 2. Alertness are the result of two AI function of SIBA and SCBA

Links between Theories of Consciousness and Theories of Attention

As can be seen from the brief overview given above, spotlight theory of attention is closely linked to both GW theory of consciousness and its neuronal counterpart. In that sense, GW can be seen as the stage defined by the spotlight of attention for bringing the information to our awareness. The other association could be traced to the unison theory of attention that resonates with the recurrent processing theory and IIT. In both theories, information processing is the keystone of attention and consciousness, respectively. Another connection emerges between the higher-order theory of consciousness and the competition theory of attention, as both recognize the role of agency. In fact, as in most theories attention is seen as the means (i.e., an algorithm) for increasing the information processing efficiency, these models can be adapted to the functional data processing systems such as AI. Indeed, only trilogy separates the attention into intentional attention (DSIA) and algorithmic attention (SIBA) that is engaged in conscious and unconscious entities, respectively.

Theories of Consciousness and AI

As computer science has made significant advances in recent decades, it is necessary to examine the aforementioned theories and their relevance for AI. This is particularly important given the ongoing debate regarding the AI’s ability to think and be conscious or self-conscious.

There is no doubt that AI is capable of sensing, reasoning, rendering a judgment, and/or making decisions, which may suggest that it is an entity that can think. Therefore, according to the cogito “I think therefore, I am,” AI can be considered a conscious being. In fact, based on Tononi et al.’s (2016) IIT theory of consciousness, since AI uses integrated information, it is a conscious entity, albeit not at the same level of consciousness as a human being. Other theories of consciousness such as GW theory of consciousness or recurrent processing theory may also resonate with IIT and consider consciousness as a property of AI. For example, while GW theory posits that information integration in a special module of mind is a prerequisite for consciousness, this rule can easily apply to AI since its processor fully complies with this requirement. Likewise, the definition of recurrent processing theory can be adopted to AI since data processing can use a particular circuit in a recurrent manner and expand the processing to other subsystems in its processing unit. Given that AI is already more efficient than human beings in performing many tasks, even the boundaries defined by neuronal GW theory for consciousness could be easily expanded to offer consciousness to hybrid AI where a neural network is incorporated as the core neuromorphic architecture on an electronic chip (Wang, 2021). It however remains to be seen if advanced programing of AI can encompass a schema for attention or introject the AI as a subject into the experience and meet the criteria for consciousness proposed by higher-order and attention schema theories.

Trilogy also makes a distinction between NI and AI because it defines NI as a conscious entity due to its faculty of mind known as “I.” Without “I,” mind is an unconscious entity similar to AI. All thinking and decision-making processes in mind or AI are due to SIBA and SCBA, which results in sensations and autopilot decisions, respectively. Only through a unique action of DSIA and ABCS within “I” NI is capable of having awareness and decision making based on free will, and this combination makes NI a conscious being. Consequently, based on this theory, AI lacks consciousness not only because of its limited processing capability, but rather due to the fact that there is no "I" in AI (Farhadi, 2021).

Theories of Consciousness and the Hard Problem of Consciousness

Awareness is the pillar of our consciousness and it gives meaning to our lives as it allows us to transform objective information into subjective experience. As a part of this transformation process, sensation turns into perception (qualia), knowledge turns into knowing, memory turns into remembering, and emotion turns into feeling. What happens in this process, however, remains the hard problem of consciousness, as none of the theories of consciousness reviewed in this manuscript (including trilogy) addresses this question adequately. Still, since trilogy draws a sharp line between awareness and consciousness, according to its postulates, "hard problem of consciousness" originally proposed by Chalmers (1995) should be renamed into "hard problem of awareness."

Limitations of Theories of Consciousness

In sum, the presented theories of consciousness are conceptual models that do not provide calculations or empirical predictions but lay a platform for generating further empirical hypotheses or theories and propose a framework for visualizing the main concepts of consciousness and attention. Moreover, these theories do not provide a detailed neural mechanism for the processes of consciousness, nor do they address the hard problem of consciousness as elaborated above.

Conclusion

Consciousness is considered a state of mind while awareness is described as an experience. Although literature is loaded with subtle differences between these two terms, they are used interchangeably in many scientific and philosophical domains. The review of pertinent theories provided here shows that the line separating these two entities remains poorly defined when it comes to the theories of consciousness. Among these theories, trilogy theory of mind stands out since it considers not only awareness but also the decision making process as pillars of consciousness and at the same time adds agency as an indispensable byproduct of consciousness. As elaborated above, there are drastic differences in the way these theories define and approach consciousness such as selection of information for processing, grading the level of consciousness, and application of these theories to AI. Some of those theories could consider the alertness generated by various sensations due to its algorithmic attention and autopilot decisions in current version of AI as a sign of consciousness while some reserve the designation of consciousness to NI where awareness due to intentional attention and the capacity to make decisions based on free will can result in consciousness and sense of selfhood. In particular, trilogy presents self-consciousness as a byproduct of consciousness when there is a unique mental interaction of awareness and decision making in a faculty of mind called "I." Further studies are thus needed to explore these conceptual models of consciousness and build upon their frameworks to produce new empirical theories of mind.

Conflict of Interest: None

Funding Information: N/A

Acknowledgments: N/A

References

- Allport, A. (1993). Attention and control: Have we been asking the wrong questions? A critical review of twenty-five years. In D. E. Meyer & S. Kornblum (Eds.), *Attention and Performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 183–218). MIT Press.
- Araujo, H. F., Kaplan, J., Damasio, H., & Damasio, A. (2015). Neural correlates of different self domains.*Brain and Behavior*, *5*(12), 1–15. http://doi:10.1002/brb3.409
- Bayne, T., Hohwy, J., & Owen, A. M. (2016). Are There Levels of Consciousness?*Trends in cognitive sciences*, *20*(6), 405–413.
  https://doi.org/10.1016/j.tics.2016.03.009
- Baars, B. J. (1988). *A cognitive theory of consciousness*. Cambridge University Press.
- Bennett, M. R., & Hacker, P. M. S. (2003). *Philosophical Foundations of Neuroscience*. Blackwell.
- Bermudez, J. L., Marcel, A., & Eilan, N. (Eds.). (1995). The Body and the Self. MIT Press.
- Birch, J., Schnell, A. K., & Clayton, N. S. (2020). Dimensions of Animal Consciousness. Trends in cognitive sciences, 24(10), 789–801. https://doi.org/10.1016/j.tics.2020.07.007
- Broadbent, D. E. (1971). Decision and Stress. Academic Press.
- Budson AE, Richman KA, Kensinger EA (2022). Consciousness as a Memory System. Cogn Behav Neurol. 2022 Oct 3. doi: 10.1097/WNN.0000000000000319. PMID: 36178498.
- Carruthers, P. (2019). Human and animal minds: The consciousness questions laid to rest. Oxford University Press. Chicago
- Chalmers, D. (1995). Facing up to the problem of consciousness. Journal of Consciousness Studies, 2(3), 200–219.
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behavioral and Brain Sciences, 36(3), 181–204.
- Cleeremans A. (2011). The radical plasticity thesis: How the brain learns to be conscious. Front Psychol. 2:86. doi:10.3389/fpsyg.2011.00086
- Cory, T. (2013). Aquinas on Human Self-Knowledge. Cambridge University Press. https://doi.org/10.1017/CBO9781107337619
- Dafni-Merom A, Arzy S. (2020). The radiation of autonoetic consciousness in cognitive neuroscience: A functional neuroanatomy perspective. Neuropsychologia. 143:107477. doi:10.1016/j.neuropsychologia.2020.107477

- Dehaene, S., Kerszberg, M., & Changeux, J.-P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. Proceedings of the National Academy of Sciences of the United States of America, 95(24), 14529–14534. https://doi.org/10.1073/pnas.95.24.14529
- Dennett, D. C., & Kinsbourne, M. (1992). Time and the observer: The where and when of consciousness in the brain. Behavioral and Brain Science, 15(2), 183–201. https://doi.org/10.1017/S0140525X00068229
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. Annual Review of Neuroscience, 18, 193–222.
- Deutsch, J. A., & Deutsch, D. (1963) Attention: Some theoretical considerations. Psychological Review, 70, 80–90.
- Dijksterhuis, A. (2004). Think different: The merits of unconscious thought in preference development and decision making. Journal of Personality and Social Psychology, 87, 586–598.
- Doerig, A., Schurger, A., & Herzog, M. H. (2021). Hard criteria for empirical theories of consciousness. Cognitive neuroscience, 12(2), 41–62. https://doi.org/10.1080/17588928.2020.1772214
- Downing, L. (2020). George Berkeley. The Stanford Encyclopedia of Philosophy Spring 2020 Edition. https://plato.stanford.edu/archives/spr2020/entries/berkeley/
- Drummond, H. (1991). Effective Decision Making: A Practical Guide for Management. Kogan Page.
- Farhadi, A. (2021). There is no “I” in “AI”. AI & Society, 36, 1035–1046.https://doi.org/10.1007/s00146-020-01136-2
- Farhadi, A. (2023). Trilogy: A New Paradigm of Consciousness. J Neuropsychiatry, 13. 1-16
- Fernandez-Duque, D., & Johnson, M. L. (2002). Cause and effect theories of attention: The role of conceptual metaphors. Review of General Psychology, 6(2), 153–165.
- Freud, S. (1915). The unconscious. SE, 14, 159–204.
- Freud, S. (1924). A general introduction to psychoanalysis (J. Riviere, Trans). Washington Square Press Inc.
- Goodman, L. E. (2013). Avicenna: Arabic Thought and Culture. Routledge Press.
- Hanley, A. W., Baker, A. K., & Garland, E. L. (2017). Self-interest may not be entirely in the interest of the self: Association between selflessness, dispositional mindfulness and psychological well-being. Personality and Individual Differences, 117, 166–171. https://doi.org/10.1016/j.paid.2017.05.045
- Henry, A. (2017). Agentialism and the Objection from Attention Capture. Paper presented to Canadian Philosophical Association, Ryerson University, Toronto, 29 May 2017.
- Hirst, W., Spelke, E. S., Reaves, C. C., Caharack, G., & Neisser, U. (1980). Dividing attention without alternation or automaticity. Journal of Experimental Psychology: General, 109, 98–117.
- Hohwy, J. (2013). The Predictive Mind. Oxford University Press.
- Johnston, J. C., & McCann, R. S. (2006). On the locus of dual-task interference: Is there a bottleneck at the stimulus classification stage? The Quarterly Journal of Experimental Psychology, 59, 694–719.
- Jonkisz, J., Wierzchoń, M., & Binder, M. (2017). Four-Dimensional Graded Consciousness. Frontiers in psychology, 8, 420. https://doi.org/10.3389/fpsyg.2017.00420
- Kahneman D. (2011). Thinking, Fast and Slow. New York, New York: Farrar, Straus & Giroux.
- Lee, A. Y. (2022). Degrees of Consciousness. Nous, 00, 1– 23. https://doi.org/10.1111/nous.12421
- Lamme V. A. (2006). Towards a true neural stance on consciousness. Trends in Cognitive Sciences, 10(11), 494–501.

https://doi.org/10.1016/j.tics.2006.09.001

- Lutz, J., Brühl, A. B., Scheerer, H., Jäncke, L., & Herwig, U. (2016). Neural correlates of mindful self-awareness in mindfulness meditators and meditation-naïve subjects revisited. Biological Psychology, 119, 21–30. https://doi.org/10.1016/j.biopsycho.2016.06.010
- Mckilliam, A. K. (2020). What is a global state of consciousness? *Philosophy and the Mind Sciences*, 1 (II). https://doi.org/10.33735/phimisci.2020.II.58
- Mole, C. (2009). Attention in later modern thought. In Attention. In The Routledge Encyclopedia of Philosophy. Taylor and Francis. Retrieved 20 Sep. 2022, from https://www.rep.routledge.com/articles/thematic/attention/v-1/sections/attention-in-later-modern-thought. doi:10.4324/9780415249126-V042-1
- Mole, C. (2011). *Attention is Cognitive Unison: An Essay in Philosophical Psychology*. Oxford University Press.
- Mole, C. (2021). Attention. *The Stanford Encyclopedia of Philosophy Winter 2021 Edition*. https://plato.stanford.edu/archives/win2021/entries/attention/
- Neisser, U. (1976). *Cognition and Reality*. Freeman.
- Neumann, O. (1987). Beyond capacity: A functional view of attention. In A. Sanders & H. Heuer (Eds.), *Perspectives on perception and action* (pp. 361–394). Lawrence Erlbaum Associates.
- Norman, D. A. (1968). Toward a theory of memory and attention. *Psychological Review*, *75*(6), 522–536. https://doi.org/10.1037/h0026699
- O'Connor, D. H., Fukui, M. M., Pinsk, M. A., & Kastner, S. (2002). Attention modulates responses in the human lateral geniculate nucleus. *Nature Neuroscience*, *5*, 1203–1209.
- O'Regan, K., & Noe, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, *24*, 939–1031.
- Prinz, J. (2012). *The Conscious Brain: How Attention Engenders Experience*. Oxford University Press.
- Reynolds, J., & Desimone, R. (2000). Competitive mechanisms subserve selective visual attention. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, Language, Brain: Papers from the First Mind Articulation Project Symposium* (pp. 233–247). The MIT Press.
- Rosenthal, D. M. (2002). How many kinds of consciousness? *Consciousness and Cognition*, *11*(4), 653–665. https://doi.org/10.1016/s1053-8100(02)00017-x
- Russell, B. (1945). *A history of western philosophy and its connection with political and social circumstances from the earliest times to the present day*. Simon and Schuster.
- Schacter DL, Addis DR, Buckner RL. (2007). Remembering the past to imagine the future: the prospective brain. *Nat Rev Neurosci. 8*:657-661. doi: 10.1038/nrn2213. PMID: 17700624
- Shoemaker, S. (1986). Introspection and the self. *Midwest Studies in Philosophy*, *10*(1), 101–120.
- Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, 63, 129–138.
- Suddendorf T, Corballis MC. (2007). The evolution of foresight: What is mental time travel, and is it unique to humans? *Behav Brain Sci. 30*:299-351. doi:10.1017/S0140525X07001975
- Tononi, G., Boly, M., Massimini, M., & Koch, C. (2016). Integrated information theory: From consciousness to its physical substrate. *Nature Reviews Neuroscience*, *17*(7), 450–461. https://doi.org/10.1038/nrn.2016.44

- Treisman, A. (1999). Feature binding, attention and object perception. In G. W. Humphries, J. Duncan, & A. Treisman (Eds.), *Attention, Space and Action* (pp. 91–111). Oxford University Press.
- Vago, D. R. (2014). Mapping modalities of self-awareness in mindfulness practice: A potential mechanism for clarifying habits of mind. *Annals of the New York Academy of Sciences*, *1307*, 28–42. https://doi.org/10.1111/nyas.12270
- Wang, G., Ma, S., Wu, Y., Pei, J., Zhao, R., & Shi, L. (2021). End-to-end implementation of various hybrid neural networks on a cross-paradigm neuromorphic chip. *Frontiers in Neuroscience*, *15*, 615279. https://doi.org/10.3389/fnins.2021.615279
- Watzl, W. (2017). *Structuring Mind: The Nature of Attention and How It Shapes Consciousness* Oxford University Press.
- Webb, T. W., & Graziano, M. S. A. (2015). The attention schema theory: A mechanistic account of subjective awareness. *Frontiers in Psychology*, *6*, 500–510. https://doi.org/10.3389/fpsyg.2015.00500
- Wu, W. (2011). Attention as selection for action. In C. Mole, D. Smithies, & W. Wu (Eds.), *Attention: Philosophical and Psychological Essays* (pp. 97–116). Oxford University Press.