A Construct-First Based Approach to a Selective Unification of Theories of Consciousness

A Thesis

Presented to

The Division of Philosophy, Religion, Psychology, and Linguistics

Reed College

In Partial Fulfillment

of the Requirements for the Degree

Bachelor of Arts

Dylan Mabbutt

May 2024

Approved for the Division

(Psychology)

Michael Pitts

Acknowledgments

I would first like to say a massive thanks to my advisor, Michael Pitts, for providing a constant source of support and guidance throughout this process, as well as instilling in me a passion and curiosity in this topic that has led to this thesis.

To my family, I cannot say thank you enough for the love and kindness you have shown me my whole life. Without your support I would have never made it this far and would not have had the wonderful opportunity to attend this school or any other opportunity I have had.

Miller, thank you for being a wonderful person and for showing me an endless amount of love and support. I'm so grateful to have you in my life and I'm so excited to have you in it for a long time to come.

To my roommates, Ivy, Teagan, and Aria, thank you for being lovely roommates and even lovelier friends.

Ivy, I am so thankful for the friendship that you we have shared over our time together here. Words cannot describe how much of an impact you have had on me. I am so glad to have known you the past five years and I look forward to knowing you for many more.

I would finally like to extend a thanks to everyone I have known at Reed over the last five years, friends and strangers alike, for making this time so interesting.

List of Abbreviations

AST	Attention Schema Theory
IIT	Integrated Information Theory
GNWT	Global Neuronal Workspace Theory
UAL	Unlimited Associative Learning

Table of Contents

List of Figures

Abstract

The study of consciousness provides a constant source of intrigue and fascination in those who venture into it, resulting in a huge quantity of theories regarding its true nature. With these theories often claiming to be incompatible with each other, one approach that has been taken to make sense of this phenomenon, known as adversarial collaboration, aims to directly test the specific predictions of these theories against each other within the same experimental design. While this approach is informative and aids in evaluating these specific predictions, it places too much emphasis on the differences between theories, rather than their potential similarities. An alternate approach referred to as the Construct-First Based Approach, proposed by Fazekas et al. (2023), instead suggests that theories should be compared and contrasted by their most central constructs and the lower level constructs that serve them, representing the core argument at the center of each theory. Through this, researchers can examine the overlap between the central constructs of each theory and focus on these spaces in shaping further empricial research. The current study uses a PRISMA literature review in order to conduct an initial attempt at unification of four major theories of consciousness: Global Neuronal Workspace Theory, Attention Schema Theory, Integrated Information Theory, and Unlimited Associative Learning. The main topics on which theories are evaluated are the evolution of consciousness, why some systems are conscious while others or not, and how other processes relate to consciousness. The result of this is a unification that brings together the central constructs of these theories under one theoretical framework that stresses the importance of the biological drive to survive and flexible action and decision making on the shaping and evolution of consciousness and of phenomenological experience of current conscious systems.

Introduction

Scientific theories give researchers the ability to take data found in individual studies and fit it into the bigger picture. Loose empirical data by itself can often be confusing or misleading, and the proper application of theories can weave these threads together and provide a more comprehensive view of the phenomena of interest. Charles Darwin's data on the various features of beaks of different species of finches on the Galapagos Islands, while interesting in its own right, suddenly becomes a lot more meaningful when placed into the greater theory of evolution. A good theory should also allow researchers to make new predictions. It is therefore no wonder why many psychology, neuroscience, and philosophy researchers focus so heavily on providing a complete theory of consciousness that stands up to empirical data. One might expect that, as research goes on, the field narrows down and refines the theories into fewer, more comprehensive theories, but this is not in fact the case in modern day consciousness research. Instead, additional research appears to have instead led to a vast increase in the number of theories. While the true number would be near impossible to calculate, one scoping review, starting with over one thousand research articles published in English or Italian between the years of 2007 and 2017, analyzed 68 articles that referenced theories of consciousness, and found that they represented 29 different theories (Sattin et al., 2021). This number also crucially does not include theories that have been developed since 2017, such as Beast Machine theory (Seth & Tsakiris, 2018). Often, these theories focus on one specific topic and try to explain consciousness through just that lens, leading to trying to explain too much with just one theory. By focusing on multiple theories and attempting to unify their partial overlaps, it becomes possible to address more of the phenomenon through a framework that includes many different topics and aspects in consciousness. In this thesis, I present my attempt at a selective unification of four major theories of consciousness. These four particular theories were selected based on their prominence in the literature and because they each explain a different aspect of consciousness that together can provide a more comprehensive view. Before diving directly into this unification attempt, it will be important to first examine and summarize these four major theories of consciousness.

Theories of Consciousness

Global Neuronal Workspace Theory

Global Neuronal Workspace Theory (GNWT), as the name suggests, is built upon the concept of a “global workspace.” According to GNWT, the global workspace is made up of a large network of neurons with long range axons distributed throughout the brain, connecting a variety of unconscious processors together (Mashour et al, 2020). These processors are responsible for a large number of brain processes, including emotion, cognition, motor movement, action, and sensation. In order for an unconscious process, representation, or piece of information to be consciously experienced, it must be “broadcast” into the workspace, where it can be acted on widely by many different processors (Mashour et al., 2020). Through a process called “ignition,” subsets of neurons can be rapidly suppressed or activated, controlling what is in the current conscious moment. There is however a limit to the process of ignition, as only so many things can be in the workspace at any given moment; one process may be in the workspace while many others will be suppressed (Mashour et al., 2020). Thus, as with most other theories of consciousness, GNWT does not advocate for a sole, all-powerful brain region or structure that controls and determines consciousness, but rather a vast and dynamic network of processors and neurons that are responsible and changing over time. GNWT is a kind of “theater model” of consciousness, as it views attention as a kind of “spotlight” that guides consciousness around the stage. As the spotlight moves to different areas of the stage, lighting up different parts of the world, the current content of the global workspace, and thus consciousness, changes (Baars, 1997).

Attention Schema Theory

Attention Schema Theory (AST) deviates from other major theories of consciousness by saying that consciousness is not an emergent property of processes in the brain but is instead the experienced perception of an internal model of attention that is useful in controlling and directing attention (Graziano & Kastner, 2011). The definition of attention is key here and according to AST, "to attend to a stimulus is to have its representation win a competition, thus gaining greater signal strength, thus being more likely to influence other brain systems such as those involved in decision-making, movement control, and memory" (Webb & Graziano, 2015). Attention can either be top-down or bottom-up, where top-down attention is driven by the current goals and tasks of the individual while bottom-up attention is driven by salient objects in the environment. The brain needs to control top-down attention in order to select information relevant to the current task, and AST says that having an internal model of attention allows the brain to better control attention (Webb & Graziano, 2015). A useful starting point analogy to understand the attention schema is the body schema, which is a model of the body in its current state. By having a model of the body's position and configuration, it is possible to have a complex understanding of the relationship between different body parts and the movements and functions they can perform, easing complex motor function. AST states that this body schema is a simplified representation of the body, allowing for efficient control of the body. A consequence of this simplified model is that it often makes mistakes and errors in the judgment of the body's configuration, leading to errors that can be exploited by various laboratory scenarios (Webb & Graziano, 2015). Moving back to the attention schema, this simplified, internal model of attention, according to AST, can help to control attention and in turn guide various internal and external processes such as action and thought (Webb & Graziano, 2015). Similarly to the body schema, the simplified nature of the attention schema can lead to errors in awareness and attention, which could help to explain failure in tasks such as change and inattentional blindness. Having the model of attention be simplified rather than complex allows for efficient control of information, as it is easier to control less information than more. As a result of having this simplified model of attention, beings with an attention schema have a low-level, immediate, and intuitive experienced perception of a non-physical awareness. They have beliefs and claims about being conscious.

Integrated Information Theory

Integrated Information Theory (IIT) starts off with a series of axioms that the theory assumes to be true, and then from these creates postulates that extend from their respective axioms. The first of these assumptions is that consciousness intrinsically exists (Oizumi et al., 2014). The postulate built off of this axiom says that the "physical substrate of consciousness" (PSC) must also intrinsically exist and have some cause-effect power (Tononi et al., 2016). The second axiom says that consciousness is compositional, meaning that each experience is made up of different combinations of a variety of experiential aspects. The postulate for composition says that the elements that make up the PSC must have cause-effect power within the PSC (Tononi et al., 2016). The third axiom is for information, and it says that consciousness is informative, as each experience has certain qualities that cause it to differ from each other experience, and the reason that each experience is the way that it is because of these differences (Oizumi et al., 2014). The information postulate argues that the PSC must specify a specific form of the set of all cause-effect repertoires specified by all mechanisms in a given system, and that this specific structure must be the reason why it differs from all other possible forms (Tononi et al., 2016). Next is the axiom of integration which says that consciousness is integrated, meaning that each experience is irreducible to smaller components; any given experience cannot simply be reduced to a list of each of its smaller component parts (Oizumi et al., 2014). The integration postulate extends this axiom by saying that IIT states that these experiences are irreducible because each aspect of each experience is causally dependent on all of the other aspects (Tononi, 2004). Finally, consciousness is exclusive, which means that having any one experience excludes any other possible experience, as only one experience can happen at any given time (Oizumi, 2014), and also that consciousness is definite in both content and duration (Tononi et al., 2016). The postulate corresponding to this axiom says that the cause-effect structure of the PSC is also definite; there must be a definite list of cause-effect repertoires looking at a definite set of elements in a definite time and place (Tononi et al., 2016).

Within IIT, the key determinant of whether a system is conscious or not, and how conscious it is, is due to its capacity to integrate information (Tononi, 2004). In IIT information is gained when the outcome of some event rules out other possible outcomes of said event. Integration refers to the dependent, casual interactions present between different elements in a system. In a conscious system, the action of one element depends casually on the actions of other elements, and so changing the action of one of these elements would have repercussions on the entire system. In order to measure consciousness, IIT uses Φ (phi), which is a unit that indicates how much information a given system or subsystem can integrate (Tononi, 2004). Thus, a system is conscious if the amount of information integrated by the maximally integrated subsystem is greater than the information integrated by the sum of its parts. Within a system, there exist complexes, which are subsets of elements that are able to integrate information (have a phi value) and are not found within any subset that has a larger phi value; the complex with the maximum phi value is the main complex, and this instantiates the subjectivity of consciousness (Tononi, 2004). There are major complexes, which are the maximally integrated subset of elements, minor complexes, which are integrating information but less than major complexes, and then other parts that are doing no integration.

Unlimited Associative Learning Theory

Unlimited Associative Learning Theory (UAL) focuses on minimal consciousness, which UAL says includes the characteristics of "unification and differentiation, global accessibility and broadcast, temporal depth, flexible value attribution, attention processes, mapping capacity, goal-directed voluntary behavior, and self–other distinction from a point of view" (Zacks & Jablonka, 2023). UAL aims to identify the characteristics and functions that are jointly sufficient for the evolution of minimal consciousness. From this list of characteristics, unlimited associative learning is proposed as the marker for the complete evolution and development of minimal consciousness, due to the fact this process requires the same list of characteristics as minimal consciousness. Unlimited associative learning is a type of learning that involves discrimination learning, trace conditioning, flexibly changing predictions about patterns, and second-order conditioning (Zacks & Jablonka, 2023). UAL does not claim that unlimited associative learning gives rise to consciousness, but simply that it serves as an evolutionary transition marker that shows that minimal consciousness has been fully evolved (Birch et al., 2020). UAL argues for the existence of a "central association unit", where the actions and processes of sensory processing units, a motor subsystem focused on body mapping, a memory subsystem for event processing, and an evaluation subsystem that allows prioritization of inputs meet and come together (Birch et al., 2020). UAL argues for the existence of a system of "gating", where most inputs are suppressed while a select few others are globally broadcast in the central association unit. As a result of the inhibition and activation of various neural networks, attentional processes become possible (Birch et al., 2020). Attention is therefore a consequence of the UAL, rather than a driving process behind the phenomenon. According to UAL, the ability of unlimited associative learning is beneficial to organisms as it aids in flexible decision making, improving prediction making ability by enabling organisms to evaluate the accuracy of previous predictions. As minimal consciousness requires the same set of characteristics, it developed due to the evolution of unlimited associative learning (Zacks & Jablonka, 2023).

Adversarial Collaboration Approach

For one to take all of the major theories of consciousness and attempt to compare them based on the data that each of them has found in their own separate studies would be incredibly difficult. One approach that has been used in order to propel the search for the complete theory forward is to take an "adversarial-collaboration" approach, in which studies are designed to test leading theories directly against each other. This strategy, which has been conducted within the initiative Accelerating Research on Consciousness (Templeton World Charity Foundation, 2021) involves researchers behind many of the major theories coming together, designing experiments related to consciousness, hypothesizing what they predict to happen based on their respective theories, and then analyzing the results to see which theory the data most supports. This approach allows the theories to be directly compared within the same design, so the theories are evaluated in the same way, which is incredibly useful. One of the consortiums have recently completed a set of experiments within the "adversarial-collaboration" framework, looking at Integrated Information Theory (IIT) and Global Neuronal Workspace Theory (GNWT), which is currently in the process of peer review (Cogitate Consortium, 2023). These methods allow specific predictions made by these theories to be tested against each other, which will provide interesting insights into how their theoretical frameworks map onto empirical data.

While this approach is helpful and informative, a flaw with it is that in studies testing these theories' predictions, results that go against the predictions often lead to proponents of these theories redrafting their predictions about the theory, rather than thinking of how this evidence may instead be indicative of a falsification of the theory itself. Often, these predictions that are tested do not directly represent the central concepts of the theory, and so when they are unsupported or contradicted by data, this is often simply seen as a poor translation from the central concepts to predictions, and so new predictions are made. Beyond this, the adversarial collaboration approach focuses too heavily on the differences between theories of consciousness, rather than their overlaps.

Although it might be helpful to understand and test the contrasting claims of different theories of consciousness, equally (if not more) important are the areas that they agree on. An approach that places equal emphasis on the overlap and nonoverlap of theories of consciousness could potentially reveal much more information about the phenomena of interest.

Construct First Approach

An alternative approach that has been proposed is the "construct-first approach" (Fazekas et al., 2023). This approach argues that when analyzing theories of consciousness, researchers should focus on understanding the "central assumptions" of each of these theories. The phrase "central assumptions" refers to the "the core claim of a theory that defines the theory's target phenomena in terms of a theoretical construct that the theory relies on" (Fazekas et al., 2023). The first step in this approach is to enter the "deconstruction phase," where each theory is broken down in order to separate the higher-level, core constructs from lower-level constructs and auxiliary predictions. Higher-level constructs are differentiated from lower-level ones as they are seen as more abstract and often exaggerate the differences between theories. Lower-level constructs are typically used to define or explain higher-level constructs. Auxiliary predictions, or bridging principles, are predictions used to connect theoretical constructs to empirical data. Using Global Neuronal Workspace Theory as an example, Fazekas et al. argue that the higher-level constructs of this theory are the global workspace and global availability. These constructs are explained by lower-level constructs such as recurrent activity, signal strength, and temporal stability. This deconstruction phase allows for researchers to analyze how different constructs are related to each other both within and across theories. An important claim that this approach makes is that many of these theories rely on the same lower-level constructs, even if they are used to support different theories in different ways. Fazekas et al. (2023) make the argument that the Self-organizing Metarepresentational Account, Global

Neuronal Workspace Theory, and Local Recurrence Theory all rely on the strength and stability of first-order representations to explain the content of consciousness. While these three theories are not all making the same claims about consciousness, they each use the strength and stability of first-order representations to explain these claims. Fazekas et al. argue that a useful way to compare the higher-level constructs of different theories is to create "construct spaces'', where higher-level constructs from different theories are mapped onto graphs with lower-level constructs as the dimensions. Within the article, they provide an example of one such construct space, which maps the global workspace, fragile visual short-term memory, and feedforward sweep - which they claim are constructs of the Global Workspace Theory and Local Recurrence Theory - onto a graph with signal strength, temporal stability, and spread of recurrence as the dimensions, as shown in Figure 1. Creating many of these construct spaces with different constructs as the dimensions can allow researchers to examine which areas of the space the central constructs of these theories appear to cluster around, and design experiments to address how these lower-level constructs relate to different aspects of conscious experience, or if they do at all. It would therefore be possible to examine the core claims of these theories of consciousness while requiring less of a bridge to link higher and lower level constructs, meaning that the approach can go from theory to higher-level constructs to lower level-constructs, rather than from theory to lower-level constructs. While at first glance, it may appear that many theories of consciousness are incompatible with each other, it may be that they rely on similar lower-level constructs in order to make these claims, and they have more in common than it initially seems. It would also be possible to combine together the constructs that are shown to be related to consciousness and to create new theories from these. Whether constructs in the construct spaces overlap or not, new information and insights can be gained about how these constructs give rise to conscious experiences; there is an equal emphasis on overlap and nonoverlap.

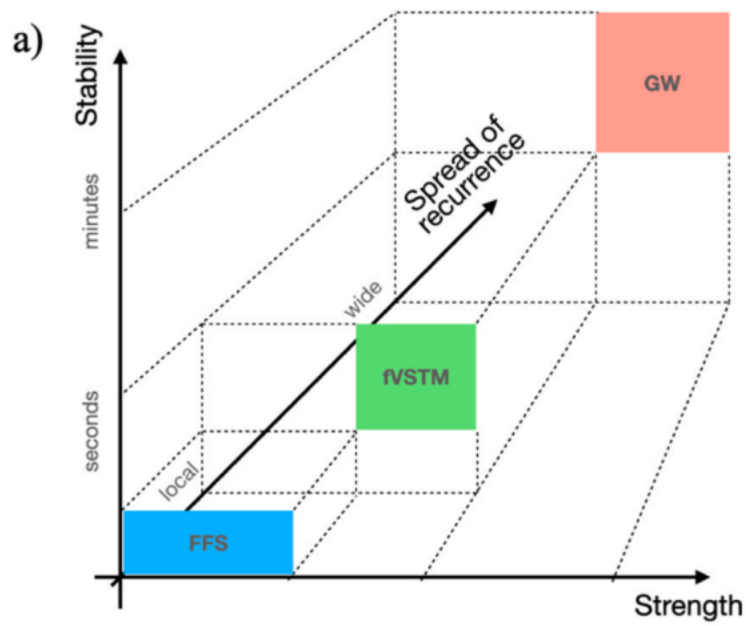

Figure 1. Example of construct space from Fazekas et al. (2023)
A plot of the lower level constructs of feedforward sweep, fragile visual short-term memory, and global workspace plotted on the dimensions of signal strength, signal stability, and spread of recurrence.

As this paper was only recently written and was only recently published, there have been as of yet no attempts at a unification of theories of consciousness using this approach. While some research has used similar methods to the construct-first approach, they often focus on only one or two theories at a time (Farisco & Changeux 2023; Leung & Tsuchiya 2023) or examine only a few aspects of conscious experience (Hommel et al. 2019; Raffone et al. 2014). Within the literature, there is a lack of studies using methods similar to these in order to examine many theories at the same time on the basis of many different aspects of consciousness. Doing so could result in a more comprehensive unification that represents a wealth of different theories and features of conscious experience. It is unlikely that any one theory fully explains all aspects of conscious experience within one framework or that consciousness can be explained by only one phenomena or process; it instead seems to be the case that it draws upon a variety of processes within the brain in order for conscious experience to exist.

In this attempt to selectively unify theories of consciousness using this construct-first-based approach, the current thesis proposes that three main questions must be addressed by the theories. Why are some systems conscious while others are not? What are the mechanisms behind the evolution of consciousness? What other processes in the brain are related to consciousness?

Why are Some Systems Conscious and Some Not?

For the most part, it is generally assumed and accepted within the field of consciousness that human beings are conscious, that it is “like something” to be a human being. It is possible to ask most human adults questions regarding what they are thinking, experiencing, and so on, and receive answers that clue you in to the conscious experiences of these people. This is not possible, however, in the case of nonhuman animals, very young human children, adult humans in unconscious states (here referring to a state in which someone is unresponsive or not awake, rather than the absence of conscious experience), such as sleep or comas, or for a variety of other reasons. Thus, when the discussion shifts towards these other systems and states, there is less of a consensus within the field regarding what is or is not conscious, and why.

An early conception of consciousness and the mind, Cartesian Dualism, the view of French philosopher René Descartes, argues that there are two separate types of substance in the world: material and immaterial. The material substance makes up the physical body and other physical objects within the world, whereas the immaterial substance is what forms the “thinking mind” (Cottingham, 2013). In this view, human consciousness is not explained by physical material, but by mental, thinking material, that is able to exist without the body. We, our thinking selves, are simply contained within physical bodies. While the body and the mind are seen as distinct by Descartes, he acknowledges that there are interactions between the two, and that there is an “intermingling of the mind with the body” (Cottingham, 2013). Descartes states that communication of thought is necessary for something to be conscious, whether verbally or through some other method. He argues that non-human animals are not capable of any such communication that is not simply mimicking human speech, and as a result, non-human animals are not rational or intelligent beings, and have no consciousness (Descartes & Cottingham, 2006). While Descartes is no longer around to comment on more advanced modern machines or artificial intelligences, he held the belief that machines would never be able to communicate sufficiently creatively or behave in ways according to any kind of reason (Descartes & Cottingham, 2006).

Many philosophers and psychologists take issue with dualism, and one counter proposal regarding the nature of the universe is materialism, which says that there is only the physical substance and that mental processes depend on physical processes, which are only made up of physical material. According to this philosophy, whether something is conscious or not depends only upon physical material and physical processes. This view also leaves many questions unanswered, the biggest of which being how these physical processes give rise to consciousness; what is it about our brains that gives us conscious experiences? This is similar, but distinct, from the "hard problem" of consciousness, which focuses on why these physical processes in the brain give rise to conscious experience (Chalmers et al., 1996). This hard problem contrasts to the "easy problems," which instead focus on various functional processes and aspects of consciousness, such as how information is integrated by a cognitive system (Chalmers et al., 1996). These problems are seen as (relatively) easy because just understanding the mechanisms involved in these processes would solve these problems, whereas the hard problem cannot be solved through an understanding of simply the functional aspects of the brain, and instead would require answering the WHY portion of the question, addressing the topics of physical processes in the brain as well as the phenomenology and subjectivity of consciousness. One of the most common methods to address the hard problem is hunting for neural correlates of consciousness (NCCs), which refer to the set of neural mechanisms and events that are "minimally sufficient" for a specific conscious experience (Koch, 2004). Identifying any potential NCCs could include pointing out specific structures, a broader answer of general regions, a specific type of process, a dynamic set of processes happening at different times and places, or an interaction of various processes within the brain that seem to be particularly connected to consciousness, or a broader answer of general regions of the brain could be provided. Discovering these NCCs would resolve one part of the problem, telling us which physical processes give rise to conscious experiences, but the question of why would still remain. Even with the functional explanation, there is still the aspect of phenomenology and subjectivity that has to be explained. All of our conscious experiences are rooted in subjectivity, and so to provide an answer that addressed only functional aspects of consciousness would be unfulfilling and insufficient. The importance of the search for NCCs should not be understated, however, as finding them will move the field closer to solving the hard problem. If information can be gained on the types of patterns of brain activity or brain regions most linked to consciousness, then insights can be made regarding the functions of these regions or activities. By understanding these functions, the WHY question could begin to be addressed.

Another view, which is not necessarily incongruent with materialism, is panpsychism. While many versions of this perspective exist, varying in degree and extremity, the most commonly discussed form says that everything made of physical material has some kind of experience in being (Strawson, 2009). Humans and other animals would all be conscious, but so too are trees, rocks, chairs, etc. Strawson points out that, according to physicalists, our conscious experiences are created by the amalgamation of some combination of physical processes and asks why this cannot too be the case with other systems, even non-living ones (Strawson, 2009). This too leaves many questions largely unanswered. In our own conscious experiences, we do not seem to experience ourselves as beings and then also experience each of our individual atoms as their own beings, so what constitutes the being; is it just the whole that is conscious or are each of the parts also conscious? If it is the case that only the whole is conscious, then what happens when parts of the whole break away? If I am erasing pencil marks with an eraser and tiny flakes of the eraser are breaking off from the pencil, do each of these now constitute their own being as tiny eraser bits? This example is a demonstration of a more extreme, widespread type of panpsychism, and there do exist more limited forms, such as IIT, which allows for a wide variety of different types of systems (but not all systems) to have consciousness.

Some of the most prominent theories of consciousness also disagree when it comes to this question of why some systems are conscious and some are not. According to GNWT, something becomes conscious once it has been broadcast into the global workspace, which makes it accessible to a wide variety of processing systems (Mashour et al., 2020). Something is able to enter the global workspace due to the ability of neurons to "receive bottom-up information from and transmit top-down information to any of the various processors, thus selecting and broadcasting information" (Mashour et al., 2020). The process of "ignition" is key to something entering the global workspace, as it is the process in which certain neurons are activated and some are inhibited, and the content of the global workspace depends upon which neurons are activated (Mashour et al., 2020). Whatever is being broadcast into the global workspace is what is conscious. Whether a system is capable of conscious experience or not therefore depends on its ability to broadcast information. If a system is unable to broadcast information, then it is not a conscious system. What processes exactly are required for a system to be able to broadcast according to this is unclear, as is whether these processes must be organic, whether non-living systems can be conscious or not. IIT advocates for different means through which something becomes conscious. In IIT, consciousness is equivalent and identical to integrated information. Within this framework, information refers to the cause-effect structure of a system that differentiates its current state from all other possible states. Integrated information is the measure of how much the cause-effect structure of a system would change if it were "partitioned (cut or reduced) along its minimum partition (the one that makes the least difference)" (Tononi, 2015), and is also referred to as "phi", or Φ. A system is conscious to the extent that its integrated information is greater than that of the sum of its parts, or the extent to which phi is greater than zero (Tononi, 2015). Thus, the reason why one system is not conscious while another is conscious is that the integrated information of the whole of the unconscious system is no greater than the sum of its parts. A famous criticism of IIT by computer scientist Scott Aaronson outlines how a system that applies a "Vandermonde matrix", the specific mathematics of which are beyond the scope of this thesis, to an input vector, has an amount of integrated information much larger than the sum of its parts; this number is so large that it not only implies that this system is conscious, but that it is far more conscious than any human being (Aaronson, 2014). A similar finding is uncovered by connecting a series of logic gates to each other in an expander graph (Aaronson, 2014). While it may intuitively seem untrue that this system could be conscious, Giulio Tononi, the creator of IIT, said that according to the math, this system would in fact be conscious (Cerullo, 2015).

It is clear that, no matter the philosophical view or theoretical model that one holds, there are many unanswered questions about the nature of consciousness and why some systems are conscious while others are not. A satisfactorily comprehensive theory of consciousness should not only answer this question, but also, within the systems that are conscious, why do some things the system does result in conscious experiences result in these experiences, while others do not? What is the difference between processes that result in conscious experiences and those that do not, and why does this difference exist?

The Evolution of Consciousness

As stated earlier, tasks used to test whether consciousness is present or not tend to depend on the ability of the subject to self-report their experiences. This is not possible with nonhuman animals, and so behavioral observations and measures are used instead. As most of these measures are not universally accepted to be directly testing the presence of consciousness, there is much debate about which nonhuman animals are conscious or not. While Cartesian Dualism argues that humans are a special case and are the only conscious beings (Descartes & Cottingham, 2006), those subscribing to panpsychism may instead argue that all nonhuman animals are conscious (Strawson, 2006). While it is largely accepted now that at least some nonhuman animals are conscious, there is still disagreement about which animals. Plant consciousness is similarly debated (Segundo-Ortin & Calvo, 2022). Whether consciousness is a solely human feature or if it is spread across the animal kingdom, there is pressure on the field of consciousness to focus on the relationship between evolution and consciousness. How has consciousness evolved differently in different species? Is there an adaptive function to consciousness, or has it been selected for because it is closely related to another trait or process? If it has been selected for, what function does it provide that makes it so beneficial? Is the evolution of consciousness a case of convergent evolution, or do all conscious species share a common ancestor with the trait? While it would be ambitious to expect one theory to answer all of these questions, it is important that answers to them are strived for in order to address the evolution of consciousness. The most important aspects for a comprehensive theory to address would be the evolutionary function of consciousness and an account of what mechanisms or processes are required for a species to be conscious.

It might seem easy to accept that animals closely related to humans, such as other primates, are conscious, but other animals that are more distantly related to us might appear to be so foreign and different from us, that it is hard to imagine what their conscious experiences would be like, if they had any. In 2012, a group of researchers convened at the University of Cambridge in order to discuss the idea of nonhuman animal consciousness. They released a document, with the main message being that nonhuman animals have the necessary "neurological substrates that generate consciousness" (Low et al., 2012). Despite these animals being genetically distant from humans, studies have discovered nociceptors and pain response in trout (Sneddon et al., 2003), crustaceans (Barr et al., 2008), fruit flies (Tracey et al., 2003), and many others. While pain response and the presence of pain receptors are not necessarily indicative of consciousness, other findings also seem to point towards the presence of nonhuman animal consciousness. A classic paradigm within the study of animal consciousness is mirror self-awareness, where animals are placed in front of a mirror, and are observed for behaviours indicating self-awareness, where they act on themselves in a way only possible with the reflection from the mirror (Gallup et al., 2002). Mirror self-awareness behaviours have been found in various primates (Gallup et al., 2002), Asian elephants (Plotnik et al., 2006), Orca whales (Delfour & Marten, 2001), and a variety of other species. Many other features, such as memory and theory of mind have also been observed in nonhuman animals (Jones, 2013). While these behaviours are often involved in the discussion of nonhuman animal consciousness, they are not necessarily direct tests of consciousness, and may be testing something else entirely. One criticism of the mirror self-recognition test is that failing the test does not necessarily mean that the subject lacks self-recognition. An alternative design of the experimental paradigm replaces the odorless substance placed on the subject with a subject that has a pleasant smell, giving the subject more motivation to engage with the test. This version is more successful at distinguishing false negatives from true negatives (Heschl & Burkart, 2004). Even with this modification, this does not resolve all of the issues with the test. As has already been stated, consciousness is a complicated phenomenon that cannot be explained just by one process, or identified by the results of one test - at least not yet - and so whether a subject can detect and interact a mark made on their face while looking in their mirror should not be used to determine whether this species has consciousness or not. This goes also for theory of mind, memory, and nociception; while these processes are often associated with human consciousness, they cannot be said to be directly indicative of consciousness. Thus, the question of which specific species of animals are conscious still remains.

Many theories of consciousness focus on human consciousness and largely ignore nonhuman animals. Other theories, however, attempt to specifically provide an account for the relationship between consciousness and evolution. One such theory is Unlimited Associative Learning (UAL) theory. UAL says that consciousness is an evolved trait, and that unlimited associative learning, which is a type of learning that allows for associations between objects, actions, and events, is the marker of this consciousness (Birch et al., 2020). According to UAL, this type of learning is only possible if animals "have a sophisticated enabling system: a central nervous system, a highly innervated body integrated at different levels, multiple feedback relations between sensory categorization programs, exploratory motor programs, and flexible values systems" (Birch et al., 2020). This theory makes specific predictions about the processes and mechanisms necessary for this marker of consciousness and further research is able to analyze the presence of these processes and mechanisms in a variety of species. While other popular theories of consciousness have some things to say about nonhuman animal consciousness, few provide as detailed an account as UAL, and most discussions of nonhuman animal consciousness are more like afterthoughts than proper inclusions.

Although it might seem too far-fetched to speculate about animal consciousness when we still have so many unanswered questions about our own consciousnesses, there are some moral implications that this knowledge could address. If it was found that livestock animals such as cows, pigs, sheep, and chickens were in fact conscious and sentient, how might this impact the livestock industry? To some, it would seem morally dubious to entrap and farm other conscious beings, just as it would be to do so to other humans. As many pharmaceutical trials and experiments use animals as subjects, drug and product testing would also be called into question. Much of human ethics focuses on conscious experiences and feelings, so learning that other animals have similar feelings and experiences could more directly apply these ethical frameworks to the consciousness of nonhuman animals. While the consciousness of these animals would not be identical to our own experiences, knowing just that they are conscious could have enormous ramifications for the way that humans interact with other beings. On the other hand, definitive proof that certain species of animals are conscious could have some practical benefits. While there would still be moral considerations, knowing that certain species of animals do not have consciousness could provide researchers more ethical means to use animals in their experiments. Due to both the moral implications of this topic and the need for a comprehensive theory of consciousness, theories of consciousness should view evolution as another aspect that is necessary to address. One route that this can be done through is by analyzing which processes are most related to consciousness and which appear to be necessary for consciousness.

Consciousness and Other Processes

A largely accepted viewpoint shared between most theories of consciousness is that consciousness is not a standalone process, there is no singular dedicated "consciousness" part of the brain. Instead, theories tend to argue that consciousness relies on and utilizes a variety of different processes. Unfortunately, this common agreement appears to end here, as different theories argue for the importance and role of many different processes related to consciousness. Even if there is agreement that a particular process is important, the manner in which it is necessary is often heavily debated.

One of the processes most commonly associated with consciousness is attention. Conscious perception is often seen as the cause of attention, for example, you hear someone singing, then turn towards the source of the noise to attend to it. In this example, conscious perception is followed by attention. This, however, may not always be the case. This is exemplified well in the case of loud, surprising noises. It is often the case that people turn towards where they believe the source of the noise is before they perceive what the noise is. Here, attention precedes conscious perception. This begs the question of whether consciousness is necessary for attention, attention is necessary for consciousness, or neither. To those who believe that attention is necessary for conscious perception, attention is seen as the limiting factor of perception. They argue that information that is available can go unnoticed because of some limit to our attention, and that the vast amount of information in the world "overflows" our limited systems of perception (Bronfman et al., 2014). This is highlighted in experimental paradigms such as change blindness and inattentional blindness. Due to the limitations seemingly exposed by these paradigms, some claim that human perception is "sparse," even though we believe that it is in fact "rich". One response to this complication is that the visual world is limited by attention, but that it is not "sparse" because the brain averages together groups of information, creating "ensemble statistics," which give average representations of a large percentage of the visual field (Cohen et al. 2016). Further detail is added to the average through saccadic eye movements and by directing attention to other parts of the visual field (Cohen et al. 2016). Another process implicated in this debate is memory, specifically short-term and working memory. Some researchers argue that instead of overflowing the perceptual systems, our limitations in these paradigms are due to failure to encode them into working and short-term memory (Block, 2011). To those who argue for the existence of ensemble statistics, tests of working memory seem to indicate some limitation in perception, not due to a failure in these systems, but because the tests used to look at this process examine specific, unrelated items that cannot be averaged together well by ensemble (Cohen et al., 2016). Some theories of consciousness, such as Attention Schema Theory (AST), specifically focus the study of consciousness on the process of attention. AST says that the brain constructs an internal model of attention, which represents the most important details in the visual field, and this model is perceived as subjective experience (Webb & Graziano, 2015).

There is a commonly held belief that our perceptions of the world are like photographs, directly capturing the exact nature of the "real world". While the exact mechanisms in which it works, research seems to suggest instead that our perceptions are affected by both our expectations (Malik et al., 2023) and past experiences (Kveraga et al., 2007). It therefore appears that our perceptions are not infallible and exact copies of the real world, but instead "perceptual best guesses," as described by Anil Seth (2021), that beliefs can distort. Perceptions are not only susceptible to beliefs and expectations, but also emotional states, as a person's affect can impact the dominance of faces in binocular rivalry experiments (Anderson et al., 2011). The effect of emotion on perception is not just the case in more laboratory test settings such as in binocular rivalry, but also in experiments with settings closer to real life. One study found that fear might be associated with overestimations of height (Stefanucci & Proffitt 2012). Attention is not, however, the only mechanism through which consciousness has been associated with emotion. When we feel any given emotion, we often experience it consciously, we are consciously aware of the emotional state. Using the classic consciousness definition of "what it is like to be something," it seems clear that to be in an emotional state is to be in a state where it is "like something" to be in, differently from when in a different emotional state.

However, research also shows that emotional processing can occur unconsciously (Clausi et al., 2017; Gainotti, 2012). A key question arising from this ability of emotional processing to be conducted either consciously or unconsciously is what the function of conscious processing of emotions is. If emotional processing can occur unconsciously, what function of consciousness provides a benefit to emotional processing that makes the increased demands on resources and energy worth it?

This question is equally important to the role of motor function in consciousness. When moving throughout the world during a regular day, you may be unaware of all the movements you make. Walking down the street does not usually require conscious deliberation of the movements of each leg. When you become aware of these movements, and attempt to consciously recreate them, you might suddenly find it much more difficult or awkward. It appears that consciousness is not necessary for regular body movements all of the time, but it can be used in these movements. It may even be advantageous to take conscious control over these movements in certain situations, for example to avoid absentmindedly walking too far down the street and getting lost. Specific patterns or sequences of movements may initially require consciousness in order to be performed, yet with training and repetition they may be able to be performed unconsciously. This appears to be the case with skilled motor movements, as subjective awareness of action seems to be crucial in the learning of and application of skilled motor movements (Boutin et al., 2014). In other words, it may be the case that conscious awareness of movements is necessary in the process of learning and demonstrating these movements.

These processes listed above are by no means all that are involved with consciousness. It must be remembered that consciousness is related somehow to all subjective experiences, and so every process that is involved in these experiences are in some way associated with consciousness. Many different types and patterns of thoughts, feelings, behaviours, and actions are in some way connected to consciousness, and therefore comprehensive theories of consciousness should attempt to identify the processes most associated with consciousness and investigate the why and how these processes support or are supported by consciousness. As stated earlier, consciousness is not standalone, and must not be treated as some kind of mystical force completely separate from all other physical processes in the brain.

In order to find sources relevant to these major theories of consciousness on these three topics, and to attempt a selective unification using the construct-first approach, a systematic literature review was conducted. The methodology of this search is presented below.

Methodology

Guidelines

In the aim of increasing the transparency and quality of this systematic review, the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines were used.

Eligibility Criteria

In order to be included in this review, publications must be peer reviewed and published, in English, between 1995 and 2024. While more contemporary research would be preferable, much important research was done in the late 90's, including the growth of many major theories of consciousness. As the topic of the evolution of consciousness will include discussion of non-human animals of various species, no exclusion will be made based on the species of subjects. Similarly, a comprehensive theory of consciousness should address humans across the stages of development, publications focusing on children and the elderly will not be excluded. While there may be complications due to developmental changes in both of these populations, the field of consciousness should address differences in the presence and presentation of consciousness across the lifespan. Psychopathological diseases and psychotropic drugs are beyond the scope of this paper and would add complications, so research focusing on these will be excluded from the review.

Information Sources

Searches were run on Scopus and PsycInfo databases. Results were downloaded onto Zotero, where duplicate articles were identified and manually removed. Two different databases were used in order to capture as many relevant publications as possible.

Search Strategy

In order to make sure each topic of the review is covered fairly, separate searches were made for each of the three main sections (why some systems are conscious and some are not, the evolution of consciousness, and consciousness and other processes). Each of the search terms was selected to provide a comprehensive overview of publications within each section. The search terms used did capture some topics irrelevant to the review, but not using these terms resulted in exclusion of many relevant publications, and so irrelevant topics were manually filtered out. The search terms are listed below.

Why Some Systems Conscious and Some Not

consciousness OR awareness

AND

theory OR model OR hypothesis

AND

“neural correlates of consciousness” OR “brain structure” OR “neurophysiology” OR “subjective experience” OR phenomenology OR subjectivity

AND

Creature OR machine OR nonhuman OR human OR animal OR baby OR babies OR sleep OR dream OR dreaming OR coma OR catatonia OR catatonic

The Evolution of Consciousness

consciousness or awareness

AND

evolution OR evolutionary OR adaptivity OR adaptiveness

Consciousness and Other Processes

attention OR memory OR emotion OR volition OR perception

AND

model OR theory OR hypothesis

AND

consciousness OR awareness

AND NOT

evolution OR electrophysiology OR mindfulness OR thermodynamics

After running this search, additional filters in SCOPUS were applied to limit results to papers written in English, those written in the fields of psychology or neuroscience, and those published between 1995 and 2024. A keyword filter was used for each section, limiting results to those that included "consciousness" as a keyword. For the section on evolution another filter was used, limiting results to those that included "evolution" as a keyword.

On PsycInfo, additional filters were added to limit results to results that were peer reviewed, written in English, had references available, and were written between 1995 and 2024. For the search on why some systems are conscious and some not, "consciousness states" and "awareness" were selected as subject major heading. For the section on the evolution of consciousness, major heading filters were added for the terms "awareness," "theory of evolution," and "evolutionary psychology." For the search on consciousness and other processes, subject major heading filters were added for the term "awareness."

The final searches were run on February 14, 2024. The number of results of each search were as follows.

Scopus:

-Why Some Systems Conscious and Some Not - 734 results

-Evolution of Consciousness - 633 results

-Consciousness and Other Processes - 2,337 results

PsycInfo:

-Why Some Systems Conscious and Some Not - 682 results

-Evolution of Consciousness - 254 results

-Consciousness and Other Processes - 988 results

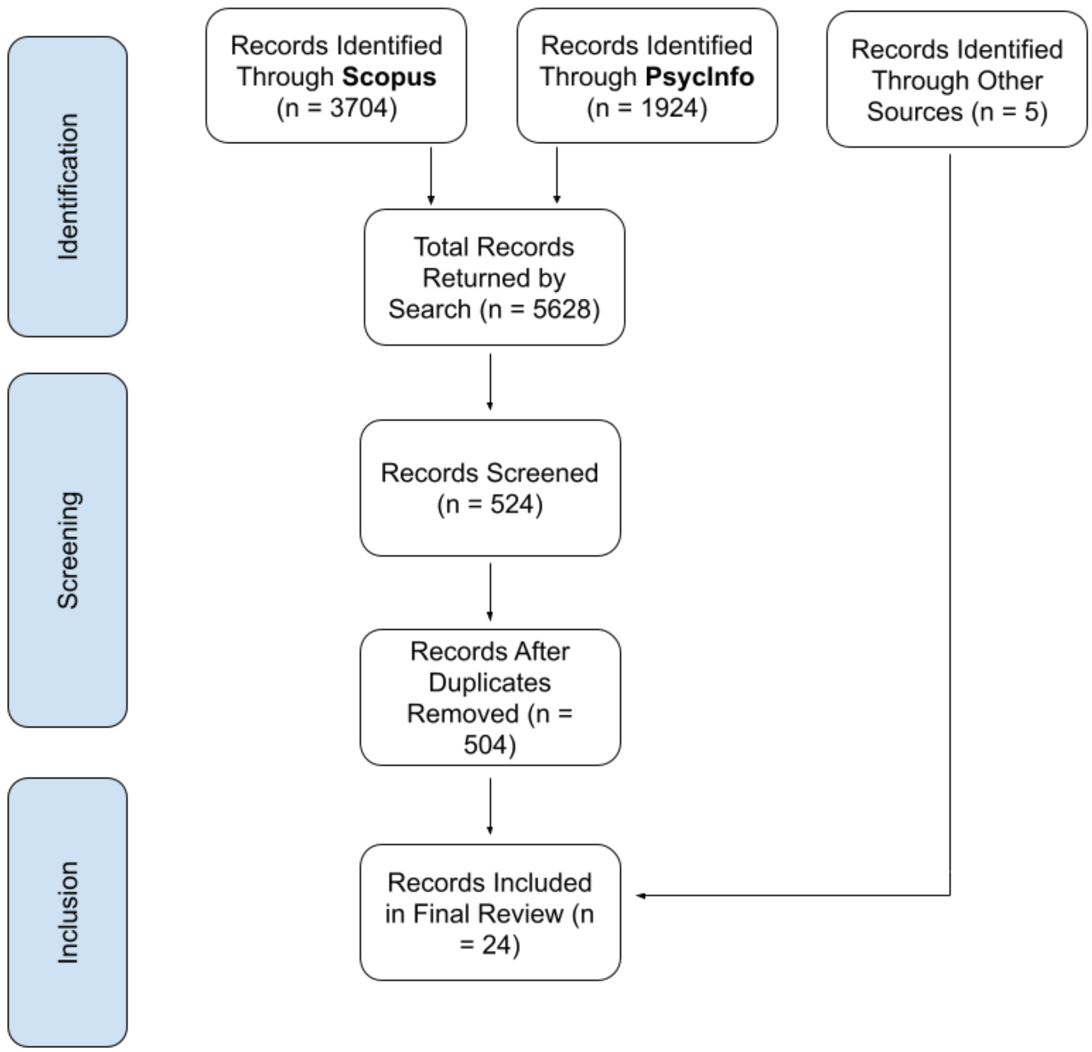

Figure 2. Flow diagram of the systematic review process

Statement of Bias

As the questions asked in this review were largely open ended, the outlining of the specific analysis of each outcome must be somewhat open-ended. In order to avoid biases that may arise from this, I aimed to be transparent in discussing why each outcome was looked at as it was and how selection of publications was made. Efforts were made to consult articles published by major proponents of each theory to make sure that theories were not being misrepresented or misunderstood.

Results and Discussion

The Evolution of Consciousness

If consciousness has an adaptive function that was selected for and resulted in the evolution of the phenomenon, then this can provide insight into the mechanism driving its evolution, the evolutionary root of consciousness, the evolutionary distribution of consciousness, and many other important questions. It could also be the case that consciousness does not have an adaptive function, or at least not one that was selected for, and instead the development of consciousness occurred due to its relation or reliance on some other feature or process that does serve an adaptive function.

The Adaptive Function of Consciousness

According to IIT, the ability of a being to integrate information provides an adaptive function. In this theoretical framework, being able to integrate a large amount of information is equivalent to having a high level of consciousness, and so beings that are very conscious are able to integrate a vast quantity of information. Thus, if the information integrated by a conscious being is relevant to the current task or situation, then this task can be completed at a higher level of performance (Tononi, 2004). IIT additionally proposes that one's ability to integrate information improves based on experience; the more practice a system has at integrating information, the higher capacity they have for integration. As a system gains a higher capacity to integrate information, it provides itself with more resources upon which to lean on and draw from in future tasks and situations demanding consciousness.

GNWT is also clear in stating that consciousness has adaptive benefits that resulted in its evolution and development. The theory does not propose that there is one crucial function of consciousness, but rather a list of purposes that revolve around the ideas of problem solving, creativity, and decision making

(Baars, 1997). In complex creatures, there are endless decisions that must be made, often simultaneously. With this in mind, GNWT states that consciousness is necessary for determining and acting upon priorities; consciousness is used to link long-term goals and priorities with certain behaviours, stimuli, and situations. The global workspace allows for systems to "go offline," escaping the inflexibility of immediate stimulus-driven unconscious sensorimotor programs in order to consciously and flexibly consider options for behaviour and weighing alternatives before acting. With these associations made, unconscious processes important and relevant to these priorities can be quickly activated and broadcast into the workspace in order to direct focus onto the most important goal at the moment, providing resources from a greater number of processors to achieve this goal (Baars, 1997). The global workspace also links processors from around the brain together in a kind of network, so that when an unfamiliar problem or situation arises, a diverse set of processors and banks of information are immediately available as resources on which to make a decision or attempt to solve the problem. Consciousness also provides the ability to analyze errors made by the system in order to shape future behaviour. While unconscious error detection can happen, consciousness and the global workspace allow this error to be examined in greater detail and depth, providing insight into how to avoid similar errors going forward. Once this information is encoded in the memory systems involved in the workspace, it can be recalled in future similar scenarios (Baars, 1997).

In AST, perceptual awareness is an internal model of the process of attention (Webb & Graziano, 2015). By having this internal model of attention, the functional benefit is that attention is able to be better controlled. AST holds that attention must be controlled, due to its complex nature; in the brain, the attention schema is the mechanism by which attention is controlled (Webb & Graziano, 2015). By having a simple model of attention, systems are able to have stable and rapid control over their attentional processes. Instead of having unattached information regarding the environment, the attention schema specifically attaches this information to the content that is being attended to (Graziano, 2014). This is not the only evolutionary advantage of the schema, according to AST, as it also allows for the construction of models of the attention and awareness of others, giving information on the minds of these others (Graziano & Kastner, 2009). By understanding the attention and awareness of others, insights can be gained on their thoughts and emotions, and predictions can be made about how this person might behave in the immediate future. AST says that this same process of social perception can be applied to oneself (Graziano & Kastner, 2009). Human beings, as well as many other species, are social creatures, and so it intuitively makes sense why it would be advantageous to be able to predict the behaviour and feelings of others and of oneself.

At its core, UAL is a framework centered on the evolution of consciousness. As this framework argues that unlimited associative learning predates and serves as the foundation for consciousness, the question of adaptive function focuses more on the function of unlimited associative learning, rather than consciousness. This type of learning gives systems the ability to learn from previous experiences and to use this learning to react with flexibility in future situations (Birch et al., 2020). The adaptive function that resulted in the selection of unlimited associative learning, and consciousness as a result, is therefore the ability of an individual system to adapt and grow constantly over its lifetime. Crucially, unlimited associative learning provides the ability to create associations between stimuli that are entirely new to the organism, meaning that changes to the environment can be addressed more easily by organisms with this functional architecture (Birch et al., 2020). This means that throughout the organism's lifespan, they are able to create large webs of predictive associations, giving a vast network of information regarding stimuli in the environment. As the framework is heavily dependent on biological history, this function provides massive benefits to the biological drive and goal of survival. Organisms with these capabilities would be able to learn from their experiences in the environment and adjust future behavior accordingly.

From these, there are clear parallels between the theories in terms of what they state is the adaptive function of consciousness that led to its evolutionary development. UAL and GNWT both argue that consciousness is crucial to adaptiveness and flexibility in novel situations, using information gained from relevant past experiences. In GNWT, when new information is gained, it can become stored in the memory processors in the workspace, and then can later be accessed easily when it becomes relevant. By being connected to the workspace, these memory processors can send this information to other processors, focused on tasks such as decision making and problem solving. Through having this wide connectedness, the information sharing between different processors can be very efficient, thus allowing quick adaptivity to the situation at hand. UAL similarly says that the abilities provided by unlimited associative learning allow organisms to make predictions about their environment, and rapidly change them when new information is gained. When future predictions are made, previous error detections and changes to predictions can be used as resources in guiding future behaviour. IIT gives a similar outline of the benefit of consciousness, saying that having more information integrated into the system allows the organism to incorporate new information gained through experience and use this to flexibly interact with new information and stimuli (Oizumi et al., 2014). Behaviour is, quite clearly, heavily related to natural selection and the evolution of certain traits. The ways in which an organism behaves in its environment are determinants of the survival of the species. Traits that improve survival will result in their selection, while traits that decrease survival chances will likely be selected against and reduced in the gene pool. The case that each theory makes for how consciousness, or the mechanisms that give rise to consciousness, affects behaviour are therefore crucial to understanding the evolution of consciousness as a trait. By having the capability of quick and flexible adaptive responses to novel situations that each of these three theories advocate for, then organisms are more able to behave in ways that will support their survival and the survival of their genes.

Out of these overlaps, three major aspects of consciousness focused on by these four theories can be seen: the controllability of attention, perceptual integration, and the use of learning and memory in making associations. From these aspects, the beginnings of a unification can be revealed. The position of the central constructs of each of the four discussed theories can be seen in Figure 3.

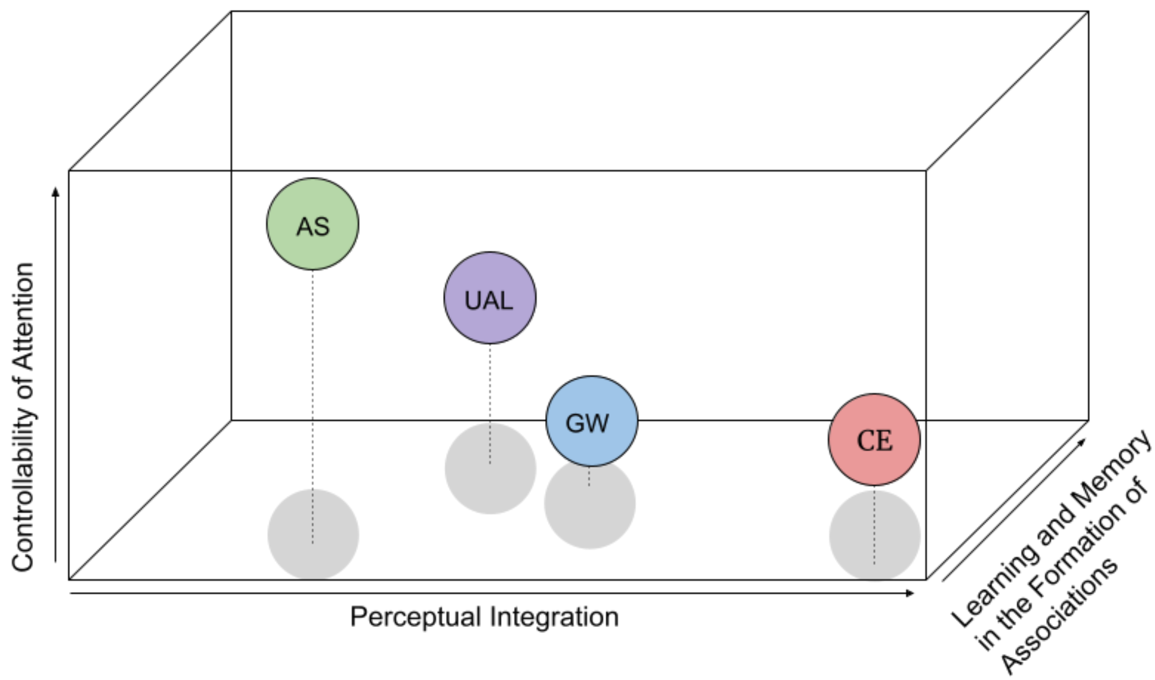

Figure 3. A space of the central constructs of AST, UAL, GNWT, and IIT
Central constructs are mapped on the dimensions of controllability of attention, perceptual integration, and learning and memory used in the formation of associations. AS refers to the attention schema, UAL to unlimited associative learning, GW to the global workspace, and CE to the maximally irreducible cause-effect structure. Each axis ranges from minimal to maximal in terms of how each construct explains the variable of interest in relation to consciousness.

As seen in the figure, the maximally irreducible cause-effect structure specified by IIT provides a maximal explanation for the role of perceptual integration in consciousness and the mechanisms required for it, the attention schema provides a maximal explanation for the role and method of controllability of the process of attention, and UAL displays a maximal explanation for the purpose and mechanism through which learning and memory are used in order to for associations. The global workspace provides a mid-range explanation on all of these dimensions, providing a maximal explanation for none of them. In a "complete" theory of consciousness, answers would be provided regarding the role of all of these dimensions on consciousness (i.e., in the far upper right region of the space depicted in Figure 3). As a result of this, I believe that the combination of these four central constructs under one framework would be able to provide a more thorough explanation. While the global workspace does not provide a strong explanation for any of the dimensions, it specifies a workspace that is relied on for the architecture and functioning of the unification model as a whole. In these theories, attention is crucial to determining conscious content and selecting what is in the current conscious experience. In this unification then, the attention schema serves as the first step in conscious experience. In order for consciousness to have a focal point and for there to be an object of consciousness, there must be a mechanism that controls attention, and leads to attending to the right things, those things that are important to the survival of the system. As AST says, the brain requires control of top-down attention to select information that is relevant to the current task and having an internal model of attention, the attention schema, allows for better control of the process of top-down attention (Webb & Graziano, 2015). Once attention has been controlled and is focused on a specific stimuli or piece of information, this information can become integrated into the system. The control of attention allows for certain outcomes of events to be ruled out and for information to be gained. In this unification, the conscious system has a series of major and minor complexes that are integrating information attended to and selected by the attention schema. The major complex that is maximally integrated and that has a phi value greater than the sum of its parts determines the information that is broadcast into the global workspace. The purpose of phi here is to signal to the workspace that the information integrated by this complex is significant and complex enough to warrant further resources dedicated to it, which are provided by being broadcast into the global workspace. Once this information is in the workspace, then processors from around the brain gain access to it and are able to process and act on it. The processors in the workspace are diverse, dedicated to a variety of different tasks within the brain. Contained within the workspace are systems responsible for memory and learning. In this unification, these systems are crucial, as they relate to the proposed major adaptive function of consciousness, which is to better be able to learn from the environment, and form complex associations based on stimuli within the environment. Once information has been integrated and processed by processors in the workspace, it can be stored in memory systems and accessed later when the information becomes relevant, such as when similar stimuli are encountered in the environment again. This system allows for slow, conscious deliberation of information when it is first encountered and processed, allowing for careful planning and flexible decision making to occur to determine what should be done about this novel stimulus. In order to behave flexibly, there is an evolutionary pressure to not react reflexively. Once this information has been integrated and stored in the memory systems, later access to this information is fast, allowing for rapid action in the face of this or similar stimuli. In order to create good associations that aid in survival, a mechanism is necessary to attend to the “right things,” which circles back to the attention schema. Attention selection is necessary for an organism to “decouple” from the environment and shift to a more flexible mode that allows these complex associations to be formed.

This unification does not claim that complex and vast perceptual integration is necessary in order to form associations of stimuli in the environment but that the ability to integrate a greater quantity of information allows for a greater number of more complex associations to be made, leading to advantages in survival. Other species may have lower capabilities of perceptual integration but are still able to form associations, these are just simpler and less flexible than those possible by a system with the mechanisms described above. In this unification, consciousness has an adaptive function that it provides to the system. As AST believes consciousness does not itself provide an adaptive function and is instead just the experience of the internal model, this belief of AST is inconsistent with the unification and not included in it. Consciousness in this model can be seen as an emergent property from the mechanisms included. The in-the-moment conscious experience of attention can be seen as a product of the attention schema and internal conscious thought and beliefs can be seen as a product of the processes of integration and global broadcast within the brain.

The Evolutionary Age of Consciousness

The question of when this system may have first evolved is rather complex but important to address. The date of the evolution of consciousness has important implications on how widespread it may be and also provides insight as to how it may have developed over time since its first appearance. If consciousness was first evolved a long time ago (in the evolutionary sense), then it would also be expected that this would be widespread among current living animal species. Adding onto this it is also important to note whether the presence of consciousness in different animal species is due to convergent evolution or a shared ancestor, which can be examined through these questions.

For the first of these questions, AST argues that the internal model of attention first appeared around 550 to 350 million years ago and the use of the attention schema for social attribution of awareness around 65 million years ago with the evolution of primates (Graziano, 2014). AST also predicts that selective signal enhancement, a bottom-up attentional process that allows systems to promote signals that are most relevant in the current moment, evolved roughly 550 to 500 million years ago (Graziano, 2014). These numbers, however, are merely informed speculations based on the data available on traits indicative of the attention schema in current species and knowledge of the phylogenetic history of these traits. AST argues that top-down control of attention has been observed in various current mammal and bird species. As a result of this, the evolution of the attention schema is dated at least 350 million years ago, as this was the time of the last common ancestor between mammals and birds. The 65 million years ago date for the development of social attribution of awareness is similarly selected based on constraints on available data, as at the time of this articles' publication, the only evidence available as to the presence of this trait was in primate brains (Graziano, 2014). These dates are therefore somewhat imprecise and open to change based on new evidence, but the case made for a common ancestor evolution of the attention schema is significant.

UAL dates the evolution of unlimited associative learning at approximately 540 million years ago, during the Cambrian period (Birch et al., 2020). This period is characterized by the "Cambrian explosion," where there was a sudden and massive diversification of species, resulting in more complex animal life that started to shift to appear more similarly to modern animal species. They argue that the evolution of associative learning in in fact caused this "explosion," as the ability to learn many predictive associations within the lifetime of one organism allowed for quick learning and discovery of new possible niches within the ecosystem, which overtime resulted in the divergence of species and diversification of animal life (Birch et al., 2020).

As argued above, the model of consciousness described by the unification depends first upon the attention schema, as this is necessary for better controlling of attention, which allows for the right information to be attended to and selected for integration, which in turn leads to broadcasting of the information with the maximum phi value, which then leads to the formation of memory and associations based on this information. It would therefore be expected that the attention schema would have to have evolved first, in order for these other processes to be built onto the schema. However, based on the dates provided by AST and UAL for the first evolution of their respective processes, it would appear that the ability of unlimited associative learning evolved earlier than did the attention schema. There are two possible resolutions to this conundrum. First, and simplest, is that the dates provided are incorrect. The estimation of 500 to 350 million years ago as the first appearance of the attention schema is provided just by selecting the date of the last common ancestor of birds and mammals, two groups that have been shown to display highly developed traits of the attention schema (Graziano, 2014). It could be the case that other animal classes display traits of the attention schema, but they have not been studied or observed as of yet. This would mean that the date for the common ancestor shared between the classes displaying the attention schema would shift, and this date would move backwards. On the other hand, it could be a case of convergent evolution, where mammals and birds evolved the attention schema separately, and then the dates of these separate evolutions would have to be examined. This first proposal thus suggests that the data provided by the theories could be incorrect and that it could be the case that in fact the attention schema developed prior to the evolution of unlimited associative learning.

The second explanation, that is more favored by this unification, is that the attention schema did not necessarily have to appear first evolutionarily in order to fit into this model. While it is described as the first step in the process of consciousness described in the model, this is in terms of the in-the-moment temporal relativity, and not in terms of evolutionary history. The model suggests that controllability of attention, perceptual integration, and learning and memory in the formation of associations are necessary for the version of consciousness described and experienced by humans, and this does not necessarily have to be the case for minimal consciousness. As described above, the unification model suggests that it is possible for a certain species to have more limited capabilities of perceptual integration and attention control but still form associations, just that these associations will be simpler due to these limitations. As such, if unlimited associative learning did indeed evolve before the attention schema, then this does not invalidate the model, it instead just suggests that the system was much more limited prior to the development of the attention schema. In order for the system to be at full functionality, it requires an attention schema for control of attention, the ability to integrate information and broadcast it into the workspace, and unlimited associations in order to form associations regarding stimuli in the environment and to flexibly control action.

Why are Some Systems Conscious and Some Not?

With this question of why some systems are conscious and some are not, the main aspect to be addressed is what, according to each theory, determines whether a system is conscious or not. A major part of this will be a discussion of the mechanisms involved in consciousness and how these mechanisms may give rise to conscious experiences. Brain structure is also crucial here, so the four selected theories will be examined to determine which structures of the brain each theory correlates with consciousness. The focus is not solely on the physical, neurophysiological correlates of consciousness, so the question of how activity among these structures leads to conscious experience is also crucial. Also of interest in this section is the topic of the distribution of consciousness, looking at machine and nonhuman animal or creature consciousness.

The Structure and Activity of Neurophysiological Correlates that Lead to Conscious Experiences.

In GNWT, a given conscious system is conscious because of the presence of a "global workspace," because it has information processors located around the brain that allow for information to be broadcast between and to each other, supported by the presence of a network of pyramidal neurons with long-range axons distributed around the brain (Mashour et al., 2020). This is the key idea of GNWT regarding what makes a system conscious; without the workspace, according to GNWT, a given system cannot be conscious. The process of ignition, where neurons that encode the stimuli currently relevant while are rapidly activated while the activity of all other neurons in the workspace are suppressed, is crucial to the functioning of the workspace, as it controls what is currently the focal point of activity in the workspace (Mashour et al., 2020). As GNWT advocates for consciousness being widespread around the brain, there are many different brain regions implicated in consciousness by this theory. These implicated regions and structures include the prefrontal cortex, medial parietal and inferior lateral parietal cortex, anterior temporal cortex, cingulate cortex, precuneus, thalamocortical core, as well as others (Baars et al., 2003; Mashour et al., 2020). Key here is the idea that not all of these structures or regions are involved in the workspace at any given time, as the workspace is seen as dynamic, and the processors that make it up are constantly changing as demands necessitate. Thus, the importance of neurophysiological structures on the workspace is not the specific structures themselves, but the patterns of activity and connectivity that together combine to form the global workspace.

IIT makes the claim that consciousness is identical to integrated information, and therefore what displays whether a system is capable of consciousness or not is its capacity for integrated information; a system with no capacity for integrated information is not conscious, a system with a large capacity for integrated information is very conscious. (Oizumi et al., 2014). Information is gained when alternative outcomes of a given event are ruled out based on the actual outcome of the event. Information is integrated when the system receiving the information is causally dependent on itself, as in the functioning of each individual element relies on the functioning of all of the other individual elements (Tononi, 2004). Whether a system is conscious or not is therefore dependent on whether it has a cause-effect structure in place that cannot simply be reduced to the sum of its parts. As with GNWT, in IIT the structures that make up the major complex, the complex with the largest phi value and the maximally integrated subset of elements, are somewhat dynamic, and dependent on the type of information that is being processed; the only elements contributing to the quality and content of conscious experience are those that are exchanging and integrating information within the major complex (Tononi, 2004). The specific conscious experience in any given moment is determined by the activity of the elements currently present within the major complex (Tononi, 2004). The theory does however argue that the thalamocortical network is crucial for the generation of consciousness, which includes a diverse variety of functionally specialized processors throughout the cerebral cortex (Tononi, 2004). The nature of the current integrated information determines which of these specialized processors make up the major complex, and damage to these specialized processors result in deficiencies in the conscious experience of the elements and contents that the activity of these specialized processors code for.

In UAL, unlimited associative learning is seen as the transition marker of consciousness, as it requires the same set of capabilities as consciousness, including global accessibility and broadcast, unification and differentiation, selective attention and exclusion, intentionality, integration of information over time, an evaluative system, agency and embodiment, and the "registration of a self/other distinction" (Birch et al., 2020). A system is conscious not because it is able to form unlimited association, but because it has the systems and mechanisms in place that give it the capabilities needed for both consciousness and unlimited associative learning. Without any of these features, a given system would be capable of neither consciousness nor unlimited associative learning. This model is open and supportive of consciousness and unlimited associative learning being widespread among the animal kingdom and due to the diversity in brain structure among the animal kingdom, it does not claim a specific set of brain structures are required, describing the functional architecture necessary for consciousness and unlimited associative learning, rather than the structural architecture (Birch et al., 2020; Zacks & Jablonka, 2023).

As stated, AST argues that the beliefs and claims that people have regarding a non-physical consciousness are not emergent properties of processes in the brain, but are instead the experience of the model of attention constructed in the brain; the difference between a system that has these beliefs and claims and one that does not is the presence of the attention schema (Webb and Graziano, 2015). While the theory claims that there is a physical basis of the attention schema and of attention itself, it does not provide a direct explanation of what this physical basis, what the structures and activities responsible for the creation of the attention schema are. However, the theory predicts that the inferior parietal lobule and the temporoparietal junction are likely involved (Igelström & Graziano, 2017). This makes the question of why some systems are conscious - or why some believe they have consciousness - while others do not difficult, as there is no account to be made for differences in the structural mechanisms necessary.

Largely, all four of these theories focus more on function than structure. GNWT and IIT both suggest that the structure determining the current conscious moment is constantly changing and that the full list of structures involved in consciousness are distributed widely throughout the brain, involving a diverse set of processors. In both of these theories, the current conscious content and experience is determined by the activity of the processors that happen to be either acting in the workspace or the major complex in the moment. The unification presented in this thesis holds that consciousness is not a solely human phenomena, and so the architecture provided by it must account for non-human systems. Thus, it focuses more on the functional architecture presented by these theories, rather than the specific brain structures and regions that they may argue result in consciousness. This does not, however, mean that structure will be ignored, as enough research and data exist on the human brain in order to make judgments and claims about it.

The function of consciousness according to this unification model is to provide the brain with information relevant to the systems survival so that this information can then be integrated into the system and acted on by a wide variety of processors, overall allowing for the formation of complex associations between stimuli present in the environment. In a similar fashion to GNWT, this provides systems the ability to "go offline," with many processors functioning unconsciously while a specific set are dedicated to consciously acting on the currently broadcast information, giving the system more flexibility with action and decision making. In order to support this function, a variety of functions have to be supported by structures in the system. For the attention schema, the system must have dedicated structures towards creating the internal model of attention. As AST is quiet on what exactly is necessary for an internal model of attention to be created, it is difficult to say what the system would need to create this, even in functional terms. What is known to be necessary, however, is that the system must have cognitive access to the model. In AST, access to the cognitive model is what provides the ability to make claims about conscious experiences (Webb & Graziano, 2015). The internal model cannot be constructed just by background processors, then, as it must be accessible to the system of which the model is constructed. For the integration of information, a system must have some capacity for integrated information. The system must have a network of elements that depend on each other. If it were made up of elements that were disconnected and ran individually with no dependence or cause-effect relationship with other elements, then information within the system cannot be so easily transferred between different processors around the system, diminishing the complex action of the system as a whole. The cause-effect and dependent relationship between elements allows for the activity of the system as a whole to go beyond what the sum of each individual elements would be able to do. IIT implicates the whole thalamocortical network in the generation of consciousness, and this network is equally important in this unification. In order for maximum flexibility to be achieved, the system needs a diverse set of processors with different specialized functions, so that information can be acted on in a variety of ways. The thalamocortical network is integral to this as the thalamus provides a connection to a massive number of cortical processors and structures. Research suggests that activity in the thalamocortical network can have a "temporal binding" function that allows for the binding of information from different sensory modalities, causing them to be experienced as one percept (Ribary, 2005), so it would not be unreasonable to expect this ability to apply to consciousness as well. GNWT also implicates activity in the thalamocortical network with consciousness, so this is equally important for the workspace aspect of the unification. As stated, an exact account of each structure necessary for consciousness is not able to be provided, as the unification of these four theories argues that consciousness is widespread across a variety of animal species. The functional account of this aspect, therefore, would be that a conscious system requires some kind of binding activity that allows the various processors around the network to be linked together, allowing for integration and unification of information.

UAL clearly lists the capabilities necessary for a system to have consciousness and unlimited associative learning, and some of these are addressed by the other frameworks included in the unification. Global accessibility and broadcast are possible due to the workspace and the network of processors and neurons that make it up, allowing for activation and the broadcasting of certain signals. Because the processors included in the workspace are linked together by nature of being in the workspace, this provides global accessibility of information being acted on. Selective attention is addressed by the attention schema; this capability requires that a system be able to place importance and salience of one stimulus over others, which is allowed for by the control of attention provided by the attention schema. By controlling the process of information, a system is able to focus heavily on stimuli that are deemed to be the most important within the environment. By directing attention to this particular stimulus, the attention schema provides the rest of the system with the information that should be focused on in the current moment. Unification and binding, as just discussed, are provided in the unification model by the connectivity between processors that in the human brain is provided by the thalamocortical network. Intentionality refers to the state or thing that a particular mental representation represents and is also rereferred to as "aboutness" (Ginsburg & Jablonka, 2019). In this unified model, the content of consciousness is determined by what is broadcast into the workspace. The information broadcast to the workspace could represent or be about either the environment or the internal state. UAL says that conscious contents must have neural effects that last longer than just a fleeting moment (Ginsburg and Jablonka, 2019). For these neural effects to persist in the system, UAL says that there must be some type of process where neurons send signals back and forth to each other or there must be a dedicated circuit with sequential activation of neural activation (Ginsburg & Jablonka, 2019). One possible mechanism they argue could provide this is "recurrent processing," where information is passed through a feedforward pathway then returned through a feedback pathway. These recurrent loops are a key part of GNWT as they allow information to be broadcast and become globally accessible. According to GNWT, these loops can amplify neural signals and sustain them for a period of time (Mashour et al., 2020). The capability of integration over time is therefore provided by the recurrent loops found within the global workspace architecture. The evaluative system required by UAL provides animals with the ability to label experiences with a specific valence and as positive or negative, which makes possible motivation and goal-directed behaviour possible. Looking at the human brain, the orbitofrontal cortex has been associated with reward value and emotional experience while the amygdala has been associated with responses to autonomic activity (Rolls, 2023). These activities are specifically related to responses based on rewards or punishments, but the idea of valence and evaluation still are present. With specialized processors such as these that have the function of placing value on incoming stimuli, this evaluative system can be built. The last two capabilities required for consciousness and unlimited associative learning, agency and embodiment and the registration of a self-other distinction, largely fit together. UAL states that for this sense of self to be constructed, the system requires both a mapping of the internal state of the body and the integration of sensorimotor information in order for these capabilities to be possible. In their model consciousness provides this integrating action, leading to the sense of ownership that leads to agency. By having the architecture necessary for these capabilities, the system then has the foundations necessary for unlimited associative learning.

To sum up the architecture necessary for the processes described in this unification model, the system must be able to create an internal model over the process of attention that it has cognitive access to, a large network of integrated and specialized processors that are causally dependent on each other and connected through a network of recurrent loops, long range transmitters that can send signals all around the system, processors dedicated to evaluation of incoming stimuli that can provide goal-directed behaviour, and processors dedicated to the integration of sensorimotor information that leads to a sense of ownership over the system. While various physical processes within the brain are required to support these functions, the widespread nature of consciousness among species as argued by this unification results in providing a list of required physical structures to be an impossible task. Structures within the human brain associated with these functions are discussed.

Creature Consciousness

The presentation of the unification thus far has established that consciousness is viewed as being widespread around the animal kingdom. This is drawn from UAL, which argues that the evolution of unlimited associative learning around 500 million years ago provided a hugely important adaptive benefit to animals that contained it and that enabled the animal kingdom to rapidly diversify (Birch et al., 2020). This section, then, intends to extend on this topic and investigate some animals that could be considered conscious. To test whether an animal has consciousness using the UAL framework, the species must be tested for whether they display the four main behaviours associated with UAL: discrimination learning, trace conditioning, flexibly changing predictions about patterns, and second-order conditioning (Birch et al., 2020). They argue that trace conditioning, second-order conditioning, and discriminative learning have been displayed in rats, rabbits, pigeons, and goldfish, and they say that honeybees and fruit flies display some of these behaviours (Birch et al., 2020). While these animals do not display the whole set of behaviours included in UAL, the theory states that these behaviours are tied together, and so they believe that the display of at least one of these traits indicates that the species may be able to engage in all of them (Birch et al., 2020). In order to fully test these claims, the ability of more species to display these behaviours would have to be examined more thoroughly. They say that unlimited associative learning is present in most vertebrates, as well as some arthropods and coleoid cephalopods (Birch et al., 2020).

AST says that as of yet, a strong top-down control of attention, as provided by the attention schema, has only been shown to be displayed by birds and mammals (Graziano, 2014). They do not claim that these are the only groups capable of the attention schema, instead saying that they are the only two proven to have these capabilities, and further research would need to be conducted on other types of animals. The attention schema is an important part of this unification model, so this is an important area that requires more attention. Animal groups that have shown no evidence of being capable of volitional control of attention include reptiles, amphibians, cephalopods, insects, and most fish. Interestingly, one species of fish, the archerfish, have been shown to succeed in a spatial cuing task testing for volitional attention (Nieder, 2022). Not all of these groups have failed this task per se, as there have yet to be designed experiments that test these capabilities for certain animal species. So, while it cannot be claimed that these other groups of animals are capable of controlling top-down attention, this possibility cannot be fully ruled out as of yet.

What is important to note in light of these findings, is that AST and UAL do not claim to be discussing exactly the same phenomenon. AST says that the beliefs and claims of having a non-physical consciousness arise from the subjective experience related to having an internal model of attention (Graziano, 2014). UAL, on the other hand, focuses not on the beliefs and claims experienced by humans, and instead hopes to examine the first evolution of minimal consciousness (Zacks & Jablonka, 2023). So, while the capability of volitional control of attention appears to be far less widespread than those capabilities explained by UAL, this does not necessarily indicate an incompatibility between the two theories. UAL focuses on a more limited form of consciousness, which could be seen as a necessary precursor for the development of the attention schema. In this case, UAL is more widespread, as it refers to a more basic form of consciousness than the attention schema might suggest.

Cephalopods are an interesting target of consciousness research, due to their vast differences in structure and function from mammalian species that are most commonly assigned research by theories. One model, proposed by Giovanna Ponte and colleagues, suggests that cephalopods can be considered conscious, due to meeting a set of three necessary conditions (Ponte et al., 2022). These three conditions include discriminatory and anticipatory behaviours that can provide an explanatory link between perception and memory, physical substrates that share functions with the human thalamus and cortex, and an understanding of the functional signatures and responses to different conscious states (Ponte et al., 2022). They claim that cephalopods meet all of these conditions, displaying a number of advanced cognitive and behavioural abilities. While the neural structures supporting cephalopods are quite different to certain vertebrates, such as mammals, they appear to encode a set of behaviours and abilities that would suggest some form of consciousness (Ponte et al., 2022). This model stresses that it cannot be expected for non-human species to display the same set of traits and conscious experiences as humans, and so human experiences should not be seen as the end-all-be-all of consciousness.

Panning back to the unification model, these pieces of research provide interesting expansions onto the question of creature consciousness. UAL suggests that unlimited associative learning, and thus minimal consciousness, is a process shared by a large proportion of the animal kingdom, spanning to most vertebrates, and some arthropods and celioid cephalopods (Birch et al., 2020). AST suggests that the attention schema is, based on the state of the current research, rarer than this, only appearing in mammals and birds (Graziano, 2014). So, while minimal consciousness can be viewed as widespread, the small distribution of the attention schema in the animal kingdom would imply also that the model of consciousness as described by the unification is similarly rare. Though current research only supports the ability of volitional attention in mammals, bird, and archerfish, further investigation is necessary to determine whether more species are capable of this and therefore possess some form of attention schema.

Machine Consciousness

Up to this point, the focus of this unification has been on evolution the biological aspects of consciousness. Biological conscious systems are the only ones currently known to exist and so it makes sense to focus on these systems. However, the question of machine consciousness should not be ignored, particularly with the recent societal popularity and focus on the topic of artificial intelligence. Before this consideration, however, it must be noted that in this unification, there is a shared belief with UAL that consciousness is originally a biological process that is driven by the biological need to survive, and so any artificially created consciousness would have to be assembled as a replica of this, yet would be radically different, due to the lack of embodiment within a physical body and the biological drive to survive. Nonetheless, the unification model holds that artificial consciousness is theoretically possible.

One model of artificial consciousness, proposed by researchers in Finland, suggests that consciousness cannot simply be placed into machines, and instead machine consciousness must "self-emerge" out of the operations and architecture provided to it (Fingelkurts et al., 2012). This model uses the operational architectonics theory, which believes that a pattern of brain activity emerges whenever a pattern of phenomenology occurs; this emergence of brain activity occurs as the result of many operations within the brain (Fingelkurts et al., 2012). In this view, these operations serve as the foundation for later, more complex phenomenology. In the construction of machine consciousness, they argue that these operations are what must be cloned into the machine, rather than the entire range of biological mechanisms found within living systems. Once these operations are in place, they argue that consciousness would emerge by itself from these by discovering that there is a functional advantage to having consciousness (Fingelkurts et al., 2012). From this, this artificial system would develop a sense of phenomenology in order to serve this function. What exactly this phenomenology would look like, is not argued by this theory.

Another model, proposed by Harry Haroutioun Haladijan and Carlos Montemayor, rejects the idea that any phenomenology can be implemented into an artificial system (Haladijan & Montemayor, 2016). Additionally, they say that processes such as emotions and empathy could not be placed into these artificial systems. They say that these processes rely heavily on phenomenology, and that their definitions are fundamentally built on subjective experience, and so without phenomenology, a machine could not have these processes, and would be limited to more information-based actions such as "emotion-like" responses based on environmental cues (Haladijan & Montemayor, 2016). On the other side of this, intelligence and thought can be viewed as just computation, which as they explain is what machines are designed for. This viewpoint is informed by their argument surrounding the dissociation between attention and consciousness, saying that these two processes are related, but separate. Attention by itself is not awareness, and so artificially creating an attention system in a machine does not by itself provide a machine with phenomenology and with conscious experience. Additionally, they argue that empathy and emotion arise somewhat from the biological drive towards survival and reproduction as well as social cooperation with other living beings, and so without this, artificial systems cannot be considered to have any phenomenal content (Haladijan & Montemayor, 2016).

While consciousness is held to be a biological process in all known systems with this feature, it is clear that artificial consciousness is a topic that cannot be ignored. Consciousness emerged as a biological property based on the biological need to survive, and so any artificial system that may be considered conscious is missing out on this major aspect. The lack of an embodied self and of biological needs and desires would mean that the phenomenology, if there is any, of this artificial system would be radically different to any other known conscious system. From these two models of artificial consciousness, more can be learned about what exactly this consciousness could look like. In one model, phenomenal consciousness is possible in artificial systems, but only if they derive for themselves a purpose that consciousness could serve, based on the operations cloned from biological systems. In the other, it is held that artificial systems could never have any phenomenal content, as they are too separate from the biological processes that gave rise to consciousness in biological systems. While describing the exact nature of artificial consciousness and what it might look like are beyond the scope of this thesis, either of these two ideas could make sense in light of the focus on evolution. As evolution in biological systems resulted in consciousness, it could be the case that some sort of artificial evolution, primed by an adaptive function of consciousness for machines, could result in some kind of artificial consciousness. On the other hand, this nature of consciousness as biological could mean that this could never be replicated in a machine, as the biological properties resulting from this biological drive could never be replicated in a non-biological system. Whether possible or not, it is important to note that artificial consciousness, were it to exist, would be radically different to the consciousness of living systems, due to vastly different driving factors that give rise to it.

Consciousness and Other Processes

The purpose of this section is to discuss the various processes that are typically implicated in or associated with consciousness. Specifically, the processes of attention, emotion, perception, action, and memory will be focused on. Much like how IIT argues that consciousness is identical to integrated information, there are a number of models that claim consciousness to be identical to these processes or at least primarily determined by these processes. By focusing on these accounts, information can be gained as to specifically how these processes are related to consciousness, what role they play in consciousness, and why it might be beneficial for consciousness to involve these processes.

Emotion

Robert Aumann presents an account of consciousness that views emotion as the key reason for the existence of consciousness. In this model, the adaptive function of consciousness is to provide emotions with the ability to operate properly (Aumann, 2024). This is built upon the idea that all behaviours and actions are driven by conscious emotions; emotional incentives provide the motivation for all activities, including meeting basic needs of life. Aumann says that while meeting these needs is required for maintaining life, the real in the moment reason for acting towards resolving these needs is done primarily in pursuit of the emotional incentive, rather than for the sake of survival. He provides the example of nourishment, saying that when people eat, they eat in order to resolve the feeling of hunger or because of the taste of the food, not out of knowledge that eating the food will provide their body with nutrients or sustenance (Aumann, 2024). In order for these incentives to exist, Aumann says that consciousness is required. He says that incentives are based on the emotion of desire and acting based on these incentives is volitional, and according to him desire and volition are functions of consciousness, and so are not possible without it. He also says that there must be a conscious inner world experience based on what is being desired and that this inner world experience is based on a conscious outer world experience based on sensation (Aumann, 2024). Aumann extends the importance of emotions onto nonconscious actions as well, saying that any unconscious activity has at some point been informed or instructed by consciousness, and so the purpose of this activity is still determined by emotions. This may seem to imply that all emotions are conscious and that unconscious emotions are impossible, yet Aumann stresses that this is not the case and that unconscious emotions are possible, they are just always built based on conscious emotions, seeing them as a kind of "branch" leading off of conscious emotions (Aumann, 2024).

Another model that focuses the discussion of consciousness onto the process of emotion is the "primordial emotions" model, presented by Derek Denton. This too places emotions as the driving force behind consciousness. This model, as the name suggests, focuses on "primordial emotions," described as the first emotions developed in animals, including thirst, hunger, and pain. According to Denton, these primordial emotions are signal to the animal that there is a major threat to the life of the animal that must immediately be dealt with (Denton, 2005). Once this signal is received, the animal is able to use the processes of conceptual categorization and memory to make an informed choice about the action needed to be taken in order to satisfy this emotion (Denton, 2005). The presence of a conscious awareness of the body's internal states allows these emotions to be amplified, compelling the animal to act towards addressing these emotions.

The key aspect from these models to be included in the unification outlined in this thesis is that emotions serve as a signal, priming the system as to what the immediate goals of the system should be. Whether these emotions are conscious as Aumann suggests or primordial emotions as Denton presents, they are able to cue the system to focus on a goal important to survival. In viewing the attention schema as the control mechanism for attention, emotions can be seen as the guide of what stimuli should be focused on to serve the current goal. If, for example, the being experiences a feeling of thirst, this signals to the attention schema that signs of a water source should be looked for in the environment. In the moment, these emotions do not necessarily have to be consciously accessed in order for them to be acted on, they can instead be "branches" off of previously constructed conscious emotions and goals, as Aumann suggests. When these emotions are conscious, however, it becomes possible for the system to provide extra resources to the task of addressing them. As the being would have constructed associations between stimuli in the environment related to the relief of these emotions, they can refer to these in the current moment when the emotional signal is sent, and they can therefore quickly respond to the cue.

Attention

Already captured in this unification model is AST, which makes a clear claim as to how consciousness is related to the process of attention. In this theory, consciousness is no more than the subjective experience of the attention schema. This shares similarities with the "identity thesis of consciousness," which suggests that consciousness is the same as attention. The key difference between what AST says and what this suggestion says is that AST does not claim that attention is consciousness, but that consciousness is the experience of attention. Harry Haroutioun Haladijan and Carlos Montemayor present an alternate account of the relationship between consciousness and attention, saying that they are dissociated (Haladijan & Montemayor, 2015). It is important to note that they do not advocate for a full dissociation where consciousness and attention are entirely separate and each function fully independently, but a more moderate form of dissociation. They argue that while it possible for attention to occur without consciousness, there is only weak evidence supporting consciousness without attention (Haladijan & Montemayor, 2013). They expand on this by saying that not all attention is conscious attention, and in fact there are many types of unconscious attention that evolved at various points, and so it must be the case that consciousness and attention are dissociated, as these processes evolved separately. While they do concede that it is likely the case that some forms of attention are necessary for consciousness, they hold that the majority of attentional processes occur unconsciously (Haladijan & Montemayor, 2015).

As the attention schema is included in this unification, a general basis of how attention is related to consciousness has already been provided. In this model, attention is controlled by the attention schema in order to select and attend to stimuli that are important to the survival of the system. The claim that consciousness is identical to attention is not one that is believed by this model, as instead it is argued that consciousness arises due to the internal model of attention, integration, and global broadcast. Thus, a level of dissociation between consciousness and attention is accepted in this model; they are not held to be identical processes. It is argued that some type of attention is necessary for some aspects of consciousness, as attention and the attention schema are required for the selection of stimuli that the system becomes conscious of. While it is agreed on that certain types of attention can occur unconsciously, this model holds a view of attention and consciousness that is less dissociated than the idea presented by Haladijan and Montemayor. By involving attention in consciousness, a system is better able to select information that is important to the survival of the system. Furthermore, the broadcasting of information into the workspace makes the subject of attention available to conscious processing by a variety of specialized processors around the network, allowing for more rigorous and dedicated processing. Attention is therefore viewed as a key process in consciousness.

Memory

The framework required for unlimited associative learning includes dedicated memory subsystems that allow formed associations to be stored in the long term and working memory is viewed as crucial for integration over time, a capability required for both consciousness and unlimited associative learning. While UAL makes it clear that memory is important to both of these processes, UAL is not entirely built upon memory and requires a variety of other processes as well. Contrary to this, Andrew Budson, Kenneth Richman, and Elizabeth Kensinger provide an account of consciousness that memory, specifically episodic memory and sensory, semantic, and working memory, give rise to consciousness (Budson et al., 2022). They make the claim that episodic memory provides the ability to create future plans through the "flexible recombination" of memories. In order for this flexible recombination to be possible, the system requires consciousness, and so this model makes the claim that consciousness first arose from episodic memory. They argue that episodic memory requires first the encoding of some mental representation from the past, the consolidation of the memory into storage, and then the retrieval of the memory later on (Budson et al., 2022). In this way, the processes of sensory, working, episodic, and semantic memory and consciousness are part of the same system (Budson et al., 2022). In explanation of what conscious experience is, they say that the conscious moment is not an experience of perception, but an experience of memories of the perceptions. To expand on this, they use the Cartesian theater metaphor, saying that within the mind of each conscious being, there is some "homunculi", sitting in a metaphorical theater watching and experiencing their memories (Budson et al., 2022).

This theory and UAL both share a focus on the evolutionary roots of consciousness, yet while this theory argues that consciousness evolved directly from the development of episodic memory (Budson et al., 2022), UAL argues that consciousness arose due to a number of interlinked processes, which included memory (Birch et al., 2020). Functionally, there is also a similar claim about how consciousness provides flexibility. In UAL, GNWT, and the unification model, consciousness is argued to provide the system with the ability to run various processes unconsciously while dedicating conscious resources from many processors to one specific task, allowing flexible decision making regarding the future behaviour of the system. This model proposed by Budson et al. (2022) shares this focus on flexibility in action and planning but posits that this flexibility arises from the ability to recombine memories in a flexible manner. Rather than focusing on flexibly dealing with information as it is integrated into the system, it focuses on flexibly dealing with information once it has been encoded into storage as a memory. A key part of this unification model is that it involves the actions of many specialized processors and subsystems, the functions of which are diverse and that through integration of the elements in the network, act together in support of the system as a whole. It does not argue that consciousness arises from the actions of just one type of processor or subsystem, but from the functions of the attention schema, integration, and global broadcasting of information, which encode a vast number of different processes and functions. So, while it agrees that memory is a hugely important aspect of consciousness, it does not share the belief with Budson et al. (2022) that consciousness arises from episodic memory alone. An important aspect of this model to include in the unification is the idea that conscious deliberation of memories can be used to create plans for future behaviour. After associations between stimuli in the environment have been formed and stored into the dedicated memory subsystems, they must then be assessed, in order to inform future behaviour based on what was learned. The formation and storage of these associations are not by themselves enough for the full functioning of unlimited associative learning, as this conscious deliberation is necessary in the case of complex associations.

Action

So far in this unification, action has been viewed almost as the end product of conscious thought. Information is received by the system, the system integrates and reviews the information, stores it, then examines it for how it can be used to act in future scenarios. However, it is not always necessarily the case that action always follows consciousness, as it could be that conscious thought is informed by action, rather than vice versa, such as acting in such a way to purposefully gain some sensory information. Living beings are inherently acting beings, and so how consciousness relates to action is of utmost importance. Provided in a summary of a series of their meetings on this topic, a group of researchers provide an overview of the standing of the current research field on the nature of the relationship between action and consciousness (Seth et al., 2015). One argument that they make is that consciousness is specifically designed to enable and support flexible action by providing the system with a number of methods that can be used for the planning and execution of actions within the environment. They also argue that an action is different from any other movement because of the presence of some kind of internal goal. In order to best pursue these goals, a conscious system is provided with the capability of action awareness. By consciously knowing and focusing on the action that is currently being performed, the system can more deeply analyze what is currently occurring and provide the task with more conscious resources when required. The degree of action awareness is dependent on experience with and ease of task (Seth et al., 2015). More than a specific account of how exactly consciousness and action are related, the work of these researchers provides a list of potential frameworks that could represent this relationship and a focus on the idea that consciousness and action are highly correlated with each other. They argue that, as a whole, the field of consciousness research should not ignore the role of action in shaping conscious experiences.

With these ideas in mind, it is important to note the important role action plays in this unification model. Already covered is the idea that action can be seen as the end goal of the creation of these associations. Adding onto this, action can also be seen in the earlier steps of consciousness. In order to probe the environment for important information and to understand the role that one plays in the environment, a system must take action within this space, and observe the associated results. Without action, a being can only create associations based on observation of the environment by itself. If instead the being wants to learn more about its own role within the environment, it must take actions to uncover these connections. As these actions are tied to specific motivational goals of the system, actions can be planned that directly address these goals and move the system towards their achievement.

From these various models of consciousness represented through the lens of single processes in the brain, the main takeaway is that consciousness is not a solitary phenomenon. There is no "consciousness organ" in the brain, solely responsible for the generation of subjectivity and experience. Consciousness, whatever its exact nature and mechanisms are, is built upon a variety of processes in the brain. Additionally, these processes are not always conscious or always unconscious, this depends on the current activity and the resources necessary for it. Based on these discussed models, and the four theories included in this unification, the processes of emotion, attention, memory, and action each play important and distinct roles in this unified framework.

Conclusion

What emerges from this literature search and review is a unification that attempts to pull together the central constructs of four major theories of consciousness and hopes to address the roles of controllability of attention, perceptual integration, and the role of memory and learning in the formation of associations. Through this, it tries to demonstrate a means of comparing theories by placing more weight on their similarities than differences and by focusing on their most central constructs. More specifically, it binds together the central constructs of the global workspace, attention schema, unlimited associative learning, and a maximally irreducible cause-effect structure. These constructs are brought together in the topics of the evolution of consciousness, why some systems are conscious while others are not, and how consciousness relates to other processes. In addressing these various topics, this unification presents the idea that the adaptive function of consciousness the ability to flexibly act on information through the actions of a variety of processors, which in turn allows for the formation of complex associations of stimuli in the environment that can be expanded upon and stored for long periods of time, and therefore able to increase its odds at survival.

In looking at the timing of a given subjective experience, the first step to occur in this model involves attention. For voluntary, selective attention, the attention schema is used to control attention. By controlling attention, a system is able to select information that is important to its current goal. In the lens of the proposed function of creating associations, the system can specifically attend to important stimuli in the environment that are important to the survival of the system. Also relevant to this process of controlling attention is the process of emotion. In deciding what the relevant information that should be attended to, emotion can serve as a signal of this. When a given emotion is experienced, for example hunger, attention can be controlled in a way to directly address this feeling, such as by searching for sources of food. After any given piece of information is attended to, it can be integrated into the system. The cause-effect structure allows for the system to determine the information that requires the dedication of more specialized processors through the level of integration. The maximally integrated complex within this system is indicated by a high phi value, which serves as this signal. The information in this complex, now determined to require further processing, is then broadcast into a global workspace by a process of rapid activation of the neural networks encoding this information, providing a variety of widely distributed specialized processors with access to the information. This workspace is very dynamic, meaning that the specific processors are not fixed, and can change from moment to moment. Through the actions of some of these processors, the system can create these complex associations described by UAL, and then store them in memory subsystems for future reference. In the moment of experience, the system can delegate certain tasks to be run unconsciously, while providing conscious access to a specific task that requires more dedication. When first encountering some novel stimuli, the system is therefore able to slowly and deliberately interact with it in a careful manner, while when encountering a familiar stimulus, the system is able to use the already created associations in order to react quickly.

Limitations and Future Directions

Like all studies, this one is not without its flaws. As discussed in the introduction, there are a vast number of theories of consciousness, and this review focuses primarily on just four of these. Each of these other theories is supported by some empirical research, so they each likely have important aspects that could be included in this unification. However, consulting every theoretical framework within the field would be far beyond the scope of this thesis, and additionally, the four theories presented in this thesis are each prominent and have important things to say about consciousness, and so were included for this reason. Furthermore, the topics related to consciousness included in this unification are not an exhaustive list, and there are a number of other processes and mechanisms related to consciousness that should be addressed in further unification attempts. This unification does not claim to fully explain the phenomenon of consciousness, and so further research and dedication to the study is required. Consciousness is a deeply complicated topic, and the questions it raises likely cannot be explained by the explanations and study of just one individual, but by the work of the field as a whole. This unification serves as an initial attempt at a unification of theories of consciousness using a construct-first based approach. Future research should use this focus on the central constructs of theories in order to specifically examine the similarities and differences held between major theories, paying particular attention to the overlap between them. The current research focused on adversarial collaboration between different theories is also promising, and the combination of these two methods and the concerted effort of the field could result in many important discoveries being made on consciousness.

Bibliography

Aaronson, S. (2014, May 21). Why I Am Not An Integrated Information Theorist (or, The Unconscious Expander). *Shtetl-Optimized*. https://scottaaronson.blog/?p=1799

Accelerating Research on Consciousness. (n.d.). Retrieved May 3, 2024, from https://www.templetonworldcharity.org/our-priorities/accelerating-research-consciousness

Anderson, E., Siegel, E. H., & Barrett, L. F. (2011). What you feel influences what you see: The role of affective feelings in resolving binocular rivalry. *Journal of Experimental Social Psychology*, *47*(4), 856–860. https://doi.org/10.1016/j.jesp.2011.02.009

Aumann, R. J. (2024). Why consciousness? *Neuropsychologia*, *196*, 108803. https://doi.org/10.1016/j.neuropsychologia.2024.108803

Baars, B. J. (1997). *In the Theater of Consciousness: The Workspace of the Mind* (p. 208). Scopus. https://doi.org/10.1093/acprof:oso/9780195102659.001.1

Baars, B. J., Ramsøy, T. Z., & Laureys, S. (2003). Brain, conscious experience and the observing self. *Trends in Neurosciences*, *26*(12), 671–675. Scopus. https://doi.org/10.1016/j.tins.2003.09.015

Barr, S., Laming, P. R., Dick, J. T. A., & Elwood, R. W. (2008). Nociception or pain in a decapod crustacean? *Animal Behaviour*, *75*(3), 745–751. https://doi.org/10.1016/j.anbehav.2007.07.004

Birch, J., Ginsburg, S., & Jablonka, E. (2020). Unlimited Associative Learning and the origins of consciousness: A primer and some predictions. *Biology & Philosophy*, *35*(6), 56. https://doi.org/10.1007/s10539-020-09772-0

Block, N. (2011). Perceptual consciousness overflows cognitive access. *Trends in Cognitive Sciences*, *15*(12), 567–575. https://doi.org/10.1016/j.tics.2011.11.001

Boutin, A., Blandin, Y., Massen, C., Heuer, H., & Badets, A. (2014). Conscious awareness of action potentiates sensorimotor learning. *Cognition*, *133*(1), 1–9. https://doi.org/10.1016/j.cognition.2014.05.012

Bronfman, Z. Z., Brezis, N., Jacobson, H., & Usher, M. (2014). We see more than we can report: “cost free” color phenomenality outside focal attention. *Psychological Science*, *25*(7), 1394–1403. https://doi.org/10.1177/0956797614532656

Budson, A. E., Richman, K. A., & Kensinger, E. A. (2022). Consciousness as a Memory System. *Cognitive and Behavioral Neurology*, *35*(4), 263–297. Scopus. https://doi.org/10.1097/WNN.0000000000000319

Cerullo, M. A. (2015). The Problem with Phi: A Critique of Integrated Information Theory. *PLoS Computational Biology*, *11*(9), e1004286. https://doi.org/10.1371/journal.pcbi.1004286

Chalmers, D. J., Kahn, A. J., & Everett, B. A. (1996). Facing up to the problem of consciousness. In S. R. Hameroff, A. W. Kaszniak, & A. C. Scott (Eds.), *Toward a science of consciousness: The first Tucson discussions and debates.* (1996-97651-001; pp. 5–28). The MIT Press.

Clausi, S., Iacobacci, C., Lupo, M., Olivito, G., Molinari, M., & Leggio, M. (2017). The role of the cerebellum in unconscious and conscious processing of emotions: A review. *Applied Sciences (Switzerland)*, *7*(5). Scopus. https://doi.org/10.3390/app7050521

Cohen, M. A., Dennett, D. C., & Kanwisher, N. (2016). What is the Bandwidth of Perceptual Experience? *Trends in Cognitive Sciences*, *20*(5), 324–335. https://doi.org/10.1016/j.tics.2016.03.006

Cottingham, J. (2013). *René Descartes: Meditations on First Philosophy: With Selections from the Objections and Replies*. Cambridge University Press. http://ebookcentral.proquest.com/lib/reed/detail.action?docID=1139548

Delfour, F., & Marten, K. (2001). Mirror image processing in three marine mammal species: Killer whales (Orcinus orca), false killer whales (Pseudorca crassidens) and California sea lions (Zalophus californianus).

Behavioural Processes, *53*(3), 181–190. https://doi.org/10.1016/S0376-6357(01)00134-6

Denton, D. (2005). *The Primordial Emotions: The Dawning of Consciousness*. Oxford University Press.

Descartes, R., & Cottingham, J. (2006). *The philosophical writings. Vol. 1* (18. print, Vol. 1). Cambridge Univ. Press.

Fazekas, P., Cleeremans, A., & Overgaard, M. (2024). A construct-first approach to consciousness science. *Neuroscience and Biobehavioral Reviews*, *156*, 105480. https://doi.org/10.1016/j.neubiorev.2023.105480

Fingelkurts, A. A., Fingelkurts, A. A., & Neves, C. F. H. (2012). "Machine" consciousness and "artificial" thought: An operational architectonics model guided approach. *Brain Research*, *1428*, 80–92. Scopus. https://doi.org/10.1016/j.brainres.2010.11.079

Gainotti, G. (2012). Unconscious processing of emotions and the right hemisphere. *Neuropsychologia*, *50*(2), 205–218. https://doi.org/10.1016/j.neuropsychologia.2011.12.005

Gallup Jr., G. G., Anderson, J. R., & Shillito, D. J. (2002). The mirror test. In *The cognitive animal: Empirical and theoretical perspectives on animal cognition* (pp. 325–333). MIT Press. https://doi.org/10.7551/mitpress/1885.001.0001

Ginsburg, S., & Jablonka, E. (2019). *The Evolution of the Sensitive Soul: Learning and the Origins of Consciousness*. MIT Press.

Graziano, M. S. A. (2014). Speculations on the Evolution of Awareness. *Journal of Cognitive Neuroscience*, *26*(6), 1300–1304. https://doi.org/10.1162/jocn_a_00623

Graziano, M. S. A. (2017, November 14). *Frontiers | The Attention Schema Theory: A Foundation for Engineering Artificial Consciousness*. https://www.frontiersin.org/articles/10.3389/frobt.2017.00060/full

Graziano, M. S. A., & Kastner, S. (2011). Human consciousness and its relationship to social neuroscience: A novel hypothesis. *Cognitive Neuroscience*, *2*(2), 98–113. https://doi.org/10.1080/17588928.2011.565121

Haladjian, H. H., & Montemayor, C. (2013). *Unifying Theories of Consciousness, Attention, and Conscious Attention*. 2470–2475. Scopus.

Haladjian, H. H., & Montemayor, C. (2016). Artificial consciousness and the consciousness-attention dissociation. *Consciousness and Cognition*, *45*, 210–225. Scopus. https://doi.org/10.1016/j.concog.2016.08.011

Heschl, A., & Burkart, J. (2006). A new mark test for mirror self-recognition in non-human primates. *Primates*, *47*(3), 187–198. https://doi.org/10.1007/s10329-005-0170-8

Igelström, K. M., & Graziano, M. S. A. (2017). The inferior parietal lobule and temporoparietal junction: A network perspective. *Neuropsychologia*, *105*, 70–83. Scopus. https://doi.org/10.1016/j.neuropsychologia.2017.01.001

Jones, R. C. (2013). Science, sentience, and animal welfare. *Biology & Philosophy*, *28*(1), 1–30. https://doi.org/10.1007/s10539-012-9351-1

Koch, C. (2004). *The Quest for Consciousness: A Neurobiological Approach*. Roberts and Company.

Lombardi, O., & López, C. (2018). What does "information" mean in integrated information theory? *Entropy*, *20*(12). Scopus. https://doi.org/10.3390/e20120894

Malik, A., Doerschner, K., & Boyaci, H. (2023). Unmet expectations about material properties delay perceptual decisions. *Vision Research*, *208*, 108223. https://doi.org/10.1016/j.visres.2023.108223

Mashour, G. A., Roelfsema, P., Changeux, J.-P., & Dehaene, S. (2020). Conscious Processing and the Global Neuronal Workspace Hypothesis. *Neuron*, *105*(5), 776–798. https://doi.org/10.1016/j.neuron.2020.01.026

Montemayor, C., & Haladjian, H. H. (2015). *Consciousness, Attention, and Conscious Attention on JSTOR*. https://www-jstor-org.reed.idm.oclc.org/stable/j.ctt17kk7j7

Nieder, A. (2022). In search for consciousness in animals: Using working memory and voluntary attention as behavioral indicators. *Neuroscience and*

Biobehavioral Reviews, *142*. Scopus. https://doi.org/10.1016/j.neubiorev.2022.104865

Oizumi, M., Albantakis, L., & Tononi, G. (2014). From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0. *PLoS Computational Biology*, *10*(5). Scopus. https://doi.org/10.1371/journal.pcbi.1003588

Plotnik, J. M., de Waal, F. B. M., & Reiss, D. (2006). Self-recognition in an Asian elephant. *Proceedings of the National Academy of Sciences*, *103*(45), 17053–17057. https://doi.org/10.1073/pnas.0608062103

Ponte, G., Chiandetti, C., Edelman, D. B., Imperadore, P., Pieroni, E. M., & Fiorito, G. (2022). Cephalopod Behavior: From Neural Plasticity to Consciousness. *Frontiers in Systems Neuroscience*, *15*. Scopus. https://doi.org/10.3389/fnsys.2021.787139

Povinelli, D. J., Rulf, A. B., Landau, K. R., & Bierschwale, D. T. (1993). Self-recognition in chimpanzees (Pan troglodytes): Distribution, ontogeny, and patterns of emergence. *Journal of Comparative Psychology*, *107*(4), 347–372. https://doi.org/10.1037/0735-7036.107.4.347

Ribary, U. (2005). Dynamics of thalamo-cortical network oscillations and human perception. In S. Laureys (Ed.), *Progress in Brain Research* (Vol. 150, pp. 127–142). Elsevier. https://doi.org/10.1016/S0079-6123(05)50010-4

Rolls, E. T. (2023). Emotion, motivation, decision-making, the orbitofrontal cortex, anterior cingulate cortex, and the amygdala. *Brain Structure and Function*, *228*(5), 1201–1257. Scopus. https://doi.org/10.1007/s00429-023-02644-9

Segundo-Ortin, M., & Calvo, P. (2022). Consciousness and cognition in plants. *WIREs Cognitive Science*, *13*(2), e1578. https://doi.org/10.1002/wcs.1578

Seth, A. K. (2021). *Being You: A New Science of Consciousness*. Dutton.

Seth, A. K., Verschure, P. F. M. J., Blanke, O., Butz, M. V., Ford, J. M., Frith, C. D., Jacob, P., Kyselo, M., McGann, M., Menary, R., Morsella, E., & O'Regan, J. K. (2015). Action-Oriented Understanding of Consciousness and the

Structure of Experience. In *The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science*. The MIT Press.

Sneddon, L. U., Braithwaite, V. A., & Gentle, M. J. (2003). Novel object test: Examining nociception and fear in the rainbow trout. *The Journal of Pain*, *4*(8), 431–440. https://doi.org/10.1067/S1526-5900(03)00717-X

Stefanucci, J. K., & Proffitt, D. R. (2009). The roles of altitude and fear in the perception of height. *Journal of Experimental Psychology. Human Perception and Performance*, *35*(2), 424–438. https://doi.org/10.1037/a0013894

Strawson, G. (2009). Realistic Monism: Why Physicalism Entails Panpsychism. *Journal of Consciousness Studies*, *13*. https://doi.org/10.1093/acprof:oso/9780199267422.003.0003

Tononi, G. (2004). An information integration theory of consciousness. *BMC Neuroscience*, *5*. Scopus. https://doi.org/10.1186/1471-2202-5-42

Tononi, G. (2015). Integrated information theory. *Scholarpedia*, *10*(1), 4164. https://doi.org/10.4249/scholarpedia.4164

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. Scopus. https://doi.org/10.1038/nrn.2016.44

Tracey, W. D., Wilson, R. I., Laurent, G., & Benzer, S. (2003). Painless, a Drosophila Gene Essential for Nociception. *Cell*, *113*(2), 261–273. https://doi.org/10.1016/S0092-8674(03)00272-1

Webb, T. W., & Graziano, M. S. A. (2015). The attention schema theory: A mechanistic account of subjective awareness. *Frontiers in Psychology*, *6*(APR). Scopus. https://doi.org/10.3389/fpsyg.2015.00500

Zacks, O., & Jablonka, E. (2023). The evolutionary origins of the Global Neuronal Workspace in vertebrates. *Neuroscience of Consciousness*, *2023*(2), niad020. https://doi.org/10.1093/nc/niad020