Assessing Consciousness Theory: A Systematic Scoping Review of 25 Years of Empirical Evidence for Neuroscientific 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

Cole Dembski

May 2020

Approved for the Division

(Psychology)

Michael Pitts

Acknowledgments

Many, many thanks to Michael Pitts, for providing guidance and encouragement throughout the process of putting together this thesis – and for accompanying me as I found, for the first time, an area of study that genuinely excited me.

Boundless gratitude and love for my family, without whom I would never have had the extraordinary opportunity to come to Reed and who have supported me throughout my academic experience here.

And, of course, to all those I have been fortunate enough to have in my life over the last four years: I cannot express the love and appreciation I have for you – thank you.

Table of Contents

List of Figures

Abstract

In the last 30 years, scientific interest in consciousness has grown steadily, as has the quantity and quality of relevant research. As understanding of the brain's relationship to consciousness has deepened and become more accurate, numerous scientific theories of consciousness claiming to be empirically supported have been developed. However, the data provided as evidence are often ambiguous or unclear as to their actual relevance to each of the proposed theories. Evaluating the scientific viability of these theories requires an objective review of consciousness research over the last few decades, and the subsequent assessment of the empirical support for each.

Recently, the need for such a review has been recognized, but, despite the plethora of data currently available, a comprehensive systematic review of consciousness research has yet to be performed. The combination of inconsistent terminology and the sheer variety of potentially relevant data poses a significant obstacle to merely compiling a comprehensive and appropriate selection of included studies; even after assembling such a collection, interpreting the findings and assessing their significance with regard to proposed theories of consciousness present numerous challenges, especially when those theories do not offer precise enough predictions to directly test them.

Given these difficulties, any thorough, unbiased, and accurate review of the available research on consciousness must utilize a structured methodology in its search and selection process, transparently reporting the process in detail such that it can be clearly understood and replicated. Using the PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews) methodology developed by Tricco et al. (2018), the current review surveys relevant experimental evidence from research conducted across the field to assess the empirical support for three promising theories of consciousness: integrated information theory, global neuronal workspace theory, and attention schema theory. The findings are then considered as a whole, bringing together the empirically-supported elements of the three theories to establish a basis for a unified theoretical model of consciousness.

"Each brain was just a trifle in the vast furniture of the world – a quivering jelly fitting inside a cup of bone, a tiny loaf covered by a hat, a poor sponge soaked with just a glass of wine – a fist enough to break it. How could a brain contain the sky? The brain painted the world with color … made it alive with sound, and gave it taste and smell … the brain dreamed up and forged all things that were – lutes, rooms, mountains, planets and stars."

— From *Phi*, by Giulio Tononi (2012)

A Science of Consciousness

All that is seen and felt, dreamt and imagined, believed and remembered is that which constitutes the reality of any conscious being. Joy, agony, excitement, shame – feelings so very real, so palpable – characterize the experience of being alive, yet each and every sensation depends on a numb, gray, 3-pound, tofu-textured mass inside the skull. Damage the brain in a certain way, and all the lights turn out – but damage it in another, and the only evidence might be a slight tremor in the hand; split it down the middle, and one half believes in God while the other is ardently non-religious. Scientists have observed molecules in superposition – but still none can answer how the 86 billion neurons that populate the brain combine to create conscious experience.

The Problem of Consciousness

The famous 'mind-body problem' was introduced by René Descartes almost 400 years ago in his philosophical treatise *Meditations on First Philosophy In Which the Existence of God and the Distinction between the Soul and the Body Are Demonstrated* (1641/1998). Widely considered to be the father of the modern philosophy of mind (Nagel, 1994, p. 65), Descartes reasoned that the one aspect of his experience he could be sure of was the existence of his own mind – his consciousness. In his *Meditations* (1641/1998), he expounds on what he saw to be a fundamental divide between a nonphysical mind and a corporeal body, a perspective now termed 'Cartesian dualism' (Shoemaker, 1994; Irani, 1980, p. 70). Taking consciousness as primary, he found himself confronted with the problem of attempting to explain how the mind controls the body. Descartes recognized that of all the parts of the body, the mind depended on the brain alone (1641/1998, p. 101), and suggested that the pineal gland – a small endocrine gland now known to be involved in regulating circadian rhythm – might be the seat of the interaction between the body and the mind due to its central location in the brain and its status as the only brain structure of which there is only one; the others exist bilaterally, one in each hemisphere of the brain (Edelman & Tononi, 2000, p. 8; Libet, 2004, p. 181).

In the centuries since, the mind-body problem has been one of the subjects most addressed by philosophers. Although multiple forms of dualism have been proposed, the dualist perspective began to decline in the first half of the 20th century as the scientific and intellectual spheres began to embrace the explanatory adequacy of physics, the belief that all phenomena can be explained and described by a set of governing theories that fall within the realm of standardly accepted scientific practices (Lewis, 1966). With the growing acceptance of a scientific model centered around testable hypotheses came the rise of materialistic monism, which asserts that all reality is made of one 'stuff' and that that stuff is of a physical, material nature – a definitive rejection of dualism (Searle, 2004, p. 60).

The materialist standpoint entails that the mind be wholly physical (Bernstein, 1968/2000, pp. 202-203), and over the course of the 20th century, philosophers grappled with trying to make sense of how consciousness could be explained in material terms (Fodor, 1994, p. 26). Countless versions first of behaviorist, then functionalist and identity theses, among others, were posited over the years, all deeply anchored in materialism (Searle, 2004, p. 48).

The behaviorist doctrine denies the existence of anything 'mental,' asserting that what are referred to as mental states are, in fact, really just a set of probabilistic dispositions to behave in a certain way in response to physical events (Fodor, 1994, p. 25; Irani, 1980, p. 60). From a behaviorist standpoint, if stimulus-response patterns are fully understood, the need for positing mental causes for behaviors will be eliminated (Fodor, 1994, p. 26). However, growing experimental evidence began to indicate that observed behaviors were only explicable by increasingly complex models of the internal processes mediating them, and many began to pursue a functionalist approach or turned to identity theories to explain mental events (Seager, pp. 24-25). Functionalism posits that mental states are defined by the functional architecture of a system along with its inputs and outputs, and that all systems with a given functional structure will have the same repertoire of possible mental states regardless of their physical bases (Fodor, 1994, p. 26; Seager, 1999, p. 24). Identity theories make the related but distinct proposal that every phenomenological experience is identical – not merely correlated with or caused by, but actually identical – to a physical state (Lewis, 1966; Levine, 1983).

The Explanatory Gap

Despite a general consensus that some version of functionalist or identity theories offered the best solution to the mind-body problem, many found neither to be truly satisfying, a view that persists to this day. While few would now argue against the notion that physical processes in the brain give rise to phenomenal experience, many feel that identity theories leave an "explanatory gap" between the experiential properties of the mental and the descriptive properties of the physical (Levine, 1983).

Strong identity theory can be boiled down to the following: experience X is neural process Y – not that Y causes X, nor that X must be accompanied by Y, but that they are quite literally one and the same (Nagel, 1974; Seager, 1999, p. 21; Searle, 2004, p. 55). The classic example used in discussions of identity theory is the relationship between pain and the physical processes that accompany it: according to identity theory, pain is identical with the firing of C-fibers (Searle, 2004, pp. 60-61). On one level, this can be easily accepted: C-fibers do transmit pain signals to the brain and, in one form or another, correspond strongly to the experience of pain (Levine, 1983; Searle, 2004, p. 61). However, the inadequacy of this explanation when it comes to consciousness lies in its fundamental abstraction. The knowledge that the firing of C-fibers amounts to the experience of pain is information invaluable to neurology, but this abstract concept bears no resemblance to the actual experience of being in pain, and no amount of training or mental effort will make it possible to genuinely experience pain as C-fibers firing (Seager, 1999, p. 6). Even the most precise, detailed, and accurate physical description of the neural processes behind the experience of pain will still fail to account for its actual phenomenological character, the crude feeling of it (Levine, 1983). The generation problem – how any type of brain activity might result in consciousness – is left unanswered (Seager, 1999, p. 24).

The Problems

In a 1995 article published in the Journal of Consciousness Studies titled "Facing up to the problem of consciousness," David Chalmers introduced the now-ubiquitous distinction between what he called the "easy problems" of consciousness and the "hard

problem." He proposed that the wide range of phenomena covered under the umbrella term 'consciousness' not only call for the use of multiple modes of investigation when studying it, but also demand fundamentally different types of explanations.

Some of the questions surrounding the nature of consciousness are clearly tractable within the confines of modern scientific practices, such as the neurological differences between sleeping, waking, and anesthetized states; the process of information recall; and behavioral control, to name only a few (Chalmers, 1995). These may all be adequately explained via functional or mechanistic descriptions of neural processes. For example, to understand memory, a functional description of the mechanisms underlying information recall can fully and satisfactorily account for the process. Understanding phenomena such as these are the easy problems of consciousness. Closing the explanatory gap – explaining how any of these mechanistic processes taking place in the brain are accompanied by subjective experience – is the hard problem, one which, from Chalmers' perspective, does not appear to be soluble using standard scientific methods.

It is not difficult to imagine that processes such as the focusing of attention can be fully understood through established scientific methods; in fact, many, if not most, of those involved in the study of consciousness fully expect that this will be done (Dehaene, 2014, p. 262; Damasio, 2018, pp. 159-160; Graziano, 2019, p. 146). Of course, these 'easy' problems are not easy in the sense of being effortless or even simple, but rather in that of being fairly straightforward. Though the search for the neural correlates of consciousness (NCCs) was initiated in 1990, namely by Francis Crick and Christof Koch (Crick & Koch, 1990), a full understanding of the neural activity associated with consciousness is still a goal far from being realized (Graziano, 2019, pp. 142-144). Brain imaging technologies continue to be frustratingly imprecise, artificial models of neural networks are little more than caricatures of the complex systems they attempt to copy, and roughly scanning and reconstructing a few millimeters of a thin slice of a mouse brain is as close as scientists have come to mapping the 'human connectome,' which would, in theory, be a complete map of the neural connections in the human brain (Graziano, 2019, pp. 139-146). However, with sustained effort and continued technological advances, it is reasonable to believe that it will be possible to develop

descriptions of cognitive processes that are accurate, detailed, and comprehensive enough to account for all of the functional processes associated with consciousness.

The easy problems have to do with understanding the functional brain processes correlated with consciousness; the hard problem asks how the performance of such processes is accompanied by consciousness. The hard problem is the problem of qualia, of explaining the feeling of experiencing. As of yet, there is no clear reason why certain cognitive processes should actually feel like anything as opposed to taking place "in the dark;" it seems just as likely that they might occur without any form of consciousness (Chalmers, 1995). Numerous hypotheses addressing the function of consciousness have been put forth, many centered around the potential evolutionary advantages it may offer (Feinberg and Mallatt, 2019; Koch, 2019, pp. 119-124; Damasio, 2018, pp. 156-161), while others have emphasized the importance of cultural development and memetics (Dennett, 2018b, pp. 344-345; Blackmore, 1999, pp. 237-239). Regardless of its function, however, the question remains of how consciousness 'emerges,' and many feel that even imagining a way by which it might possibly be explained is an impossible task (Nagel, 1974).

Approaching the Problems

Attitudes towards approaching the hard problem vary greatly. The primary area of disagreement among those involved in the study of consciousness is whether or not standard approaches in cognitive science will ultimately suffice to explain it or whether an "extra ingredient" will be necessary (Chalmers, 1995).

The latter perspective asserts that an entirely novel approach must be taken in order for consciousness to be understood, because no explanation of neural processes can ever explain why consciousness arises from such processes (Nagel, 1994; Chalmers, 1995). A popular proposal put forth by those who subscribe to this view is the hypothesis that consciousness may be part of the fundamental ontology of the universe, similar in nature to the basic laws of physics. This was the suggestion made by Chalmers in his original paper introducing the easy problems and the hard problem; the argument goes that, just as gravity, mass, charge and the like can be accepted without an attempt to explain why they exist, consciousness, too, may be a sort of irreducible force in the

universe. If true, the addition of a new law of this type – one which does not conflict with the known laws of the physical world – might be the solution to the hard problem. This sort of proposal, however, smacks of a new kind of dualistic thinking that attempts to construe the idea of a 'soul' as a potentially empirical claim by suggesting it could be incorporated into the scientific model of the universe – but how this proposal could ever be soundly and empirically tested is unclear, an issue explicitly acknowledged by Chalmers (1995).

The opposing view takes the approach that as more becomes known about the brain and its mechanisms, what is currently considered to be the hard problem will dissolve; it is only problematic folk intuitions that make the explanatory gap seem unbridgeable (Dehaene, 2014, pp. 261-262). Daniel Dennett argues that the very concept of the hard problem is ill-conceived and that scientists should instead focus on what he calls the "hard question," which asks, "once some item or content 'enters consciousness,' what does this cause or enable or modify?" (Dennett, 2018). Dennett denies the existence of qualia as an intrinsic aspect of conscious mental states and argues that instead, what are experienced as qualia are constructed representations that serve some purpose; understanding this purpose is the real challenge – and the key to understanding consciousness (Dennett, 2018).

The Meta-Problem

Despite the significant advances made in consciousness science over the last 25 years, many in the field – including Chalmers – feel that the hard problem appears almost as intractable as ever, and are of the opinion that attempting to explain consciousness from the angle of solving the hard problem itself may not be the right approach. Instead, a more indirect route may be necessary, useful, or, perhaps, actually solve the problem entirely, and recently Chalmers has shifted his focus away from tackling the hard problem directly to addressing what he calls the "meta-problem" of consciousness (Chalmers, 2018). The meta-problem is a problem about the hard problem that raises the question of why consciousness appears to be so mysterious. The rationale behind shifting focus to the meta-problem is that it is, fundamentally, an 'easy problem.' Chalmers (2018) suggests that since people can easily report on their experiences of consciousness

and, specifically, on their perspectives on consciousness, such as why they feel that the hard problem is (or is not) tractable, it may be useful to first try and understand why consciousness seems so inexplicable, before attempting to actually explain it. It is widely accepted that behavior can be explained via physical, mechanistic processes; it is also probable that the process of forming subjective phenomenological reports is closely linked to the hard problem itself – but collecting, analyzing, and interpreting such reports and the processes associated with generating them is a clearly tractable endeavor. Studying the meta-problem involves the explanation of physical mechanisms, technically making it one of the easy problems. The hope is that, if the meta-problem can be understood, it will shed some light on the hard problem, or at least constrain its potential solutions (Chalmers, 2018).

Overarching Issues in the Field

Just over a century ago, the phenomenon of light was a mystery. Some thought it was a wave; some thought it was a particle. Now, elementary science textbooks teach that light is photons: quantum entities whose characteristics waver somewhere between those of particles and waves (Hentschel, 2018, p. 133).

Scientists who study optics, the branch of physics devoted to the understanding of light, now conduct their research based on a fundamental understanding of light as photons. Until the mid-1920s, however, not only was light still poorly understood, but the field itself lacked a methodological or theoretical structure (Hentschel, 2018). While many innovative scientists posited theories about rays and particles and waves, there was little agreement between scientists not only on the nature of light itself, but also on the methods that should be used to study it, the phenomena that should be studied, and the questions future research should seek to address (Hentschel, 2018, pp. 93-132). Thomas Kuhn remarked in his seminal monograph *The Structure of Scientific Revolutions* (1970) that “though the field’s [optics] practitioners were scientists, the net result of their activity was something less than science. Being able to take no common belief for granted, each writer on physical optics felt forced to build his field anew from its foundations” (p. 13). His comment in no way disparages the contributions of these scientists, but rather points

out a basic obstacle for optical science that hindered and delayed the discovery of photons. Today, scientists who devote themselves to the study of optics operate from a shared photonic paradigm in which light is made up of photons, quantum entities exhibiting wave-particle duality, despite the fact that there is still significant controversy over the 'true' identity of light (Hentschel, 2018, p. 182). The quantum electrodynamic model, however, has been indispensable in its practical applications and theoretical contributions to quantum field theory (Hentschel, 2018, p. 182), and, crucially, was the result of synthesizing multiple theories of light to form a conceptual blend of models, which eventually developed into the current understanding of photons as neither exactly a particle nor exactly a wave but something else entirely – something that shares some properties with both (Henschel, 2018, p. 141).

A Mature Science

The state of optical science in the early 20th century bears a striking resemblance to that of consciousness science today. Light appeared utterly mysterious, perhaps unknowable. Scientists posited various theories, most of which turned out to be wrong; nevertheless, many of these erroneous theories contained at least a grain of truth which ultimately contributed to the formation of today's theory of quantum electrodynamics, though there was little communication across the field (Henschel, 2018, p. 185).

Consciousness has mystified scientists and philosophers for centuries, and there is no shortage of those who claim that it may actually be impossible to understand. Over the last thirty years, scientists have proposed theory after theory based ostensibly in contemporary neuroscience but, like the optical science of the early 1900s, consciousness science has yet to develop upon an accepted theoretical basis from which to operate. In *The Structure of Scientific Revolutions* (1970), quoted above, Kuhn explores the way in which specific sciences develop and mature over time, focusing on the concept of scientific paradigms. A paradigm, in Kuhn's sense, is an overarching conceptual model used in a field which has been developed based on previous findings and serves to unite members of a specific scientific community and guide future research in the field (1970, pp. 23-25). The existence of an accepted paradigm is the hallmark of what Kuhn (1970) calls 'normal' or 'mature' science (p. 10). Lacking a shared paradigm, researchers in a

particular field are forced to start from scratch, without a methodological precedent and with little to anchor their future work: “in the absence of a paradigm or some candidate for a paradigm, all of the facts that could possibly pertain to the development of a given science are likely to seem equally relevant” (Kuhn, 1970, p. 15).

Paradigms need not be proven to be entirely correct prior to being adopted by a field; as new discoveries are made, they should be updated, improved – and sometimes, in light of an unprecedented scientific achievement, revolutionized (Kuhn, 1970, p. 12). The purpose of a shared paradigm is to provide a system of scientific operation that brings together scientists in a given field of study, facilitates collaborative and focused research, and allows for more linear, streamlined scientific progress due to the presence of an accepted body of data that has already been synthesized such that it need not be explained or justified at every turn (Kuhn, 1970, pp. 19-20).

Consciousness science, however, has remained staunchly in the pre-paradigmatic stage. Myriad scientific theories of consciousness have been proposed, many of which are tightly linked to the few researchers who developed them and subsequently base their empirical work on their personal paradigm – not to mention the countless theories put forward by philosophers over the years (Revonsuo, 2010, pp. 177-225) As a result, much of the research on consciousness from the last few decades has been grounded in the specific interests and agendas of each coalition, producing a slew of available data that lacks a clear focus.

One should not be overly critical of the field for its non-paradigmatic state: given how recently researchers began to seriously investigate consciousness and how little was known at the start, it may be that only in the last few years has the development of a shared paradigm for consciousness science become possible, or even desirable. The different paths pursued by various scientists have resulted in a wealth of literature that covers a broad scope of potentially relevant areas – a necessary precursor to the formation of a paradigm. It is now time for consciousness science to avail itself of this abundance of data and move towards the formation of a scientific paradigm.

A Clear Terminology

In the first few pages of any of the multitude of books written on the subject, almost every author deems it necessary to define ‘consciousness’ – or at least, their own personal usage of the term (Dehaene, 2014, pp. 8-10; Graziano, 2019, pp. 3-5; Koch, 2019, p. 1; Prinz, 2012, pp. 4-7). Perhaps this can be explained by the fact that many such books are written to be accessible to the general public, but the authors of many articles published in academic journals feel obliged to provide a definition of consciousness as well (Dehaene & Changeux, 2011; Graziano, 2016; Tononi et al., 2016). The difficulties caused by this confused terminology have been acknowledged (Dehaene & Changeux, 2011; Graziano et al., 2019; Wiese, 2018; Chalmers, 1995), but there has been a strange apathy to actually rectifying this problem.

The root of the issue is that the word ‘consciousness’ is neither a scientific term nor a unified concept (Wiese, 2018). Colloquially, consciousness can mean all sorts of things – but, just as the terms ‘depression’ and ‘anxiety’ have clinical definitions that diverge from their typical quotidian usage, so the ‘consciousness’ of consciousness science must be given a formal definition that can be broadly understood by all engaged in the scientific investigation of it. Without a common language by which scientists in a field can communicate, the collaborative exchange of ideas becomes significantly more difficult and inefficient, if not impossible.

A trendy approach taken by neuroscientists in recent years has been to assign different ‘types’ of consciousness catchy labels such as “i-consciousness” and “m-consciousness” (Graziano et al., 2019), or “C0-,” “C1-,” and “C2-consciousness” (Dehaene et al., 2017). While making such distinctions is valuable – if not necessary, given the wide range of phenomena to which the word ‘consciousness’ may refer – this approach of independently differentiating between various aspects of consciousness using a novel, esoteric vocabulary can further confuse what is meant by the term.

It is not unlikely that what is currently referred to as ‘consciousness’ is not a unified entity but rather a cluster concept consisting of many different overlapping elements (Metzinger, 2004, p. 107). Attempting to pare down the meaning of consciousness as a scientific term to a single, highly specific definition may be not only

unrealistic but also undesirable. As the specificity of the definition increases, so does the specificity of its application. Excluding from a formal definition certain elements of consciousness would result in an incomplete conception of the object of study for the science of consciousness; instead, the diversity of consciousness should be embraced and represented in an agreed-upon, paradigmatic definition that can be used across the field.

This definition must be both informative and inclusive. It must make clear what consciousness science is about while being careful not to exclude relevant areas of study, yet still avoid falling into the circular trap of claiming that 'consciousness is being conscious.' The adoption of a such a definition would allow for the accommodation of new, novel, and possibly unprecedented approaches to the study of consciousness and facilitate the integration of multiple modes of thought by providing a shared understanding of what the 'consciousness' being addressed is, regardless of any extra-specific definitions of consciousness that may be used by individual scientists and research groups.

Some of the definitions that have been proposed include the act of being conscious (Graziano, 2019, pp. 3-6); conscious access, or the repertoire of content that may be conscious at a given time (Dehaene, 2014, p. 20); and "a state of mind in which there is knowledge of one's own existence and of the existence of surroundings" (Damasio, 2010, p. 157). None of these suggestions meet the requirements outlined above: the first definition comes very close to falling into the circular trap; the second seems to refer more to the range of things one can be conscious of in any given moment; and the third introduces additional conditions that stipulate the existence of a sense of self, the presence of an external environment, and the capability for higher-order cognition about the self and the world.

The intricacies of the aforementioned definitions which attempt to make specific the notion of consciousness end up being divisive more than anything else. If it is accepted that consciousness has characteristics that may be both described and investigated in different ways, it should likewise be accepted that the definition of 'consciousness' as a subject of scientific investigation must be inclusive of all of those characteristics. Christof Koch (2019) offers a more satisfactory definition: "consciousness is experience" (p. 1).

The generality of Koch's definition is exactly what makes it an appropriate definition for the field to adopt. Just as 'neuroscience' encompasses a number of specialized areas of study, 'consciousness,' as an object of scientific inquiry, need not refer to only one precise phenomenon – in fact, it should not. Its multiplicity must be encompassed in a general definition, while its numerous facets should be treated as such: they are diverse aspects of a diverse subject, not obstacles to the development of a mature science of consciousness.

Koch's wording is not without fault, however – 'experience' can itself be an ambiguous word which may have different meanings depending on the context in which it is used. The experience to which Koch refers is not that of acquiring knowledge in a specific area or of a 'learning experience,' but rather to that of having a subjective experience. This potential confusion can be easily eliminated with a slight rephrasing of his definition: consciousness is *experiencing*.

Challenges in Consciousness Science.

For decades, the development of a science of consciousness was impeded by a mistaken belief that the subjective nature of consciousness means that it cannot be studied scientifically: 'the science of consciousness' is a paradoxical term because science must strive for objectivity – the subjective has no place in it (Searle, 2000). Therefore, the argument goes, a science of consciousness is not possible. This claim, however, relies on a logic that is fundamentally fallacious and confuses the object of study (consciousness) with the mode of study (the scientific method).

The philosopher John Searle offers a useful terminology for thinking about this 'problem' by distinguishing between epistemic and ontological subjectivity and objectivity (Searle, 1997, pp. 113-114). The epistemic is concerned with knowledge; the ontological is concerned with existence. The subjective is observer-dependent, while the objective is observer-independent.

Something epistemically objective is a fact, like the statement 'dogs are kept as pets.' On the other hand, the claim 'dogs are the best pet' is epistemically subjective – a

matter of opinion. Each contains information, but the first statement is universally true, while the truth of the second statement is dependent on the individual.

Similarly, the ontological can also be either objective or subjective. An objective ontological statement might be ‘the sun exists.’ The sun exists regardless of whether there is anyone around to know that it exists. On the other hand, the pain of a bad sunburn is observer-dependent: it exists, but only for the sunburnt individual.

Science strives to be epistemically objective: if a claim cannot be verified independently of the person making that claim, it cannot be considered to be a serious scientific assertion (Gauch, 2012, pp. 26-28). Consciousness is ontologically subjective and exists only for the conscious entity in question, but its ontological subjectivity does not preclude the possibility of having an epistemically objective study of it. The science of consciousness is “the science of subjectivity,” as Anitti Revonsuo aptly subtitled his 2010 book on consciousness – but it need not be a subjective science. It is the latter that is a contradiction in terms, not the former.

Basic Difficulties

Nevertheless, merely establishing that a science of consciousness is possible does little to address the difficulties of actually collecting and interpreting the relevant empirical data, foremost being that any comprehensive research program on consciousness must utilize measures that assess the internal experience of an individual in tandem with external measures of brain activity and behavior (Dehaene & Changeux, 2011; Koch et al., 2016). Internal experience is most commonly measured via self-report, such as confidence ratings or reports of stimulus visibility (Dehaene & Changeaux, 2011). Objective observations of behavior may also be used to assess consciousness and internal experience, as in forced-choice paradigms when the subject is presented with a stimulus and subsequently must choose between two stimuli and select the one that was presented (Boly et al., 2013). Other external measures include electroencephalography (EEG), the monitoring of neural synchrony, and observations of widespread brain activation using functional magnetic resonance imaging (fMRI) (Seth et al., 2008). Data collected using these various techniques can be difficult to compare due to their fundamentally different natures, especially when considering observations of neural

activity alongside data collected via self-report, for example. The resultant findings must be interpreted together in order to identify possible ways in which the internally and externally descriptive data may be connected, but great care must be taken to avoid making unfounded inferences or inaccurate assumptions (Wiese, 2018). As such, it is crucial that the measures of consciousness used in an experiment be chosen conscientiously. Specifically, careful attention should be paid to the assumptions underlying the use of a specific measure and the level of description it seeks to provide (Seth et al., 2008; Metzinger, 2003, p. 110).

Biased and Conflicting Measures

The threat of theory-laden data is prevalent throughout the realm of science (Gauch, 2012 pp. 57-58). The subject of observation in an experimental paradigm is necessarily specific, selected because it is believed to be relevant to the question being investigated; the process of choosing what that subject is rests inevitably on a previous frame of reference (Popper, 1968, pp. 46-47). The wide variety of proposed theoretical frameworks for consciousness greatly exacerbates this concern, since many methods of measuring consciousness presuppose the validity of one or more frameworks (Seth et al., 2008). Self-report methods measuring awareness of a stimulus, for example, equate conscious states with states in which an individual is aware of being aware, the signature feature of higher-order theories (HOTs) (Seth et al., 2008). Measures of functional brain connectivity, on the other hand, have little to say about the metacognitive processes that HOTs emphasize and instead presume that the structure of neural networks is key to consciousness, a core tenet of some theories such as integrated information theory (IIT) (Oizumi et al., 2014).

Given the conflicts that exist between certain theories, different measures often provide conflicting results (Seth et al., 2008). For instance, measures of widespread brain activation, associated with integration-based theories, are typically employed with the assumption that greater widespread activation is correlated with consciousness, a hypothesis that has been broadly supported (Dehaene et al., 2001; Del Cul et al., 2007). On the other hand, there is evidence that consciousness as measured via subjective self-report – awareness of being aware, the cognitive process implicated in HOTs – correlates

specifically with activity in the mid-dorsolateral prefrontal cortex, while widespread activation correlates with task performance but not subjective awareness (Lau & Passingham, 2006).

Despite the complications that result from these conflicts, they can be used to identify future directions for research which may clarify the accuracy of aspects of different theories – for example, the need to investigate further the correlation between widespread activation and subjective awareness, as indicated by the example described above. Furthermore, attending to such conflicts facilitates the assessment of the validity of measures used: for example, objective behavioral measures of awareness assume that the ability to respond correctly at an above-chance level corresponds to consciousness, but subjects have consistently been shown to be able to perform at above-chance levels while claiming to have no subjective awareness (Dehaene & Changeaux, 2011). This suggests that forced-choice tasks may not be reliable indicators of conscious awareness but are instead often reflective of unconscious processing; for this reason, many researchers have gradually shifted away from using objective behavioral reports to measure consciousness (Dehaene & Changeaux, 2011). As such, experimental paradigms should aim to use multiple distinct measures of consciousness and compare the resultant data to validate or call into question the use of one or another, facilitating the refinement of both the measures and their theoretical bases (Seth et al., 2008).

Levels of Description and the Matching Problem

Regardless of their theoretical presuppositions, measures of consciousness can be divided into different categories based on their levels of description (Metzinger, 2003, p. 108; Wiese, 2018). The utility of any given measure differs based on the context in which it is used, and the method implemented should correspond to the domain to which it is applied (Goldman, 1997). Just as water can be described both on a micro level as an H_20 molecule and on a macro level as a liquid, consciousness can also be described from multiple perspectives. Imaging techniques such as EEG or fMRI, for example, can track brain activity at a cell-assembly level, giving a neural description of cognitive processes. Analyses of neural networks and their interactive behavior, on the other hand, offer a computational perspective of how information is processed, shared, and retrieved in the

brain, while subjective reports of awareness or introspection offer experiential descriptions at a phenomenological level (Metzinger, 2003, p. 110).

Making sense of the connections between these different levels of description is vital to developing a complete understanding of consciousness, but rarely is it clear what those connections are or how they function. This ‘matching problem’ becomes more and more difficult as the distinctions between levels of description grow: linking phenomenological self-reports to third-person neurobiological observations is a more dubious task than connecting observations of individual neurons to observations of neural systems (Wiese, 2018). Nevertheless, the unification of multiple descriptive levels is key to developing a comprehensive understanding of consciousness and should be considered a primary goal within the field.

The Modern Science of Consciousness

Despite the multiplicity of questions that still persist in the field of consciousness science, much has been learned since the basis of the modern approach to consciousness research was introduced in 1990 by Francis Crick and Christof Koch in a paper titled “Towards a neurobiological theory of consciousness,” in which they proposed that the most practical approach to attacking the problem of consciousness would be to direct research towards identifying the neural correlates of consciousness (NCCs), and suggested that visual awareness, specifically, might be a promising first area of investigation.

The framework that they proposed for future research on consciousness was founded on a few assumptions and laid out certain boundaries for which topics were appropriate to address at the time and which should be set aside. Those assumptions remain relevant today, asserting first that a scientific explanation of consciousness is needed, one which relies on an explanation of what processes in the brain correlate directly with awareness at any given time; and, second, that while there are many forms of consciousness, its varied facets rely on common mechanisms that apply across sensory and cognitive modalities (Crick & Koch, 1990). More specifically, they posited that research should be conducted under the assumption that some species of animals

experience a form of consciousness, at least those who display more extensive and complex cognitive abilities; as such, research on animal awareness should be considered relevant to the study of consciousness, and language should not be considered a crucial element of consciousness.

The limitations they proposed, on the other hand, have become dated – a testament to the progress that has been made in the field over the last thirty years. In 1990, Crick and Koch held that it was premature to pursue a precise definition of consciousness; that, initially, no neural theory of consciousness would explain all aspects of consciousness; and that the problem of qualia, or the subjective feeling of experiencing, would best be left aside for the time being. It may be that this last restriction still holds, since establishing whether the red seen by one person is the 'same' as the red seen by another goes a step further than understanding the mechanisms behind why someone consciously perceives red at all (Crick & Koch, 1990); however, the field has progressed enough in the last 25 years that both a definition of consciousness and a fully explanatory theory of consciousness are now possible and should be pursued.

Theoretical Directions

The multiplicity of theories of consciousness that have been proposed over the years poses a challenge to pinpointing which are most scientifically viable. Nevertheless, since the turn of the millennium, a few prominent theories have been posited, developed, and updated based on empirical observations: integrated information theory (IIT), which is based on the hypothesis that it is the structure of the connections within the brain which are crucial to consciousness; global neuronal workspace theory (GNWT), which posits that consciousness is tied to the widespread availability of information across the brain; and attention schema theory (AST), which emphasizes computational processes in the brain that construct a model of attention on a moment-by-moment basis. The diversity of these approaches and the maturity of the theories make them appropriate starting points from which to construct a framework for consciousness science.

Not to be considered here are theories invoking quantum mechanics, free energy, or microtubules; while alluring – chiefly due to the similarly mysterious nature of these phenomena and of consciousness (Chalmers, 1995) – such theories stray far from the

neurobiological approach taken by most scientists in the field. Turning away from investigating the neural substrates of consciousness and refocusing on entirely different areas of research is unwise and certainly premature: consciousness science does not appear to be so off-track that the neurally-based research program which has been pursued to date should be discarded in favor of pursuing possible connections between widely differing physical domains, such consciousness and its place in space-time geometry and quantum mechanics (e.g. Hameroff & Penrose, 2014). No research to date has convincingly indicated that any sort of quantum computations take place in the brain, and – while the brain is assuredly subject to the laws of both quantum and classical physics – there has been scant evidence that any special quantum properties influence neural processes (Koch & Hepp, 2006).

Theories of Consciousness

The three theories of consciousness assessed in this review – integrated information theory (IIT), global neuronal workspace theory (GNWT), and attention schema theory (AST) – were formulated by neuroscientists and research groups with varied backgrounds and areas of specialty. The approaches taken by the different theories are reflective of the diverse backgrounds of those who developed them, and each offers a different perspective that makes distinct hypotheses regarding the nature, function, and bases of consciousness and conscious experience.

Approaches to Theorizing Consciousness

Both scientific and philosophical approaches to consciousness can be broadly classified in a few ways. Few theories fit neatly into the categories discussed here – nor are the categories themselves necessarily mutually exclusive – but it is nevertheless useful to consider directly the general modes of thought underlying the contemporary development of consciousness theory.

Neurobiological Naturalism

The approach taken by neurobiological naturalism, a term coined by Todd Feinberg and John Mallatt, firmly rejects the conception of consciousness as mysterious and possessing a nature fundamentally different from that of all other physical phenomena, asserting that – while certainly unique – it is fully explicable via current scientific practices (Feinberg and Mallatt, 2018, p. 5). Instead of viewing consciousness as having a genuinely mystical character, neurobiological naturalism takes it as a natural phenomenon that emerged during the course of evolution as organisms grew more and more complex, positing that it possesses adaptive functions such as facilitating flexible responses to new and varied situations and the integration of large amounts of sensory input such that it can be used to guide behavior (Feinberg & Mallatt, 2019).

According to neurobiological naturalism, the reason that consciousness appears to be so inexplicable is its intrinsic complexity, but this complexity simply reflects the uniquely intricate structure of the brain (Feinberg & Mallatt, 2019). Consciousness is a highly diverse phenomenon, distinct even within the realm of biology. Feinberg and Mallatt (2018) point out that even recently, life was considered to be so exceptional that it might only be explained via some kind of vital force, the *élan vital*, but advances in the biological sciences have eliminated this idea and provided a natural, scientific explanation of life which most find to be entirely satisfactory (p. 1). Similarly, instead of being truly inexplicable, consciousness is merely an immensely multi-faceted feature of certain organisms which, as a result of its complexity, appears to be unfathomable. Instead of taking this unfathomability as a legitimate indication of the truly inexplicable and baffling nature of consciousness, the scientific community should proceed, undiscouraged, with the understanding that the solution to the hard problem will be just as complex and multifaceted as consciousness itself (Feinberg & Mallatt, 2019).

Illusionism

The illusionist perspective on consciousness is one that has only recently been taken seriously, partly due to an oft-held fundamental misconception of what an 'illusion' actually is (Graziano, 2019, pp. 96-97). The word 'illusion' does not refer to a false representation of something that genuinely does not exist, but instead to a misrepresentation of something that is itself ontologically real (Blackmore & Troscianko, 2018, p. 221). Illusionism is often interpreted as a full-on denial of the existence of any experience of consciousness, which is at best a misleading over-simplification of the approach, and, at worst, can lead to a deeply skewed understanding of it which tends to end in a vehement condemnation of the approach as asinine, bizarre, and "ducking the question" of the hard problem (Dennett, 2016; Chalmers, 1995).

Instead, illusionism asserts that the experience of consciousness is a fundamental misrepresentation of the actual processes occurring in the brain that is interpreted as phenomenal experience (Frankish, 2016). In other words, the brain's processors receive first-order inputs from the outside world via the senses which are interpreted by the brain as having phenomenal properties due to its incomplete interoceptive knowledge of the

nature of these sensory experiences. The warped, partial nature of sensory awareness causes the brain to feel as if it is having a subjective experience (Frankish, 2016). Most illusionists make the claim that actual experience is non-phenomenological and non-subjective and that, thus, there is no true subjective experience or phenomenological consciousness; however, it appears as if phenomenal consciousness exists because of the incomplete character of the brain's representation of its processes.

Illusionism comes in various forms which differ somewhat in one way or another, but, overall, it proposes that the experience of subjectivity and individuality of experience is illusory and that the 'qualia' agonized over by philosophers of mind over the years do not actually exist *per se*, but only as something an agent believes itself to be experiencing (Frankish, 2012). An important implication of taking this approach is that illusionists generally believe that the scientific study of consciousness should focus on the mechanistic and functional processes of the brain that lead it to believe it has consciousness, and that searching for what 'gives rise' to consciousness is hopelessly futile because consciousness does not actually exist for itself (Frankish, 2016). Searching for an answer to the hard problem will amount to nothing: there is no answer, because there is no problem.

Higher-Order Approaches

Often presented as "higher-order theories" (HOTs), various approaches to consciousness that posit that metacognitive, self-reflective processes are the key to consciousness have been popular over the years, including "higher-order thought," "higher-order representation," "higher-order representations of representations," and "higher-order perception" theories, to name only a few (LeDoux & Brown, 2017; Gennaro, 2012). Metacognition, as typically understood in higher-order approaches, refers to cognition about cognition (Lau & Rosenthal, 2011a); in other words, it is 'thinking about thinking.'

From this perspective, a mental state becomes conscious only when a higher-order thought process is directed towards it and forms a higher-order representations of that state (Gennaro, 2012, p. 14). Essentially, it is the state of being aware of one's own state of consciousness. A primary argument against HOTs is that they lead to infinite regress:

if consciousness is the product of a higher-order thought process that represent conscious experiences, then there must be a higher-order thought to accompany that representation, which must in turn be accompanied by another higher-order thought, and so on. However, a key, oft-overlooked feature of the HOT approach that is that the higher-order thought is not itself conscious, but is instead a representation of a given first-order state that allows for awareness of being in that state (Lau & Rosenthal, 2011a, 2011b). As such, the higher-order approach does not, in fact, lead to such infinite regress, though it does allow for the existence of 'third-order' thoughts directed at lower, 'second-order' thoughts (Gennaro, 2012, p. 30). Regardless, however, the higher-order approach is firmly grounded in the belief that a mental state is only conscious if it is the subject of a higher-order representation; in and of itself, however, that representation is not conscious.

Global Neuronal Workspace Theory

The global neuronal workspace theory (GNWT) first described in 1998 by Stanislas Dehaene, Michel Kerszberg, and Jean-Pierre Changeux is an expanded, updated version of the global workspace theory (GWT) proposed by Bernard Baars a decade earlier (Baars, 1988). When Baars first proposed the theory, consciousness research was still a fringe area of neuroscience and there was little direct evidence for the theory (Baars, 2002). However, as consciousness science became more widely accepted as a serious area of study and brain imaging techniques improved, empirical support for the theory began to accumulate (Newman et al., 1997). These advances allowed for the development of a theory firmly grounded in neural processes in the brain: the global neuronal workspace theory, described in detail in Dehaene's 2014 book *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts*.

Global Workspace Theory

Baars' GWT was initially based on a theoretical principle of integration within the brain, suggesting that a primary feature of consciousness is the widespread 'broadcasting' of information (Baars, 1997). The theory was grounded in the idea that there are two

types of processing which occur in the brain. One is unconscious and handles raw information coming in via the sensory system, while the other is responsible for the construction of a unified cognitive structure which integrates that information (Revonsuo, 2010, p. 206).

The brain can be divided into 'modules,' or low-level processing systems designed to handle only one type of input (Revonsuo, 2010, p. 206). These modules have little direct communication between them, but the different output messages they send to the brain are unified into a single experience in a process known as binding (Koch & Crick, 1994).

GWT posits that there is a conceptual space where this binding happens, a 'global workspace' where information is made widely available to distributed cortical structures in brain (Baars, 2002). The metaphor of a theater in the brain or a spotlight is often used to illustrate the theory. Attentional processes select specific stimuli for amplification and broadcasting which 'take the stage' momentarily (Baars, 1997, p. 44). Attention is selective, since the working memory capacity of the brain is limited: only a certain amount of information can be accessed at one time (Mole et al., 2011, p. xii). The process of attention shines a 'spotlight' on a specific piece of information which then becomes globally available to the many different regions of the brain, allowing for intermodular communication that facilitates the binding process. The information in the global workspace – in the spotlight – is the content of consciousness (Baars, 2005).

While Baars has remained active in theorizing GWT (e.g. Baars et al., 2013; Baars, 2017), the most prominent proponents of the global workspace model of consciousness are Stanislas Dehaene and Jean-Pierre Changeaux (Baars et al., 2013), who have worked to formulate an updated version of GWT that takes into account more recent research on neural networks conducted since GWT was first proposed in 1988 (Baars, 2002). Called the global neuronal workspace theory (GNWT), the model is essentially loyal to the fundamental concepts behind the initial GWT but is significantly more detailed and provides the theory with a firm grounding in empirical evidence.

Pyramidal Neurons and Global Availability

Like Baars' theory, GNWT posits that global availability is the crucial ingredient of consciousness. Additionally, it specifically implicates a type of neuron as being key to this process of global broadcasting, neurons whose long axons, measuring up to a meter in length, connect distant regions of the brain (Dehaene, 2014, p. 169). Called pyramidal neurons, their many branching dendrites project up into layers of the cortex, forming a complex network that connects the temporal, parietal, and prefrontal areas of the brain, among others (Dehaene, 2014, pp. 170-173). These neurons perform much of the work of relaying signals between regions of the brain as well as to the spinal cord (Koch, 2019, pp. 58-59). For this reason, discussions of GNWT often refer to these as 'workspace neurons,' positing that they are responsible for most of the binding that occurs after the entrance of stimuli into the lower processing areas (Dehaene & Naccache, 2001).

Feedback Systems, Top-Down Attention, and Accessing the Global Workspace

The complex connections between these neurons form a recurrent loop in which signals are sent from low-level modules to higher-order areas of the cortex and, crucially, are also sent in the opposite direction (Dehaene & Changeux, 2011). According to the theory, this feedback system allows for the sustained amplification of certain signals – essentially, the top-down control of attention (Dehaene & Naccache, 2001).

This sustained amplification of signals acts as a 'gateway' that enables content to become the main actor in the global workspace; neural activation, without the top-down control of attention and, correspondingly, the amplification of certain signals, is not sufficient for content to enter the global workspace, or consciousness (Dehaene and Naccache, 2001; Dehaene & Changeux, 2011). However, once a content does enter into the global workspace, it can maintain its position in the spotlight after the stimulus which originally triggered that content has become inactive, suggesting that top-down attentional control and the corresponding amplification of certain signals play a major

role in determining what content is conscious at any given moment (Dehaene et al., 1998).

Sustained Contents in the Global Workspace

Information that enters the global workspace and is retained after the initial stimulus is no longer present can be used in the brain independently of the time or space in which it was originally globally broadcast (Dehaene, 2014, p. 165). The storage and prolonged usage of conscious information allows for complex thought, the content of which is primarily determined not by immediate stimuli, but instead by mental contents previously stored in the brain. According to GNWT, when the resultant content of the widespread integration of information input from modular processors in the brain enters the global workspace, it becomes immediately accessible for use in modulatory cognitive processes, including the top-down control of attention, which are then able to exact a significant amount of control over the contents of the global workspace at any given time (Dehaene, 2014, p. 167).

Integrated Information Theory

Integrated information theory (IIT) was first proposed by Giulio Tononi in 2004 (Tononi, 2004), and has since continued to be developed by Tononi and others including Masafumi Oizumi and Larissa Albantakis, who, with Tononi, co-authored the most recently updated, detailed account of IIT (Oizumi et al., 2014). Other proponents of the theory include Melanie Boly, Marcello Massimini, and Christof Koch (Tononi et al., 2016), the latter of whom worked closely with Francis Crick in searching for the neural correlates of consciousness from the early 1990s until Crick's death in 2004.

The Dynamic Core Hypothesis

IIT is a direct descendant of the dynamic core hypothesis which Edelman and Tononi had previously introduced (Tononi & Edelman, 1998; Edelman & Tononi, 2000).

An initial understanding of the basic tenets of dynamic core theory is useful in understanding the more complex yet fundamentally consistent IIT.

By the turn of the century, the neuroscientific study of consciousness had been well underway for close to a decade (Revonsuo, 2010, p. 66; Crick & Koch, 1990), but the hard problem that had deterred neuroscientists for years continued to plague researchers in the field. It was becoming clearer and clearer to many that even an intricate understanding of the neural correlates of consciousness (a goal that even now, 20 years later, has yet to be attained) would not be sufficient to fully explain the phenomenological experience of being conscious; no matter how detailed, a description of the mechanisms underlying consciousness could never be translated into the actual feeling of being conscious (Edelman & Tononi, 2000, p. 11). While not a belief held by all, for Edelman, whose work on consciousness had centered primarily around the properties of neural systems as opposed to its specific neural correlates (Nunez, 2016), the 'explanatory gap' still stretched wide, and would not be bridged by an understanding of the NCCs alone.

In an attempt to address this, Edelman and Tononi took a decidedly non-reductive approach to forming the dynamic core hypothesis, beginning first by establishing the fundamental phenomenological properties of every conscious experience and then seeking to identify neural mechanisms that might explain those properties (Tononi & Edelman, 2000, pp. 18-19). They pinpointed integration and differentiation as two universal characteristics of every conscious experience. Integration refers to the unity of consciousness – it is impossible to have multiple experiences at the same time – while differentiation is the extreme specificity and informative nature of every conscious experience: any conscious state is only one of billions of possible states and, as such, is highly informative (Tononi & Edelman, 1998). Taking these two properties as fundamental, the dynamic core theory made a number of predictions regarding the neural processes that might contribute to consciousness, proposed a method of computing the level of integration of a system, and offered methods by which it might be tested (Tononi & Edelman, 2000). Over the next four years, Tononi developed and expanded on the theory and, in 2004, introduced his information integration theory of consciousness, which has since been updated multiple times to form the current version addressed in this

review, dubbed “Integrated Information Theory 3.0” (Tononi, 2004; Tononi, 2008; Tononi, 2012a; Oizumi et al., 2014).

The Structure of IIT

Integrated information theory takes the same basic approach as the dynamic core hypothesis: start with the phenomenology of consciousness, then ask what physical mechanisms might account for that phenomenology. IIT, however, meticulously formalizes this process and uses a logic-based methodology to orchestrate the move from the phenomenological to the mechanistic. It begins by establishing self-evident ‘axioms’ describing the fundamental features of any subjective experience. From these axioms, corresponding ‘postulates’ are derived. These postulates are structured assumptions about the physical substrates of consciousness that are logically inferred to be necessary properties of conscious systems based on the axioms (Oizumi et al., 2014).

Applying the postulates to proposed underlying mechanisms in the brain, IIT introduces a mathematical framework for measuring consciousness based on the level of information integration of any given system, denoted by **Φ**, or phi (Oizumi et al., 2014). A core implication of this model is that consciousness is graded as opposed to an all-or-nothing feature of certain systems. According to the theory, conscious experience is integrated information, and the greater the **Φ** value, the greater the level of consciousness (Koch, 2019, p. 88). It is this mathematized structure that most notably sets IIT apart from other contemporary theories, and significantly, offers a direct route towards the practical, clinical application of it; in fact, this is already being done in assessing the state of patients with disorders of consciousness whose level of awareness is unclear from an external perspective and difficult to assess (Koch, 2019, pp. 93-104).

Known as the ‘zap-and-zip’ technique, this method uses transcranial magnetic stimulation (TMS) and EEG to calculate a number known as the perturbational complexity index (PCI) (Koch, 2019, p. 100), which roughly reflects the complexity of integration and differentiation in the brain at the time of TMS and is used to measure the level of consciousness in a given subject (Massimini & Tononi, 2018, p. 120). TMS is a non-invasive technique that briefly stimulates cortical neurons by sending an electrical current through a coil of wire held to the scalp, which generates a magnetic field that, in

turn, stimulates electrical activity in cortical neurons (Massimini & Tononi, 2018, pp. 103-104). EEG recordings can provide spatiotemporal descriptions of signals across the cortex: dozens of electrodes placed on the scalp simultaneously record electrical activity and allow for the observation of patterns of brain activation (Koch, 2019, p. 99).

Upon administration of TMS – the 'zap' of zap-and-zip – stimulated neurons send signals to other cortical neurons which then impact the activity of other neurons, forming a network of neuronal activation (Massimini & Tononi, 2018, p. 105). EEG recordings spatiotemporally map this pattern of activity across the cortex, and the collected data is analyzed to determine how compressible the cortical response is – the 'zip,' borrowed from the algorithm used in file compression (Koch, 2019, p. 101). The compressibility of the data reflects the complexity and breadth of the patterns of neural activation induced by TMS: the more compressible the data, the less complex the pattern and the lower the PCI. High complexity and low data compressibility reflect a cortical response that displays high integration and high differentiation, and correspond to a high PCI value – and a high level of consciousness (Koch, 2019, p. 101).

The validity of this numerical index of complexity has been tested in healthy control subjects who could (or could not) report their experiences while awake, asleep, or under anesthesia, as well as in responsive brain-damaged patients (Massimini & Tononi, 2018, p. 122; Koch, 2019, p. 102). This allowed for the identification of a 'threshold' PCI value (called PCI*) for measuring consciousness in which values below PCI* indicate an unconscious subject, while those above indicate a conscious subject (Koch, 2019, pp. 101-102).

Following the establishment of the PCI*, the zap-and-zip technique was then applied to 81 patients who had been classified as either being in a vegetative state (VS) or a minimally-conscious state (MCS), correctly identifying all but two of the MCS patients and confirming the diagnosis of 34 of the 43 VS patients – alarmingly, however, nine of the VS patients had a PCI above PCI* (Massimini & Tononi, pp. 126-132). Of these nine patients, 60% recovered responsiveness in the following months, compared to 20% of patients with low-complexity PCI values and 0% of patients who displayed virtually no cortical response upon TMS, suggesting that these patients may have been misdiagnosed,

and also demonstrating the sensitivity of the PCI to conscious level (Massimini & Tononi, p. 131).

The zap-and-zip method shows great promise for detecting consciousness in patients who are either unresponsive or unable to respond; it also makes concrete and measurable the properties of integration and differentiation introduced by Edelman and Tononi two decades ago with their dynamic core theory, properties which now form the foundation of IIT.

Information in IIT

'Information,' as it is used in IIT, has a specific definition that may not be immediately obvious. It is neither information in the colloquial sense (e.g. "An encyclopedia contains a large amount of information") nor in the sense of 'Shannon information' which was introduced by Claude Shannon in his mathematical theory of communication (Tononi et al., 2016). Shannon information refers to the efficiency and accuracy with which a content is transmitted, but the meaning of the content is irrelevant; the informativeness of a given system is defined by the exactitude of the communication within it, or the extent to which a transmission specifies a given message from a range of possible messages (Shannon, 1948). Both these types of information are extrinsic and can be measured objectively in some form or another.

The information of IIT, on the other hand, refers neither to the process of information transmission nor to colloquial information. Instead, it is an intrinsic feature of a system which causally specifies the state of that system in the past, present, and future (Tononi et al., 2016). It must have a qualitative impact on the system; it is the "differences that make a difference" (Oizumi et al., 2014, p. 4).

The Axioms

Descartes' classic "I think, therefore I am" sums up the general premise behind establishment of the five phenomenological axioms of IIT: the only thing that can be immediately, definitively known – about anything – is the existence of one's own consciousness (Tononi et al., 2016). The axioms specify and partition the broad

characteristics of any conscious experience into concrete, intrinsic elements: existence, composition, information, integration, and exclusion (Oizumi et al., 2014).

> Existence: Consciousness exists for any conscious entity.
>
> Composition: Every conscious experience is structured, containing varied elements which can be combined in a multitude of ways; for example, one might have the experience of perceiving a green circle or a green triangle, a green circle and a green triangle, or a red circle and a green square.
>
> Information: Any conscious experience is highly informative in that it differs from every other possible experience – perceiving a green circle is necessarily different than perceiving a red circle.
>
> Integration: Conscious experiences cannot be reduced to their component parts without also reducing the extent to which they are informative. Perceiving a red circle and a green square cannot be reduced to the individual experiences of red, of a circle, of green, and of a square; instead, the color and shape are integrated into a single percept. In other words, any conscious experience is greater than the sum of its individual parts.
>
> Exclusion: It is not possible for one conscious entity to have multiple experiences at the same time. The same circle cannot be experienced as both entirely red and entirely green simultaneously.
>
> (Oizumi et al., 2014)

These axioms are often viewed as being too simplistic or trivial to be used as the basis of a scientific theory, but this is exactly the intention behind them: a dispute with a single axiom is, by default, a dispute with the entire theory. They are the foundation on which IIT is built.

The Postulates

The axioms act as an outline of the phenomenological qualia that a comprehensive theory of consciousness must account for. Their corresponding postulates are posited properties of physical mechanisms or systems of mechanisms which, in light of the axioms, should be true if they generate conscious experience; they are logically

deduced – but unproven – assumptions (Oizumi et al., 2014). The first two postulates paralleling the axioms are simple:

Existence: mechanisms underlying consciousness must exist.

Composition: mechanisms may be combined to form more complex mechanisms.

> The mechanism that creates the experience of green and the mechanism that creates the experience of a circle may be combined to create the experience of a green circle.

The subsequent postulates of information, integration, and exclusion rely on the first two postulates and apply their corresponding axioms to the mechanisms and systems of such mechanisms.

Information: Since each conscious experience is informative, the mechanisms underlying consciousness must have a causal effect on the functioning of the system. As such, these mechanisms must have an input-output structure and be able to exist in two or more states that are affected by inputs. The current state of the mechanism must likewise affect the output of that mechanism. The sum of the causal power of these inputs and outputs – in other words, how much the state of a mechanism may affect the system – is relative to the amount of information contained within the mechanism, called 'cause-effect information.'

Integration: To directly contribute to consciousness, a mechanism must be irreducible to its component parts. A mechanism that only specifies 'green' or only specifies 'circle' and does not integrate the two into the experience of a green circle cannot contribute to consciousness per se, since the phenomenological experience of a green circle is qualitatively different than the separate experiences of green and circle individually.

Exclusion: a mechanism that contributes to consciousness may not simultaneously specify multiple states of a system. Conscious experience is exclusive, so its key substrates must specify a similarly exclusive state.

(Oizumi et al., 2014)

From these postulates and axioms, IIT derives its mathematical propositions based on the structure of systems and the causal power contained within them. The intricate connections between neurons makes the brain by far the most complex system currently known to science (Squire, 2013, p. 15), but IIT does not suggest that any system – even those that are highly complex – is conscious. Instead, the theory posits that it is the structure of these neural networks that is key to conscious experience.

Causal Power, Complexes, and Integrated Information

Causal power is the extent to which the past and future states of a mechanism are constrained by its current state; in other words, to have causal power, the state of a given element within a mechanism must impact its future state and the future state of the mechanism. If an element has no effect on the mechanism, it adds no causal power to the mechanism; conversely, the more the state of an element or elements in a mechanism constrains the past and future states of that mechanism, the more causal power the mechanism has. As such, to have causal power, a mechanism must be structured in such a way that it forms a continuous feedback loop in which its past, current, and future states are intricately connected.

Figure 1 depicts a simple mechanism PQR consisting of three 'logic gates,' or switches that may be either in an 'off' state or an 'on' state, similar to the way in which neurons are, at any given moment, firing or not firing. The arrows show the direction of communication: P and R mutually affect each other, as do R and Q, but P has no direct impact on Q.

The state of each gate at a given time constrains the future state of another gate. Each gate in PQR has a distinct mode of function. P is an OR gate, meaning that if either R or Q is on, P will be on in the next time step. Q is a copy gate and will take the current state of R in the next step. R is an exclusive OR gate (XOR), meaning that if either Q or P is on, but not both, it will be on in the next time step.

In the initial state of PQR mechanism in Figure 1 (left; t_0), both Q and R are off, while P is on. Since P is the only gate that is on, R will be on in the next time step. Q will take the state of R, so Q will be off in the next time step. Since both R and Q are off in the current time step, P will be off in the subsequent one.

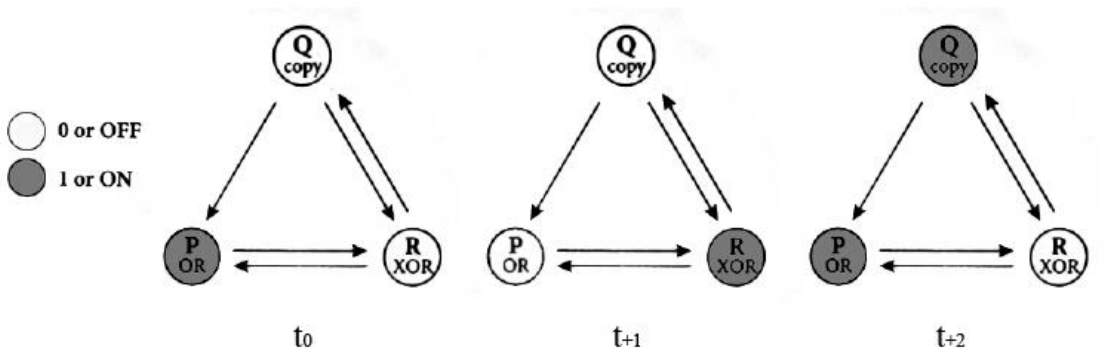

Figure 1: A three-node integrated mechanism

A simple mechanism PQR made of three logic gates which contains integrated information. In the initial state of the mechanism (t_0), P is on, while Q and R are off. In the next time step (t_{+1}) P and Q are off, while R is on. At t_{+2}, P and Q are on, while R is off. (Adapted from Koch, 2019, p. 83.)

The state of PQR in the next time step (Figure 1, center; t_{+1}), then, will be: P is off, Q is off, and R is on. In the time step after that (Figure 1, right; t_{+2}), Q will be on (copying R's state), P will be on (R was on; if R, Q or both are on in one time step, P will be on in next), and R will be off (neither P nor Q was on). Finally, the mechanism will return to its initial state: since both Q and P were on, R will remain off; since R was off, Q will copy that state; since Q was on, P will remain on.

In this way, the current state of the PQR completely specifies the state it will be on in the following one. While there are only three elements, the bidirectional connections between them and the conditions specifying the functioning of each form a complex mechanism with multiple potential states that are entirely constrained by its current state. Applied to neural mechanisms, PQR may be thought of as a highly simplified model of interconnected neurons whose interactions are specified by the structure of their synaptic connections.

The nature of this sort of feedback loop is such that the total quantity of information contained within a mechanism will be more than the total quantity of information contained in its constituent parts; this is known as integrated information. Individually, P, Q, and R being on or off specifies little about the state of the system; together, however, their mutual constraints give them causal power far and above the sum of their individual causal power.

As demonstrated by this PQR mechanism whose informative capacity is greater than the sum of the informative capacities of its parts, integrated information is irreducible – it cannot be divided into its individual elements. Just as seeing the word 'dog' is informative above and beyond the information contained in individually seeing the letters 'd', 'o', and 'g' when added together, for a mechanism to contain integrated information, its cause-effect information must be informative above and beyond the sum of the information contained in its component parts.

Systems containing integrated information may be made up of multiple mechanisms, each with their own level of integrated information, which is measured as **Φ**, or phi. The higher the **Φ** value of a mechanism or system, the greater the amount of integrated information contained within it. The precise method of calculating **Φ** is not necessary to understanding the general theory and will not be described in depth here (see Oizumi et al. (2014) for details on the mathematics of calculating **Φ**), and, in any case, calculating **Φ** for systems even slightly more complex than PQR is currently infeasible: only recently, it took several hours for a computer to calculate the maximal amount of integration possible in a system with only 8 elements (Massimini & Tononi, 2018, p. 76). The key concept is that the level of integrated information is determined by the structure of the connections within a system or mechanism.

The set of elements in a system that generates a local maximum of integrated information, known as the "maximally irreducible conceptual structure" (MICS), is called a complex (Oizumi et al., 2014). The MICS is the structure of integrated information with the highest **Φ** value that is generated by a complex in a specific state. According to IIT, integrated information is consciousness. It does not merely correspond to consciousness – it has an identity relationship with it; that is, consciousness is integrated information and integrated information is consciousness. As such, the MICS of a complex is itself conscious, and the structure of the MICS – known as a "quale" or "constellation of concepts" – at any moment in time defines the quality of its conscious experience (Oizumi et al., 2014).

Crucially, not all mechanisms in a system containing integrated information are conscious – it is only the set of elements with the highest **Φ** value. This is based on the principle of exclusion: it is only possible to have one experience at a time. If every

mechanism with a **Φ** value were to have its own individual consciousness, any mechanism with integrated information contained within another mechanism with a higher **Φ** value – for example, the MICS – would have its own experience, yet it would also contribute to the integrated information structure of the MICS. Logically, it follows that the experience of the MICS would somehow consist of both the experiences of its sub-mechanisms and its own experience – a possibility precluded by the exclusion principle. As such, only the mechanism with the highest **Φ** value, containing the greatest quantity of irreducible, integrated information, is conscious.

While the above is only a rough summary of IIT, it should be sufficient to understand its significance and its unique contributions to theorizing consciousness. At its core, IIT posits that the causal connections within neural networks form conceptual structures that cannot be reduced to their individual parts. The causal power of a network corresponds to the amount of information contained within it; information is considered to be 'integrated' if contained within an irreducible conceptual structure. The system or mechanism with the greatest amount of integrated information (the highest **Φ** value) is the system that is conscious, and the structure of that integrated information specifies the quality of conscious experience.

Attention Schema Theory

One of the more recently conceptualized theories of consciousness, attention schema theory (AST) has been posited and developed by Michael Graziano over the last decade, with its general tenets first described in *God Soul Mind Brain* (Graziano, 2010, pp. 51-92) and later in a 2011 article that Graziano co-authored with his wife, the neuroscientist Sabine Kastner. It was further expounded upon in detail and given its name in Graziano's 2013 book *Consciousness and the Social Brain*, as well as in numerous other articles in the years following (Graziano & Webb, 2015; Graziano, 2016; Graziano, 2017; Graziano, 2018a). Most recently, *Rethinking Consciousness: A Scientific Theory of Subjective Experience* (Graziano, 2019) gives a comprehensive and up-to-date account of the theory.

The Body Schema

Graziano spent the first 20 years of his neuroscientific career studying the way in which the brain monitors and controls the body in space by constructing a 'body schema,' an internal representation of the location and structure of the body (Graziano, 2019, p. 2). As is perhaps evident given the name of the theory, it is on this concept of schemas which AST is built. Understanding the body schema will facilitate an understanding of Graziano's theory of consciousness; it can even be considered to be the inspirational progenitor of AST (Graziano, 2019, pp. 2-3).

The body schema is a mental construct that facilitates efficient monitoring of the location, shape, and state of the body, chiefly in order to flexibly and accurately control movement (Graziano, 2013, p. 63). However, despite functioning to enhance the accuracy of movement, the body schema is not a technically accurate representation of the body. It contains no information regarding the muscular processes occurring during movement or the dilation of blood vessels during exercise (Graziano, 2019, p. 103). Rather, it approximates the weight and position of the body such that it may be effectively controlled. When reaching for a glass of water, the brain's body schema is utilized to accurately estimate the distance that the arm needs to move in order to pick up the glass, but the actual physical processes taking place are left entirely mysterious, no matter how much one might try to be aware of the changes in individual muscle fibers.

This failure to experience the mechanistic happenings in the body is no careless mistake, however. To operate effectively in the world – to run or scratch an itch or shake someone's hand – being constantly aware of such operational intricacies has little practical utility. On the contrary, closely monitoring all those processes would consume a huge portion of the brain's limited capacity and ultimately be detrimental to functioning effectively and efficiently (Graziano, 2019, pp. 40-41; Graziano, 2013, p. 28). Furthermore, the approximate nature of the body schema allows it to adapt so flexibly that external objects can be incorporated into the schema as 'extensions' of the body, like the hockey stick of a professional hockey player or the hammer of a skilled carpenter (Graziano, 2018b, pp. 110-111).

The brain utilizes schemas of similar types to facilitate the execution of numerous processes (Holland & Goodman, 2003). Recently, the predictive processing account of perception has gained traction in the field, proposing that the brain constructs a probability-based model of the sensory inputs it expects to receive (Kwisthout et al., 2017). This generated model may alter perception by adjusting what is subjectively experienced to match its expectations, and it constantly updates itself when prediction errors occur, a process called active inference (Seth, 2015). Naturally, this form of perception is imprecise and often inaccurate, and the level of detail of the predictive model and the likelihood of making a prediction error vary inversely (Kwisthout et al., 2017). The more detailed a prediction, the more likely it is to be incorrect in one respect or another – but, as predictions get simpler, they also become less informative. The predictive processing model attempts to strike a balance between these two. One way in which it does this is by increasing the salience of events which violate expectations; events matching the prediction can be effectively filtered out if they coincide with the predictive model and, as such, do not provide much in the way of useful information (Kwisthout et al., 2017). This attentional neglect of expected percepts allows for the direction of mental resources not only to noticing these violations, but also to updating the model as more information is acquired (Kwisthout et al., 2017).

Models of this type facilitate the speed and efficiency with which the brain may process and react to events going on around it (Holland & Goodman, 2003). The takeaway relevant to AST is that the brain generates rough, imperfect models of its processes, whose imprecision is exactly what makes them useful. These schemas are formed to filter out information irrelevant to the task at hand – broadly speaking, staying alive. They provide a sufficient basis from which to operate without crowding the brain with unnecessary information which would serve only to detract from its practical, day-to-day functioning.

The basic concept behind AST is that the brain constructs for itself not only perceptual schemas and a body schema, but also an attention schema. According to the theory, the brain builds a schematic, imprecise model of its process of attention, and this is what is experienced as consciousness.

The Process of Attention

Colloquially, attention is typically understood as concentrating on a specific task, i.e. 'paying attention.' This is not the modern scientific definition of attention nor that of 'attention' as used in AST. Instead, attention, in its most basic form, refers to the mechanistic process of selecting certain stimuli for deeper processing; the greater the salience of a stimulus or the more relevant it is to the current goal, the greater the likelihood it will be attended (Desimone & Duncan, 1995). This selection occurs at the expense of attending to other stimuli, a process broadly referred to as selective inhibition which, in the context of spatial attention, is known to involve 'lateral inhibition' (Graziano, 2019, p. 10). In lateral inhibition, since many areas of the brain are organized via spatial maps (e.g., of the visual field or the body), the most strongly activated neurons inhibit the firing of those surrounding them, leading to the perceptual enhancement of the attended stimulus and the suppression of unattended stimuli. This signal enhancement is the fundamental essence of basic attention: signals from attended stimuli are amplified, leading to the simultaneous inhibition of the firing of neurons responding to other stimuli. In Graziano's words, attention is "a data handling method" (2019, p. 110).

The limited capacity of working memory and cognition in general precludes the possibility of making sense of all the information coming into the senses at any given time, so the selection of certain stimuli for deeper processing is crucial for efficiently and accurately responding to important inputs in a complex sensory environment (Itti & Koch, 2001). As such, attention is essential to effectively monitoring one's surroundings. It can be a bottom-up process triggered by a salient external stimulus, or it can be consciously controlled in a top-down manner based on what is relevant to fulfilling a current goal (Graziano, 2019, p. 33). For example, while the brain might initially attend to the sudden sound of an ambulance's siren outside, it can also subsequently direct its attention to back reading a book.

Monitoring Attention

The brain is constantly monitoring its process of attention in order to exert some amount of control over it; otherwise, attention would shift willy-nilly from target to target

and result in complete perceptual chaos. When attention occurs without awareness, it can actually be detrimental to the performance of other tasks. For example, a low salience stimulus that is not strong enough to be consciously noticed may nevertheless be strong enough to attract some attention; this leads to the direction of mental resources away from the task at hand and is detrimental to overall performance (Tsushima et al., 2006). Somewhat counterintuitively, being aware of a distractor stimulus can actually improve task performance, because it is then possible to explicitly direct attention away from the distractor (Webb et al., 2016). In this way, the monitoring of attention is greatly useful for successfully functioning in the world: the ability to track attention allows the brain to direct it towards certain stimuli and away from others, filtering out unimportant information while processing that which is important

Attentional control makes it possible to attend to stimuli that are not immediately present, like trying to remember the words to a song or drawing an object from memory. Similarly, it makes it possible to, for example, listen in on a conversation taking place at a different table in a coffee shop while apparently reading the newspaper. While the newspaper is the most immediately salient stimulus, attention can be directed away from it and towards a different stimulus. This is a high-level cognitive process known as covert attention and may be closely tied to what is subjectively experienced (Graziano, 2019, p. 42). While overt attention, its more primitive counterpart, is anchored to physical stimuli, such as visually tracking a moving object, covert attention has no 'material' grounding. Of course, it is a physical process on a neural level, but, just as being aware of the cellular details of muscular processes would have no functional benefits, the brain has no reason to be aware of the neural processes underlying covert attention.

The Attention Schema

Attention is a neural process that is key to functioning properly; however, the brain has no need to be aware of inner workings of that process. Nevertheless, the ability to exert control over attention makes possible higher-level cognitive processes – namely, covert attention. Internal models are highly useful in facilitating efficient functioning in a host of different scenarios, from bodily movement to the navigation of self-driving cars. As such, the brain constructs an internal model that roughly traces its process of covert

attention. However, since this model is incomplete and the brain does not know the neural underpinnings of it, it must also manufacture for itself a representation of the process without any of those details. Since covert attention lacks a clear physical anchor, this representation gets interpreted as a nonphysical essence residing somewhere inside the body. This fabricated, apparently soul-like 'substance' is what is experienced as the self, as consciousness.

The attention schema is a sketchy model of attention, constructed to effectively control it – not to literally understand how it works. As attention shifts from item to item and thought to thought, it feels as if there is some inner, metaphysical entity having subjective experiences, even though no such entity exists.

William James wrote in *The Principles of Psychology* (1890) that the nature of consciousness is "exactly such as we might expect in an organ added for the sake of steering a nervous system grown too complex to regulate itself" (p. 144). This is the core thesis of AST: consciousness emerged when the brain, faced with an ever-growing barrage of stimuli and information, needed a simplified model of its attention in order to function properly. Instead of attempting to constantly track its entire attentional process, which would overload its limited capacity, the brain builds a rough representation of it that allows for the monitoring of relevant contents and the efficient direction of attention to important items. Since this representation does not provide information about the mechanistic underpinnings of attention, however, the brain does not perceive its basic computational structure and instead interprets the dynamic, incomplete model of its attention as its own consciousness.

Between-Theory Compatibility

The theories outlined above each have their strengths and weaknesses and call for different directions for future research, all of which should be considered. Notably, there are many areas of overlap and few direct conflicts between the three theories. GNWT's emphasis on global availability does not conflict with IIT's suggestion that the intricacies of neural connectivity and the informative capacity of such connections are crucial factors in consciousness. Nor does either preclude the possibility that it is the brain's

schematic modeling of its processes that lead to the feeling of phenomenal consciousness, as posited by AST.

What remains to be seen is which of these proposals – or, more likely, which aspects of them – will prove to be most directly relevant to consciousness, and to which features of it. Widespread neural connectivity is likely to be crucial in facilitating the integration of experience, but how global availability actually gives rise to conscious experience is not specified by GNWT. IIT has already shown promise as a measure of consciousness (Massimini & Tononi, 2018), yet whether its assertion that integrated information is consciousness can actually provide an answer to the hard problem is questionable. AST's suggestion that the rough, continuous modeling of attentional processes gets interpreted by the brain as consciousness may offer an answer to why consciousness feels ephemeral, abstract, and inscrutable, yet it does little to address issues such as why consciousness is often present in rapid eye movement (REM) sleep but absent in dreamless, non-REM sleep, for example (Boly et al., 2013).

Speculation and debate over the merits of each will only go so far, and progress in developing a paradigm for consciousness science continues to be hindered by competitive theorizing (Graziano et al., 2019). A comprehensive, evenhanded, careful assessment of the available empirical evidence will shed light on the strengths, weaknesses, conflicts, and areas of agreement between the varied theoretical approaches and allow for the establishment a framework around which future research on consciousness may be conducted.

Methods

Protocol

The protocol used in this review is based on the Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR), which was developed to serve as a reporting guideline for systematic scoping reviews to ensure both methodological transparency and reporting quality (Tricco et al., 2018).

Eligibility Criteria

Publications included in this review must have been available in English, have been subject to peer-review to assure the quality of the research presented, and have reported on original empirical research. To maintain contemporality, papers must have been published between 1995 and 2019. The development of many of prominent theories of consciousness taken seriously today began towards the end of the 20th century (Dehaene, Kerszberg, & Changeux, 1998; Tononi & Edelman, 1998; Damasio, 1999); thus, research beginning in 1995 was included to allow for the consideration of foundational research on which much of the current empirical and theoretical work being done in the field today is based.

Only research conducted on primate species was considered in this review due to the significant differences in cognitive ability and capacity to self-report between primates and non-primate animals. Non-human primates display a greater capability to self-report than non-primates, and there is evidence that they possess higher cognitive abilities than most other animals, making it easier to compare data from studies conducted using non-human primates and those conducted using human participants. Since research using non-human animals is advantageous due to the possibility of conducting experiments that cannot be carried out on human participants, studies using

non-human primates were included in this review, but those conducted on non-primate mammals and non-mammals were excluded.

Pilot studies, retrospective studies, case reports, proof-of-concept studies, and studies relying wholly on virtual models or simulations were excluded from this review due to the difficulties associated with assessing the validity and soundness of such research.

Studies centered around psychopathological disorders such as schizophrenia and dissociative disorders were excluded due to the potential for additional confounding variables. Studies using non-adult participants were excluded as well: the incomplete brain development of children complicates the comparison of neurological research on adults and children, above and beyond the need to control for age. Similarly, studies focusing on gerontology were excluded due to the possible neurological differences related to aging which can be difficult to account for. Finally, research using psychotropic drugs was not included in this review, since the confounding effects of such drugs are difficult to assess. While it is acknowledged that these exclusions may lead to the loss of some pertinent research, the pre-existing difficulties of conducting a comprehensive review of a wide array of consciousness research spanning 25 years call for the minimization of additional confounds, as well as for a general narrowing of scope to only the most relevant data. An in-depth account of the inclusion and exclusion criteria used and further details regarding the rationale behind the establishment of each can be found in Appendix A.

Information Sources

To identify relevant papers, searches were run on both the Scopus database and the Web of Science (WOS) citation index. The WOS search included results from the following databases: the Web of Science Core Collection, the BIOSIS Citation Index, MEDLINE®, and the SciELO Citation Index. The results were downloaded as .ua files and subsequently imported to Zotero, which performed an automatic duplicate check that was reviewed, and duplicates were manually removed as appropriate.

Search Method

Search terms were selected with the objective of compiling a comprehensive index of research papers pertinent to consciousness science published during the years 1995-2019. This presented a number of challenges due to the nature of the keywords necessary to retrieve the germane literature: for example, ‘consciousness,’ ‘awareness,’ and ‘perception’ are terms necessarily relevant to this review, but they may also be widely applied to a broad range of neuroscientific research which may not be directly related to the present subject of inquiry. However, attempting to refine the search by using more specific terms resulted in the exclusion of relevant articles, which was noted by cross-referencing the first 50 results of various searches conducted using narrower keywords with the first 50 results of a more general search. As such, the final terms used in the search query were chosen with the understanding that numerous irrelevant results would be returned upon running the search and would need to be manually filtered out. The following initial search was conducted (shown as entered on Web of Science, prior to any additional filtering):

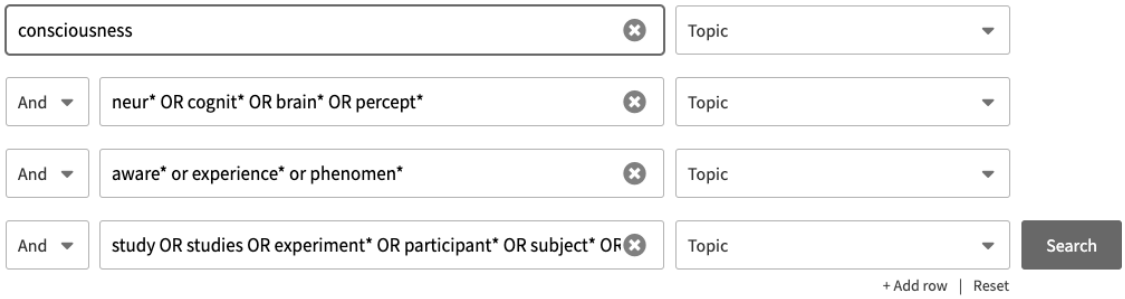

Figure 2: Initial Web of Science search query
An asterisk indicates a wildcard search term (for example, neur* will return results for neural, neurology, neuron, etc.). The ‘Topic’ field searches the title, abstract, and keywords of bibliographic entries.

The search terms were selected on the grounds that papers reporting empirical, neurological research should contain at least one of the terms from each row. The rationale for each is listed below.

1. Consciousness: Research relevant to consciousness science should contain the word “consciousness” in the abstract, title, or keywords.

2. Neur* OR cognit* OR brain* OR percept*: This review is primarily concerned with neurological research; the selected wildcard terms are intended to exclude blatantly non-scientific papers while remaining inclusive enough to avoid the unwanted removal of relevant work.
3. Aware* or experience* or phenomen*: These wildcard keywords were chosen to select for research centered around subjective experience and phenomenology.
4. Study OR studies OR experiment* OR participant* OR subject* OR volunteer*: Since the present systematic review is concerned exclusively with original empirical evidence, these terms were used to filter out review articles, under the assumption that any empirical report should include at least one of these terms or their variations in either the title, abstract, or tags.

After running this initial search, the analytical tools available on the databases were used to limit results to those available in English and those whose language was unspecified. A further filter was applied to include only results categorized as falling within the research areas of neuroscience or neurology, and papers tagged as pertaining specifically to pediatrics or gerontology were excluded due to the established exclusion criteria for the review. Finally, to filter out non-empirical articles, the following document types were excluded: reviews, meetings, case reports, books, editorials, letters, biographies, news, corrections, and retracted publications. Including filters, the final search criteria was as follows (as conducted on Web of Science):

TOPIC: (consciousness) *AND* **TOPIC:** (neur* OR cognit* OR brain* OR percept*) *AND* **TOPIC:** (aware* OR experience* OR phenomen*) *AND* **TOPIC:** (study OR studies OR experiment* OR participant* OR subject* OR volunteer*)
Refined by: RESEARCH AREAS: (NEUROSCIENCES NEUROLOGY) AND [excluding] **RESEARCH AREAS:** (PEDIATRICS OR GERIATRICS GERONTOLOGY) AND [excluding] **DOCUMENT TYPES:** (REVIEW OR MEETING OR CASE REPORT OR BOOK OR EDITORIAL OR LETTER OR BIOGRAPHY OR NEWS OR CORRECTION OR RETRACTED PUBLICATION) AND **LANGUAGES:** (ENGLISH OR UNSPECIFIED)
Databases= WOS, BCI, MEDLINE, SCIELO, ZOOREC Timespan=1995-2019
Search language=Auto

Figure 3: Final Web of Science search query
"Refined by" indicates the refinement criteria applied using the citation index's analytical tools.

Selection of Sources of Evidence

Searches were run on both databases on January 2, 2020, returning 2164 results from WOS and 1508 from Scopus for a total of 3040 citations, after removing duplicates; these were then filtered to remove papers that did not meet criteria. The abstracts of the remaining 1618 were screened for significance and relevance, of which 462 were selected for further consideration. 80 papers from the search were included in the review along with 26 identified through other sources; these 106 articles were included in the final review (see Appendix B for a full index of the articles included in the review).

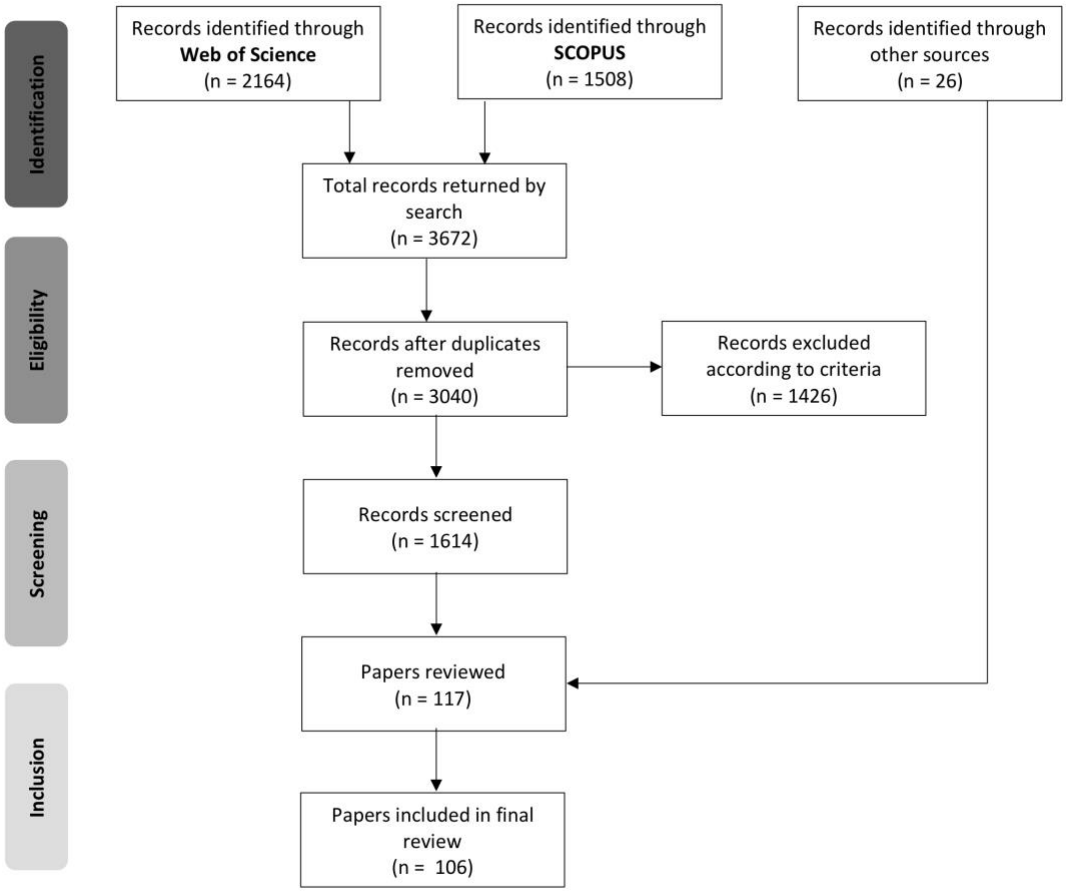

Figure 4: Flow diagram of the systematic review process

Data Charting Process

For each article, 12 fields of information were recorded in an Excel spreadsheet: the year of publication, title, authors, focus of the paper, quality (if notable), number of experiments, content assessed, task type, measure(s) used, and manipulation(s), as well as an overview of the experimental design and a summary of the main results. Three additional fields for noting the relevance of the results to each of the three theories were also filled where applicable, along with any further notes.

Results

The empirical evidence surveyed in this review principally applies to six different areas of interest relevant to consciousness science: the significance neural network properties; the mechanisms of perception; attentional processes; the effects of awareness; brain regions and activity correlated with conscious processing; and the bases of phenomenology. Given the distinct foci of IIT, GNWT, and AST, these subject areas are not equally relevant to all of the theories. Research investigating the properties of neural networks and the generation of phenomenology is most relevant to IIT due to the theory's emphasis on functional connectivity and its relationship to 'qualia' (Oizumi et al., 2014), while AST's focus on attentional processes, their interactions with awareness, and their contributions to subjective experience (Graziano et al., 2019) make findings related to attention, the processes facilitated by conscious awareness, selective mechanisms of perception, and phenomenal consciousness most applicable when assessing its robustness. The scope of GNWT's predictions mean that a broad range of research is germane to evaluating the strength of its hypotheses (Dehaene & Changeux, 2011), but the most telling is that surrounding the mechanisms of perception, attention, and the effects of awareness. These six categories of findings are discussed in the following sections and elucidate the diverse, complex, and dynamic nature of consciousness and the processes related to it.

Neural Network Properties

The properties of neural networks in the brain have been shown to be crucial to supporting basic consciousness and have long been posited to be primary markers of its presence or absence (Laureys et al., 1999). Evaluating neural network properties typically involves the use of one or more measures of integration, modularity and dynamic patterns of neural activity; indices of functional connectivity, complexity, and information sharing in the brain have been established as some of the most reliable neural indicators of levels of consciousness in patients with disorders of consciousness (DOC), as well as in healthy

individuals under anesthesia and in different stages of sleep (Casali et al., 2013; King et al., 2013).

Integration & Modularity

In neural networks, the extent to which information is exchanged between distributed areas in the brain and allows for long-distance neuronal interactions is referred to as integration, and is frequently quantified using characteristic path length (CPL) and clustering coefficients. CPL is the average functional distance between two nodes in a network, defined as the number of distinct connections required for a signal from the first node to be transmitted to the second; the shorter the path length, the more efficient the information exchange of a network (Sporns et al., 2004). The clustering coefficient of a system reflects its modularity, or the extent to which it is segregated into 'clusters' of local interaction as opposed to larger-scale networks. An integrated network is characterized by low clustering and short CPL, and swiftly transmits information throughout a widespread system.

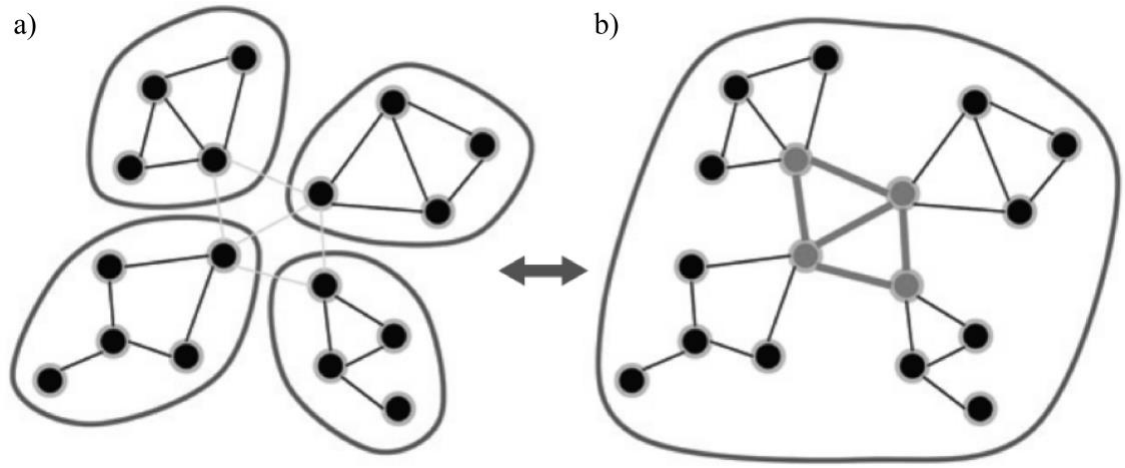

Figure 5: Network modularity and integration
a) A modular, segregated network. b) An integrated network with interconnected local modular structures. (Adapted from Sporns, 2013.)

A reduction in integration consistently accompanies loss of consciousness (LOC) in sleep (Uehara et al., 2014), anesthesia (Schrouff et al., 2011), and DOC (Zhou et al.,

2011; Silva et al., 2010). However, high integration does not directly correspond to consciousness: instead, a balance between functional integration and segregation is crucial, as predicted by IIT (Mäki-Marttunen et al., 2016). Boly et al. (2012) observed that during dreamless non-rapid eye movement (NREM) sleep, overall brain integration actually increases, but does so differentially for local and global networks. They found that both within-network (local) and between-network (global) measures of integration were higher in NREM sleep than in wakefulness, but integration within networks increased more in proportion to the increase in between-network integration, reflecting greater modularity in the functional structure of the brain. However, this specific pattern of differential changes in local and global connectivity during LOC has not been observed for all unconscious states: LOC induced by propofol anesthesia does not appear to differentially affect local and global connectivity (Monti et al., 2013). Furthermore, while modularity does increase in propofol-induced unconsciousness, it is maintained during emergence from anesthesia, suggesting that while greater clustering is associated with LOC, it may not play the pivotal role (Monti et al., 2013). On the other hand, CPL increases under anesthesia and decreases again upon recovery (Monti et al., 2013), suggesting that, in line with the GNWT model, the efficiency with which information is broadly transmitted across the brain may be a key factor supporting consciousness.

Decreased functional network efficiency characterized by increased CPL has also been observed in stage 1 NREM sleep (Uehara et al., 2014). In contrast to the results of Boly et al. (2012) discussed above, Uehara et al. (2014) found no significant difference in clustering or modular organization between stage 1 sleep and wakefulness. Importantly, however, Boly et al. (2012) analyzed data collected during deeper stages of NREM sleep (stages 2-4), which offers additional support for the hypothesis that modularity, while tightly associated with consciousness, does not play a critical role in determining conscious level. While recent research comparing brain activity in dreaming (conscious) NREM sleep and dreamless (unconscious) NREM has indicated that conscious NREM sleep is characterized by decreased local connectivity compared to unconscious NREM (Lee et al., 2019), it is well-known that even dreamless NREM sleep does not correspond to a complete lack of consciousness: some responsiveness to external stimuli is maintained, even in the absence of wakefulness (Boly et al, 2012). Scientific

understanding of states of consciousness during sleep is still incomplete; as such, it is unclear if or how these results can be generalized.

The importance of global interactions to consciousness has been further underscored in research with DOC patients. Some measures of information sharing in the brain such as weighted symbolic mutual information (wSMI) can accurately discriminate between coma-recovery patients in various states of consciousness when analyzed over medium- to long-distance neural connections, but fail to do so over very short distances (King et al., 2013). Upon emergence from a coma, patients may enter a vegetative state (VS; also referred as unresponsive wakefulness syndrome, or UWS) or a minimally conscious state (MCS) (Giacino et al., 2014). VS patients will awaken, move their eyes, and exhibit reflex responses, yet lack awareness of and responsiveness to their environment, while MCS patients are not fully conscious but inconsistently display signs of awareness, such as tracking the movements of others with their eyes (Laureys et al., 2010). The distinction between VS and MCS patients is a subtle one, yet crucial: while VS patients are effectively absent, MCS patients are evidently 'there,' despite possessing impaired awareness of their surroundings (Massimini & Tononi, 2018, pp 31-33). As such, investigating the neural distinctions between the two conditions can provide invaluable information about critical differences between consciousness and unconsciousness. The success of wSMI in distinguishing VS and MCS patients in analyses of mid- to long-range connections, and its failure to do so over short-range connections, offers significant evidence for the importance of global connectivity to consciousness.

Dynamic Activity and Complexity

In recent years, the mounting evidence that no single measure can accurately and reliably account for the network properties associated with conscious states has led to a shift towards investigating complex, dynamic activity in the brain to capture the nuances of the neural processes corresponding to consciousness. Propofol-induced unconsciousness in monkeys is accompanied by a change in functional connectivity such that it more closely follows the structure of anatomical connectivity (Barttfeld et al., 2015), a pattern also observed in patients with DOC (Demertzi et al., 2019). Anatomical

connectivity reflects neuroanatomical architecture, or the underlying connections in the basic structure of the brain independent of neural activity (Deco et al., 2011). In healthy wakefulness, functional connectivity, though modulated by the brain's underlying anatomical structure, is more dynamic and differentiated than in anesthetized states, influenced by spontaneous activity fluctuations as well as cognitive processes (Bartfeld et al., 2015). One characteristic of these changes in functional connectivity observed under anesthesia is a reduction in 'small-world' network properties (Bartfeld et al., 2015). Small-world networks have high levels of clustering (modular segregation) and short (efficient) paths globally connecting these clusters, known as 'nodes' (Sporns et al., 2004). An optimally efficient small-world network possesses a balance between segregation and global integration; the drop in network efficiency observed under anesthesia suggests that this balance is an important feature of conscious states.

The overall complexity of spontaneous neural interactions is also reduced under general anesthesia (Schartner et al., 2015), and the IIT-based perturbational complexity index (PCI), which assesses network integration and informational capacity, can successfully discriminate between levels of consciousness in wakefulness, sleep, anesthesia, and coma patients in various stages of recovery (Casali et al., 2013).

Multiple patterns of interareal coordination have been identified in healthy individuals and DOC patients (Demertzi et al., 2019). Conscious states are characterized by most frequently displaying a dynamic pattern of long-range positive and negative coherence across widespread areas of the brain, while VS patients overwhelming display low coherence the majority of the time; MCS patients are more likely to display dynamic patterns of coherence than VS patients, but are also more likely to have periods of low-coherence brain activity than healthy controls (Demertzi et al., 2019). Furthermore, increasing levels of consciousness correspond to more frequent transitioning between patterns of neural activity (Demertzi et al., 2019), indicating a greater ability to dynamically switch between different modes of functioning.

Brain Regions and Markers of Consciousness

Neural activity associated with sensation and perception begins to take place immediately after the onset of a stimulus and can be observed across a range of time windows and brain regions. However, not all of these process and the regions or time windows in which they are observed are necessarily directly linked to awareness: some are critical to both conscious and unconscious perception; others correlate strongly with awareness but may be more reflective of post-perceptual processing; and still others may not be necessary for any form of perception at all. Distinguishing between these markers is no simple task, especially considering the dynamic changes in the functional structure of the brain and the preexisting difficulties associated with assessing subjective experience. However, understanding which processes are necessary for perception will facilitate the construction of a framework for conscious processing.

Early Posterior Activity

Early processing in the primary visual cortex (V1) is crucial to perception (Hurme et al, 2017), but activity prior to ~120 ms has not been demonstrated to correlate reliably with conscious awareness of stimuli. Much work has been focused on identifying event-related potential (ERP) components, which are electrophysiological neural responses to specific sensory, cognitive, or motor events (Luck, 2014, p. 4). While some have posited an early positive ERP known as the P1 as a correlate of awareness (Pins & Ffytche, 2003), the results are inconsistent (Koivisto et al., 2008; Koivisto & Grassini, 2016; Andersen, 2016), and where a correlational effect has been observed, it is weaker than other neural signatures of conscious perception (Davoodi et al., 2015) and is now thought to reflect early attentional processes (Zhang & Luck, 2009). On the other hand, a negative posterior ERP component occurring from approximately 120-300 ms known as the visual awareness negativity (VAN) has been consistently found to be the earliest correlate of visual awareness (Koivisto & Grassini, 2016; Andersen et al, 2016; Koivisto et al., 2008).

The VAN is observed most strongly in posterior areas, especially in the occipital lobe (Koivisto et al., 2018; Meijs et al., 2018); activity in these regions has been shown to be the most reliable predictor of the quality of perceptual experience, including graded

changes in awareness (Andersen et al., 2016; Koivisto & Grassini, 2016), and appears to operate independently of attentional load (Koivisto et al., 2018). This temporo-parietal-occipital region is associated with integrative processes and has been suggested to be a "posterior hot zone" crucial to the formation of conscious percepts and modulation of consciousness (Koch et al., 2016; Boly et al., 2017).

The Ventral and Dorsal Visual Streams

Visual information moves from area V1 through higher visual areas and association cortices and ultimately towards the frontal lobe; this visual system is thought to have two 'streams,' dorsal and ventral. The dorsal stream encodes spatial and landmark information, and the ventral stream is involved in symbolic and representational processing (Lambert et al., 2018). While the extent to which the two visual pathways are differentiated in terms of their contributions to conscious perception is not fully established, it has been suggested that ventral stream activity relates to the

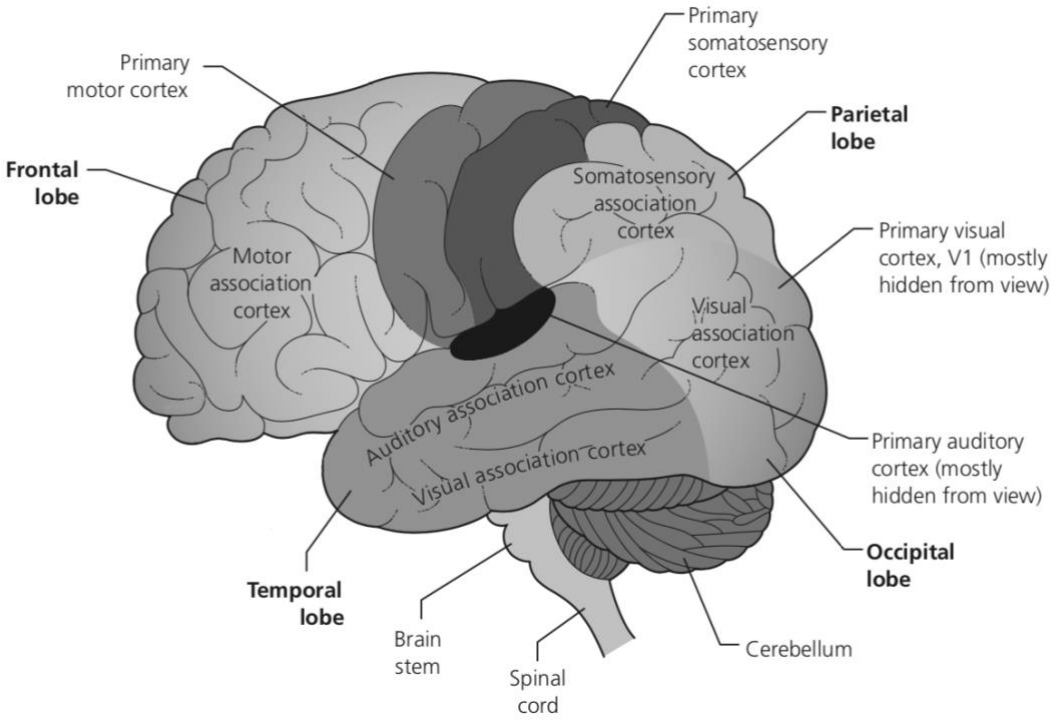

Figure 6: Brain regions and sensory areas

(Adapted from Blackmore & Troscianko, 2018, p. 78.)

formation of conscious visual representations and that activity in the dorsal stream corresponds primarily to unconscious processing related to visually-guided actions. Activity in the ventral stream correlates closely with conscious perception, decreasing linearly with subjective visibility (Zhan et al, 2018; Ludwig et al., 2016), and applying TMS to this area reduces the subjective clarity of stimuli, specifically features such as shape or color (Overgaard, 2004). The ventral pathway is populated by long-distance feedback projections to V1 from higher-order areas, and top-down attentional effects are known to progress down the visual areas of the ventral stream from V4 to V2 and finally V1 (Clavagnier et al., 2004; Buffalo et al., 2010),

However, activity in the dorsal stream does not correspond exclusively to unconscious processing. Visual masking is known to affect processing in the dorsal stream, and when stimuli are suppressed using a technique known as continuous flash suppression (CFS), a similar reduction in activation can be observed in both the dorsal and ventral streams (Hesselmann & Malach, 2011; Hesselmann et al., 2018). However, while activity in the ventral stream reflects graded levels of awareness corresponding to visibility reports, dorsal activity decreases in a step-like fashion when stimulus contrast reaches a certain threshold (Ludwig et al., 2016). As such, while both the dorsal and ventral streams are likely to be involved to some extent in conscious processing, ventral activity is more closely linked to reports of graded conscious awareness.

Late Frontal Activity

After moving through the higher visual areas, activation related to conscious processing progresses to the frontal regions of the brain. The frontoparietal network is the other brain region (along with the aforementioned posterior hot zone) commonly proposed to be critical to consciousness (Odegaard et al., 2017). Part of this network, the prefrontal cortex (PFC), is involved in attentional control, executive functioning, and working memory, and is the brain region most emphasized in GNWT (Mashour et al., 2020). Due its role in cognitive control as well as its extensive connectivity within the frontoparietal network, it has often been considered to be a prime candidate to play a causal role – if not *the* causal role – in consciousness (Koch et al., 2017; Odegaard et al., 2017).

However, the relationship between consciousness and the frontoparietal network is hotly debated, and mounting evidence suggests that, while frontal areas do play a role in modulating the contents of consciousness (Del Cul et al., 2009), anterior brain regions and neural activity may not be direct correlates of consciousness. Network activity in posterior regions distinguishes MCS from VS patients better than in activity frontal areas (King et al., 2013), and neither frontal network properties nor fronto-parietal connections reliably index reports of consciousness vs. unconsciousness in NREM sleep, whereas networks in the posterior parietal-occipital region display greater transitivity in correlation with reports of unconsciousness (Lee et al., 2019).

Activity correlated with conscious processing in this frontoparietal region typically begins around 300 ms post-stimulus onset and is commonly characterized by an ERP component known as the P3, a positive electrophysiological waveform observed from approximately 300-500 ms (Koivisto & Grassini, 2016; Andersen et al., 2016). While the P3 tends to display a strong correlation with stimulus awareness in experimental paradigms and has been proposed to be a principal marker of conscious awareness (Sergent et al., 2005; Del Cul et al., 2009), mounting evidence suggests that P3 activity reflects post-perceptual processes, not awareness per se, and is more closely associated with confidence evaluations and decision making (Pitts et al., 2014; Pitts et al., 2012; Koivisto et al., 2018; Koivisto et al., 2016). However, the question remains open, and some findings suggest that the P3 may reflect both confidence and conscious perception (Ye et al., 2019).

General findings around frontal activity are similar to those found with regard to the P3: while frontoparietal activity has been shown to correlate with conscious awareness of a stimulus (Del Cul et al., 2007), evidence suggests that this activation is not directly linked to consciousness. Stimulus-driven activity in the anterior cingulate cortex and PFC can be observed during inattentional blindness when a critical stimulus is presented, yet unnoticed, and this neural activity is similar to the stimulus-driven activation that is present when the critical stimulus is consciously perceived, suggesting that the observed frontal activity is not a direct correlate of awareness (Thakral, 2011). Furthermore, when presented with a sequence of either low- or high-contrast task-relevant arrow stimuli, frontal activity in response to the first stimulus does not differ

between the two contrast conditions, but diverges significantly for the following arrows (de Lange et al, 2011), indicating that frontal activity is reflective of top-down modulatory processes, not perception itself.

With regard to specific percepts, TMS-induced phosphene perception is not reflected in frontal activation, but is correlated with activity in posterior and central areas (Taylor et al., 2010). On the other hand, some illusory percepts are represented in frontal areas. In the double-drift illusion, a vertically-striped patch moving straight up or down is perceived as following a diagonal path if the stripes simultaneously drift horizontally across the patch; perception of this illusion can be decoded reliably in frontal areas – but also in a number of other regions, including the temporo-parietal junction, located in the posterior-central regions of the brain (Liu et al., 2019).

Core Networks

The default mode network (DMN) and the attentional network (AN) are two key networks associated with awareness that have contrasting functions, and the network connectivity between the two has been implicated as being closely related to states of consciousness (Mäki-Marttunen et al., 2016; Tagliazucchi et al., 2013). The DMN is a centrally located network with significant global connectivity that is associated with resting-state activity and self-related processing (Uehara et al., 2014; Tacikowski et al., 2017), displaying increased activation during internally-directed thought, introspection, and mind wandering (Vanhaudenhuyse et al., 2011). In contrast, the frontoparietal AN is engaged in cognitively demanding tasks and executive control (Vanhaudenhuyse et al., 2011; Mäki-Marttunen et al., 2016). The two are typically anticorrelated in their activity due to their opposing functions (Vanhaudenhuyse et al., 2011), but this relationship is disrupted in patients with DOC: while healthy individuals typically display high within-network correlations and low between-network correlations, in DOC the two networks display reduced clustering and increased inter-network correlations (Mäki-Marttunen et al., 2016). This excessive integration reflects a decrease in the functional specificity of each, resulting in less differentiation and over-correlated neural activity. In deep sleep, the DMN and frontal regions are known to decouple and display reduced complexity, which suggests that dynamic control over the functional interactions between these two

networks may play a key role in modulating consciousness (Tagliazucchi et al., 2013; Horovitz et al., 2009), and offers support for the hypothesis that disrupted connectivity between the DMN and AN may be an important contributing factor in DOC.

Mechanisms of Perception

The majority of research on perception has focused on vision due to the dominant nature of the visual modality in humans, its high level of detail, and the ease with which visual stimuli may be precisely manipulated in experimental paradigms (Crick & Koch, 1998). This approach was taken under the tentative assumption that different sensory modalities share common mechanisms (Crick & Koch, 1990), a supposition that has recently found empirical support (Sanchez et al., 2019). As such, the majority of the findings reviewed below relate to visual perception and to the extent to which they can be extrapolated for use in understanding perceptual processes in other sensory modalities.

Basic Processes of Visual Perception

Immediately following the presentation of a visual stimulus, the primary visual cortex (V1) becomes activated, followed by area V2 (Bullier et al., 1996; Boehler et al., 2008); this very early activity in V1 is critical for both conscious and unconscious perception, but by 90 ms post-stimulus onset, only conscious perception is blocked if activity in V1 is disrupted (Hurme et al., 2017). A feed-forward sweep of activation starting around 70-100 ms after stimulus onset moves through the higher visual association cortices, and conscious detection of a stimulus is typically characterized by the onset of rapid recurrent processing between the lower and higher visual areas around 100-120 ms after stimulus presentation (Boehler et al., 2008; Levy et al., 2016).

Until around 250 ms, significant neural activation can be observed regardless of subsequent awareness, at which point activity reflecting non-perceived stimuli typically begins to fade out, while full conscious perception is characterized by a sustained wave of neural activation towards frontal areas of the brain, consistent with GNWT's prediction that distributed cortical activation correlates with consciousness (Herman et

al., 2019). For a stimulus to be completely perceived, it must evoke a sufficiently strong neural response in V1 to reach a threshold level of activation necessary to trigger this swift, widespread propagation of neural activity, known as neural ignition (van Vugt et al., 2018). The likelihood that stimulus-induced activation 'ignites' is impacted by both stimulus strength and the pre-stimulus state of the brain: stronger stimuli elicit more activity in V1, and higher levels of neural activity in V1 prior to stimulus onset increase the likelihood that the stimulus-induced neural response will reach the necessary threshold level of activation (van Vugt et al., 2018).

Despite the non-linear nature of the neural ignition correlated with stimulus awareness, however, consciousness is not a unilaterally all-or-none phenomenon (Levy et al., 2016). Partial perception is characterized by 'piecemeal' awareness of a stimulus, in which information about some but not all of the features of a stimulus are available for conscious access (Kouider et al., 2010). Many complex stimuli are processed at different representational levels: for example, a written word can be processed at the low level of its spatial characteristics, the intermediate level of single letters, or at the high semantic level involving the meaning of the word (Levy et al., 2016). Under conditions of low visibility, stimuli may be perceived in a degraded or fractionated form in which only some of these representations are consciously accessible (Zadbood et al., 2011; Kouider et al., 2010).

Alone, the ignition model does not offer an account for this type of piecemeal perception. However, in the early stages of visual processing, around 100 ms post-stimulus presentation, early high gamma-band activity that is present in the absence of stimulus awareness (i.e. stimulus-absent or null perception) is suppressed, effectively filtering out irrelevant or nonspecific information and allowing for piecemeal perception. Around 250 ms, and consistent with the neural ignition model, middle gamma-band activity is enhanced in full but not partial perception (Levy et al., 2016). These findings suggest that there is a two-step mechanism modulating the contents of consciousness: first, an early inhibition of non-informative activation allowing for partial perception, and a second increase in activity corresponding to full perception. In line with this dual-mechanism model, behavioral evidence has indicated that discrimination accuracy follows a linear pattern corresponding to subjective visibility ratings when visibility is

low, but displays a step-like improvement when visibility is high (King & Dehaene, 2014).

In summary, neural activity following stimulus presentation is first observed in V1 and then moves to the higher visual areas; subsequent conscious perception is correlated with recurrent activity between these areas starting around 100 ms post-stimulus onset. Higher-salience stimuli induce more neural activation, which increases the likelihood that neural activity will reach the threshold level necessary to trigger the widespread neural ignition that is correlated with full conscious perception. When a stimulus is not perceived, the corresponding stimulus-induced activation begins to fade out around 250 ms, but if non-sensory high gamma activity is suppressed following stimulus presentation, the information contained in that activity is made available for subsequent partial or full perception. If this activation is propagated in a forward-sweeping wave of sustained activation that spreads widely throughout the brain, full conscious perception ensues.

Visual Short-Term Memory

Information regarding visual stimuli can be stored in the brain for short period of time following stimulus presentation, known visual short-term memory (VSTM) (Sligte et al., 2008). This store is generally considered to consist of three memory types, known as iconic memory (IM), fragile VSTM (FM), and visual working memory (WM) (Sligte et al., 2008; Vandenbroucke at al., 2011; Pinto et al., 2013). IM has a high-to-unlimited capacity, but lasts less than half a second and depends largely on retinal afterimages, being significantly impaired when stimuli are isoluminant (low-contrast), and instantly erased as soon as any new visual content is presented (Sligte et al., 2008). FM has a lower capacity than IM (10-16 items) and can also be easily overwritten, but only when replaced by a stimulus in the same location that shares features with the current content stored in FM; IM, on the other hand, can be erased by the presentation of a light mask (Sligte et al., 2008; Pinto et al., 2013). In the absence of an interfering stimulus, FM may last up to 4 s, decaying linearly over time (Sligte et al., 2008). WM can remain robust until 5 s post-stimulus offset and is resistant to visual interference, but has a capacity of only 2-5 items (Vandenbroucke et al., 2011).

FM and WM may also differ in their relationship to attention: while FM remains mostly intact when attention is divided, the capacity of WM deteriorates significantly under dual-task conditions and when attention is otherwise preoccupied (Vandenbroucke et al., 2011; Pinto et al., 2017). Other work has indicated that IM and FM are also affected by attention, however, while bottom-up, exogenous attention appears to have similar effects on all three memory stores, endogenously-directed attention impacts WM to a greater extent than FM, and FM more than IM (Botta et al., 2019). As such, while more research is needed to fully understand the differences between these memory stores, it is likely that while attention does have a modulatory effect on all three, it has a greater impact on WM, especially when diverted due to top-down processes.

Multisensory Perception

As noted before, the majority of research on perception has focused on the visual system under the assumption that common mechanisms underlie perception across modalities. Recent work by Sanchez et al. (2019) has taken steps towards confirming this by using decoding analyses to identify common supramodal correlates of conscious perception. By analyzing brain activity during tasks involving near-threshold auditory, visual, and tactile stimuli, they found that certain neural patterns predicted stimulus awareness regardless of the modality; furthermore, these patterns could be observed in primary sensory regions outside of that corresponding to the stimulus in question. Notably, these cross-modal signatures in primary sensory areas were mostly observed in a late (>400 ms) time window following the first period of frontoparietal activation. Long-distance feedback projections through higher cortical areas are known to connect primary sensory areas including V1, the auditory cortex, and multisensory regions (Clavagnier et al., 2004), and the findings of Sanchez et al. (2019) that supramodal markers of conscious perception are most strongly observed after frontoparietal activation support the GNWT hypothesis that conscious perception involves the global broadcasting of information to widespread areas of the brain.

Attention

The relationship between attention and awareness has long been debated, but an accumulating array of evidence has definitively demonstrated that the two, while closely linked, are fundamentally distinct. Not only is attention insufficient for conscious access (Kentridge et al., 2008; Gayet et al., 2018), it also can have strong modulatory effects on unconscious processing (Schmidt & Schmidt, 2010; Bahrami et al., 2007). Neural signatures of selective attention can be observed when task-relevant stimuli are presented even when they are not reported as having been consciously perceived (Koivisto et al., 2005; Chen et al., 2017; Sergent et al., 2005; Travis et al., 2019); furthermore, neural responses to task-irrelevant, invisible stimuli are attenuated during performance of an attentionally-demanding task (Bahrami et al., 2007)

Role of Attention in Modulating the Contents of Awareness

A primary way in which attention interacts with awareness is by increasing perceptual sensitivity. Directing attention to a specific feature, such as color or motion, enhances baseline neural activation in the corresponding brain areas (Chawla et al., 1999) and reduces interneuronal noise correlations (i.e. reducing the effect of single-neuron firing variability on the rest of the population), thereby increasing the differentiation of neural activity within a population of neurons and, accordingly, increasing the information contained in stimulus-driven neural activation (Cohen & Maunsell, 2009). Enhancing activation in a given sensory area increases perceptual sensitivity by bringing baseline activity closer to the threshold level of activation necessary to trigger the neural ignition that has been implicated as being critical for conscious awareness (van Vugt et al. 2018; Herman et al., 2019).

For this reason, divided or withdrawn attention can lead to a failure to perceive even fully visible, salient stimuli, a phenomenon known as ‘inattentional blindness’ (Mack & Rock, 1998; Persuh & Melara, 2016); even when a target stimulus is consciously noticed, features that are task-irrelevant often cannot be reported (Tapal et al., 2019). On the other hand, attention that is diffused or broadly distributed – known as a ‘global attentional scope’ – facilitates the emergence from suppression of images of

faces masked under CFS (Sun et al., 2016). Furthermore, when looking for a CFS-suppressed target stimulus that is defined by a single feature – color, for example – distractor stimuli that share that feature with the target emerge from suppression more quickly; however, when more features of the target are known, this effect disappears (Gayet et al., 2018). While the mechanisms behind these phenomena are not fully understood, they are likely to be driven chiefly by attentional modulations of neural activity in early areas of the visual cortex beginning approximately 200-300 ms post-stimulus onset (Koivisto et al., 2010; Sergent et al., 2005; Travis et al., 2019): activation driven by task-relevant stimuli or features of such stimuli is selected for amplification which, in turn, increases the likelihood that a given stimulus will be consciously perceived (Marti & Dehaene, 2017; van Vugt et al., 2018). Even after a stimulus has disappeared, modulation of neural activity in V1 and V2 driven by top-down attentional processes can enhance conscious perception of a stimulus if it occurs less than 200-300 ms post-offset (Sergent et al., 2011).

In an experimental technique known as rapid serial visual presentation (RSVP), participants are presented with a rapid stream of visual stimuli (e.g. 10 successive stimuli, each presented for 100 ms), typically containing one or more target stimuli that are to be responded to and/or identified. Each stimulus is initially processed in a systematic way, beginning with representational neural activation in V1 corresponding to individual stimuli which ascends through higher areas of the visual cortex for a few 100 ms (Marti & Dehaene, 2017). These neural representations can coexist at different levels of the visual processing hierarchy until about 350 ms, at which point activation corresponding to task-irrelevant stimuli fades out, while activity corresponding to a perceived target stimulus is amplified and sustained (Marti & Dehaene, 2017).

This pattern is likely to account for another well-documented phenomenon known as the attentional blink, in which a second target stimulus (T2) that is presented between approximately 200-400 ms following the onset of a first target (T1) is often not consciously perceived (Sergent et al., 2005). EEG recordings of brain activity have indicated that residual activation corresponding to T1 may overpower T2-induced activation such that it cannot be selected for and amplified sufficiently to be consciously perceived: while early activation (<150 ms) is preserved for both seen and unseen T2s, a

divergence in activation corresponding to seen vs. unseen T2s beginning around 170 ms can be observed (Sergent et al., 2005). While some processing of unseen T2s continues to take place, seen T2s are characterized by a late, sustained, and amplified wave of activation that spreads through distributed cortical areas – the same pattern of activation observed for seen T1s (Sergent et al., 2005).

These extensive findings indicating that top-down attention can directly impact conscious perception underscore the extent to which attention shapes subjective experience and phenomenology, as predicted by AST. Furthermore, the powerful modulatory effect of attentional processes on what content is consciously perceived is consistent with the gateway hypothesis of GNWT, which posits the existence of an early mechanism in the visual cortex that 'gates' the entry of a given content into awareness. Attention, by selectively enhancing perceptual sensitivity to certain content, may act as this gate into awareness, exerting a strong influence on what content one is conscious of at any given moment (i.e. the global workspace, in the GNWT framework).

Attention and Consciousness

Despite the significant evidence pointing to the crucial and potentially causal role that attention may play in determining the contents of awareness, it is still hotly debated whether or not attention is necessary for conscious perception. While evidence from studies on inattentional and change blindness suggest that awareness may be dependent on attention (Simons & Chabris, 1999; Koivisto & Revonsuo, 2008; Rensink et al., 1997), even a high attentional load does not necessarily preclude conscious awareness and the recognition of certain stimuli. When engaged in an attentionally demanding central task, participants were able to accurately perform a second task simultaneously in which they had to respond to natural scenes presented in the periphery if they contained animals or, in a separate experiment, vehicles (Li et al., 2002). Importantly, however, this type of dual-task paradigm fails to ensure that zero attention is allocated to the peripheral task; on the contrary, despite the challenging central task, some residual attention may be directed to surrounding stimuli due to their relevance to the peripheral task.

Evidence from brain-imaging studies is similarly ambiguous. While some research has suggested that subsequent awareness of a stimulus can be predicted by

neural events in the early visual cortex prior to the onset of attentional modulations (Koivisto et al., 2005; Boehler et al., 2008), other work has indicated that this early activity is necessary for both conscious and unconscious perception (Koivisto et al., 2010). Furthermore, observations that awareness can be predicted prior to the onset of attentional effects cannot be taken as definitive evidence that attentional processes are not key to consciousness. Neither correlation nor predictive capacity are the same as causation, and when considering the relationship between attention and awareness, such evidence must be considered in the context of other knowledge about the processes related to conscious perception.

Attentional modulations of perception do not take place solely following a stimulus, but also influence prestimulus neural activity that impacts subsequent conscious perception (Chawla et al., 1999). As such, focusing solely on the effects of attention that are observed after a stimulus is presented fails to take into account the full scope of the impact of attention.

In addition, some rudimentary processing must take place following stimulus presentation before post-stimulus attentional effects can take place; to discriminate between the stimulus-driven neural activation that should be attended to and amplified and the activation that should be ignored, a neural representation of the stimulus must first be present in the brain. Without such a representation, modulatory attentional effects triggered by specific stimulus features could not take place, because no information about the stimulus would be present in the brain.

As reviewed earlier, stimulus-related activation must reach a certain threshold in order for the stimulus to be consciously perceived; whether or not this threshold level of activation is reached depends chiefly on pre-stimulus neural activity and stimulus strength (van Vugt et al., 2018; Weisz et al., 2014). Top-down attention increases pre-stimulus neural activity, and increased stimulus salience not only triggers stronger neural activation but also draws involuntary, bottom-up attention more strongly (Chawla et al., 1999; Botta et al., 2017).

Observations of the capacity of early neural events to predict awareness prior to attentional modulations, therefore, cannot be taken as evidence of the independence of awareness from attention. While the quality of early activation cannot be determined by

late attentional processes, it does depend on both stimulus strength – which impacts attention – and pre-stimulus neural activity, which is itself impacted by attention.

Effects of Awareness

It is well-established that many cognitive functions may be performed unconsciously; nevertheless, conscious awareness has been hypothesized to either facilitate or enable numerous cognitive processes, principally the integration of sensory information; the maintenance and manipulation of content; and attentional control (Dehaene, 2014, pp. 89-114).

Visual and Temporal Binding

A key question in understanding visual perception is the so-called 'binding problem,' or how information about different features from distinct sensory modalities collected across time is integrated to create the perception of a unified experience (Feldman, 2013). Consciousness has been hypothesized to be necessary for visual binding (Revonsuo, 1999), but priming effects that rely on multiple features of a stimulus (e.g. location and orientation) have been observed even when the prime stimulus is not consciously perceived (Keizer et al., 2015). Furthermore, complex scenes including an incongruent object (for example, a person drinking from a hairbrush as opposed to a glass) emerge from suppression faster than those that do not contain an incongruent element (Mudrik et al., 2011), indicating that some processing of these scenes as a whole can occur prior to consciousness. In other instances, however, certain high-level stimulus features do not appear to be processed in the absence of awareness. Implied motion in a visible static image can induce repulsive directional adaptation, in which a subsequent real-motion probe is more likely to be perceived as moving in the opposite direction as the adaptor stimulus; however, this effect does not occur when the same implied motion adaptors are rendered invisible under CFS (Faivre & Koch, 2014a).

Beyond the integration of visual percepts, another aspect of binding is that of temporal integration (Blake & Lee, 2005). This can be observed in the case of apparent

motion, in which two dots that are alternately presented on either side of a rectangle are perceived not as two pairs of flickering dots but instead as one pair of dots moving back and forth across the rectangle if the alternations occur within a certain time period, known as the temporal integration window (Faivre & Koch, 2014b). Apparent motion stimuli of this type can induce directional adaptation effects when visible as well as when they are not consciously perceived; however, the temporal integration window decreases drastically if the apparent motion adaptor is not visible (Faivre & Koch, 2014b). Directional adaptation effects occur during non-conscious apparent motion perception only when the apparent motion adaptors are alternating at least every 100 ms; in contrast, these adaptation effects can occur under visible conditions even when the adaptors only alternate every 1200 ms (Faivre & Koch, 2014b). This suggests that, though temporal integration does take place in the absence of awareness, the time window across which non-perceived events can be integrated is a short one, while temporal integration of percepts that are consciously perceived can be performed across larger intervals.

Content Maintenance and Manipulation

Awareness has been thought to be a necessary prerequisite for the maintenance and manipulation of content in the brain (Mashour et al., 2020); however, recent findings have called this into question, suggesting that both the maintenance and manipulation of neural representations can take place unconsciously to some degree. Decoding analyses of brain activity recorded using magnetoencephalography (MEG) have shown that information regarding unseen stimuli can be stored for up to 800 ms (King et al., 2016). However, only task-relevant information is retained for this period of time; unconscious representations of task-irrelevant stimuli dissipate around 250 ms, consistent with the notion that processes of selective attention take place in the absence of awareness (King et al., 2016). Other work has found that while maintaining content in working memory does not impact the ability to perceive threshold-level stimuli, content manipulation does negatively affect perception (Koivisto et al., 2018).

Research by Trübutschek et al. (2019) investigating working memory has extended this evidence by demonstrating that when asked to perform a forced-choice mental-rotation task with stimuli of varying visibility levels, participants were able to

mentally rotate targets that were reported as unseen at an above-chance level. The MEG data collected during the same experiment, however, indicate that the behavioral results may reflect neural processes that are more nuanced than might be immediately obvious. In the task, a target square was flashed briefly on 80% of the trials in one of 24 locations equidistant from the fixation point, followed by a mask and a 3 s interval before the response screen. 1500 ms into the delay interval, a cue signaled whether the participants were to report the location that the target had appeared or a location 120° clockwise or counterclockwise from the location of the target; participants were to guess if they were unsure, and then rated the subjective visibility of the target. When target stimuli were reported as unseen, neural activity initially resembled the activity observed during stimulus-absent trials, and it was not possible to decode the pre-rotation location of the stimulus. However, around 1000 ms into the delay period, neural activity during unseen trials diverged in anticipation of the cue to resemble the patterns of activation in seen trials, and decoding analyses revealed that a neural representation of the unseen target stimulus was reinstated prior to manipulation.

Trübutschek et al. (2019) suggest that this reflects a process by which neural representations of unseen stimuli are reactivated in preparation for manipulation, and take the resemblance of neural activity in unseen and seen trials as an indicator that the target stimulus must be consciously re-represented in order to be manipulated. This interpretation offers support for GNWT, since it would indicate that content must be reinstated in the global workspace in order to be manipulated. However, their interpretation of the findings hinges on the assumption that the neural activity observed during manipulation in aware trials necessarily reflects conscious processing, which is not clear from the results. Participants were required to report the subjective clarity of their experience of the target, and, in spite of the reinstatement of a neural representation of the target, they still reported the stimulus as unseen. If the observed neural activity truly reflected a conscious process, subjective reports – currently the only method of assessing what was actually experienced by the participant – should reflect this, which was not observed by Trübutschek et al. (2019). Future work might include an additional measure of confidence ratings, which could reflect the extent to which participants felt they were consciously manipulating content.

These findings call into question two principal predictions of GWNT: first, that stimulus-related information must be conscious in order to be maintained over time; and second, that only content that is conscious can be manipulated (unless one fully embraces the interpretation offered by Trübutschek et al. (2019)). On the other hand, the activity-silent maintenance of information offers peripheral support for an important element of IIT, namely that silent neural activity can be as informative as neural firing.

Attentional Control

Stimulus-driven attentional processes can take place in the absence of conscious perception, but stimulus awareness allows for more effective and complex control of attention. Behavioral evidence has indicated that awareness of a task-relevant cue allows for greater stability in attention towards that cue compared to when that cue is not consciously perceived; furthermore, attention to a task-irrelevant stimulus can be effectively suppressed only in conditions of awareness (Webb et al., 2016). Consistent with AST's hypotheses regarding the function of consciousness, the combination of these two effects leads to a marked lack of ability to monitor and modulate the allocation of attention towards multiple stimuli in the absence of conscious perception (Webb et al., 2016).

Aside from facilitating effective attentional modulation, awareness also allows for the use of higher-level information when determining where attention should be directed. Low-level stimulus features such as luminance tend to drive attention more in the absence of conscious perception, even when they are task-irrelevant (Webb et al., 2016). In contrast, symbolic or non-intuitive predictive cues can facilitate target detection in RSVP tasks (Meijs et al., 2018) and lead to faster reaction times (RTs) in speeded-response tasks (Hsu et al., 2011), but only if the cue is consciously perceived; in fact, counterintuitive cues that are not perceived can actually negatively impact RTs (Hsu et al., 2011). Only if one is aware of the cue can information from non-intuitive yet predictive cues be used; otherwise, such cues lack any beneficial effects and may even hinder task performance, as in the case of the counterintuitive predictive cues.

Brain imaging data examining the neural processes associated with selective attention in the presence or absence of awareness are largely consistent with these

behavioral findings. Travis et al. (2019) found that neural activity associated with goal-directed activity differed depending on awareness: ERP signatures of feature enhancement (N_T component at 200-300 ms post-onset) were largely present only when cues were visible, while ERP responses correlated with distractor suppression (P_D component at 350-500 ms) were present at similar magnitudes regardless of cue visibility. This latter ERP result does not directly align with the behavioral findings by Webb et al. (2016) which suggest that distractor suppression is impaired to a greater extent in unaware trials; however, a number of considerations must be taken into account when considering behavioral and neural findings together.

First, the design of the two experiments differed significantly. In addition, Webb et al. (2016) used relative RT comparisons as their primary measure, which reflect the 'end result' of numerous neural processes that take place after stimulus onset, whereas ERP data tracks neural activity at precise moments throughout processing stages. The N_T component found by Travis et al. (2019) occurred at 200-300 ms post-onset and the P_D component at 350-500 ms, the latter of which ended 250 ms before the fastest average RTs from their behavioral findings. Notably, their behavioral results did not indicate differential modulation of suppression and enhancement for aware and unaware trials, despite the neural ERP signatures observed during the task. Expectations that neural and behavioral findings will line up neatly according to previously observed correlations (i.e. the N_T being associated with distractor enhancement and the P_D with suppression) should not be formed without careful consideration of the factors involved.

Finally, it must be reiterated that correlation cannot be taken to be indicative of a causal relationship. Since stimuli must reach a threshold level of activation in order to be consciously perceived (van Vugt et al., 2019), it may be that the amplified N_T component that displays differential changes in magnitude depending on awareness is reflective of the necessary level of activation for a stimulus to be consciously perceived; alternatively – but compatible with this first possibility – the N_T component may relate to the increased perceptual sensitivity that top-down attention can induce in lower sensory regions, whereas the P_D component may reflect other, later processes. Ultimately, relating behavioral and neural findings should be done with careful consideration; the most

effective way to explore these relationships further is to combine brain imaging techniques with behavioral tasks.

Sensory Integration

GNWT predicts that only conscious content is integrated at a high level because the corresponding neural representation must enter the 'workspace' in order to be globally accessible due to the modularity of different sensory regions; it does, however, allow for limited unconscious multimodal interactions (Mashour et al., 2020; Mudrik et al., 2014). Accumulating evidence largely supports this view, though the relationship between consciousness and multisensory perception may be more complex than accounted for in the GWNT framework.

Multimodal integration is characterized by the combination of information from multiple sensory modalities to form a unitary representation of content that is qualitatively different from each individual unisensory percept, and it has been suggested that this process can only take place when the multisensory stimuli are consciously perceived (Mudrik et al., 2014). Some evidence has suggested, however, that a different type of multisensory interaction can be observed in the complete absence of awareness (Ching et al., 2019). A classic example of multimodal integration is the McGurk effect, in which co-presentation of an auditory syllable with a silent video of a different syllable being spoken results in an illusory auditory percept of a third syllable. For example, when an auditory /pa/ is presented with a visual /ka/, the auditory stimulus is often perceived as /ta/, an auditory percept that is qualitatively different from either of the stimulus syllables.

In two studies investigating whether the McGurk effect could be induced by non-perceived stimuli, Ching et al. (2019) failed to observe true nonconscious multimodal integration, but their results indicated that another form of multisensory interaction could take place unconsciously, which the authors refer to as "multimodal alignment." In their proposed framework, a non-perceived stimulus from one sensory modality can influence conscious perception of a stimulus in a different modality if the two stimuli originate from the same source. These findings are ultimately compatible with the GWNT model – true multimodal integration was not observed to occur nonconsciously – but the presence

of early, non-conscious interactions between different sensory modalities indicates that the theory's framework must explicitly allow for cross-modal communication in the absence of consciousness.

In contrast, other research by Scott et al. (2018) has demonstrated that novel, cross-modal associations between auditory and visual stimuli can be formed unconsciously and subsequently induce inverse priming effects when presented supraliminally in a categorization task. However, the task type used in their experimental paradigm was qualitatively different than that of Ching et al. (2019), whose work examined cross-modal alignment and integration of sensory information. The research by Scott et al. (2018), on the other hand, investigated the potential for unconscious associative learning across modalities over time, and the two phenomena are distinct enough that they are likely to be driven by similarly distinct mechanisms. Furthermore, if true multimodal integration necessitates that the multisensory information lead to a qualitatively different perceptual experience, it is unclear that the work by Scott et al. (2018) meets this criteria.

Notably, certain types of conscious multisensory integration can occur even when the multisensory stimuli are consciously experienced as occurring at different times. In another study investigating the McGurk effect, supraliminal McGurk-inducing auditory and visual stimuli were presented asynchronously to participants (Soto-Faraco & Alsius, 2007). The McGurk effect was experienced even when the auditory stimulus was presented up to 240 ms prior to the visual McGurk stimulus, even though participants could reliably report that the auditory stimulus came prior to the onset of the visual stimulus if it was presented at least 160 ms prior to the stimulus, and sometimes as little as 90 ms (Soto-Faraco & Alsius, 2007). While the mechanisms behind this are unclear, these findings suggest that multisensory integration may not be a unitary process that corresponds directly to conscious experience. However, they do support the GNWT hypothesis that representation in the global workspace allows for sophisticated inter-modular communication: the earlier auditory stimulus may be consciously represented prior to the visual stimulus, but it is able to access pre-conscious information in the visual system even before that information is itself conscious. On the other hand, this interaction would likely need to take place prior to the actual representation of the auditory percept,

calling into question the proposal that representation in a ‘global workspace’ that has access to multiple modular structures corresponds to consciousness.

Multisensory neurons (i.e. those that respond to input from more than one modality) can behave in ways that are either ‘convergent,’ responding to input from multiple modalities, or ‘integrative,’ also responding to input from multiple modalities, but doing so differentially according to whether or not the input is unisensory or multisensory (Noel et al., 2019). For example, a convergent neuron might fire in response to both auditory and visual stimulation, but will behave similarly regardless of whether it is receiving auditory, visual, or simultaneous auditory-visual input. An integrative neuron, on the other hand, will respond more strongly to concurrent auditory and visual stimulation than to either one alone. While IIT offers few explicit predictions regarding multimodal integration, it hypothesizes that a system’s integrative capacity – including its ability to integrate multisensory input – should be reflective of conscious level.

Recent work by Noel et al. (2019) investigating the way in which convergent and integrative multisensory neurons behave depending on the level of consciousness has suggested that, contrary to the IIT framework, integrative neurons are less impacted by loss of consciousness than convergent neurons. Single-unit recordings from multisensory neurons in the primary somatosensory cortex (S1) and the ventral premotor cortex (vPM) of macaque monkeys undergoing propofol anesthesia indicated that while only 31% of convergent neurons remained convergent when the monkeys were unconscious, 62.9% of integrative neurons maintained their integrative functioning; the remaining neurons became either unresponsive or responded solely to stimulation from one modality (Noel et al., 2019).

Importantly, loss of consciousness was associated with a decrease in the probability of coactivation of neurons in S1 and vPM, confirming that widespread informational exchange between brain regions is correlated with consciousness, offering clear support for GNWT. On the other hand, the findings call into question the core tenet of IIT: that integration corresponds to consciousness.

This research by Noel et al. (2019) is the first of its kind, however, and has a number of limitations, specifically with respect to evaluating IIT. The authors note that only three nodes were used in the evaluation of the integrative nature of neurons, which is

the most simplified application of IIT (see Figure 1), and – perhaps most importantly – the brain areas they recorded from are not those implicated as being primary to consciousness in the IIT framework, whose proponents emphasize the role of the posterior hot zone (Boly et al., 2017).

Phenomenology

Perceptual mechanisms, neural networks, NCCs, and attentional processes are all crucial areas of investigation in consciousness research. Without a solid understanding of these components, developing a sound understanding of consciousness will be impossible. Nevertheless, the evidence reviewed so far does more to address Dennett's "hard question" than Chalmers' "hard problem" (Dennett, 2018a; Chalmers, 1995). The actual phenomenology of consciousness – the experience of experiencing – feels far more nuanced than whether certain stimuli are perceived or what neural structures support consciousness.

Illusory Perception

Contrary to the subjective phenomenology of experience, what is perceived is seldom a veridical representation of the world. Visual scientists have long known that only a small portion of the visual field – from 1-2° – can be clearly perceived in high resolution at any given time, despite the subjectively rich and seemingly complete nature of visual experience (Feldman, 2013). This is especially striking considering that an area known as the 'blind spot,' a spot on the retina without photoreceptors where the optic nerve is located, takes up approximately 6° of the visual field, but no corresponding gap is perceived during visual experience due to a phenomenon known as 'perceptual filling-in' (Komatsu et al., 2000).

The 'uniformity illusion' is a further demonstration of the power of internally-generated changes in perception: peripheral regions surrounding a central, differently colored or patterned patch are reliably perceived to gradually change to match the color or pattern of the patch after fixating centrally for several seconds (Otten et al., 2017).

When the uniformity illusion is simulated experimentally (i.e. the periphery gradually changes to match the central pattern, as it does in the naturally-occurring illusion), participants are typically equally confident about the veridical nature of the peripheral change regardless of whether it was simulated or naturally occurring (Otten et al., 2017). Certain stimuli can induce a similar phenomenon known as color spreading, in which the center of a 'square' formed by color wedges at the corners appears to change to match the color of the wedges (Shimojo et al., 2001). This illusory color spreading has been observed to induce afterimages – themselves illusory percepts – in what appears to be a widespread cortical process (Shimojo et al., 2001). The mechanisms underlying some other illusory percepts appear to be similarly global in nature and involve higher-order brain areas: activity reflecting phosphene perception induced by occipital TMS is not observed until approximately 160 ms post-TMS application, at which point phosphene-related activation can be observed in widespread brain regions, primarily comprising posterior and central regions (Taylor et al., 2010). An fMRI study by Liu et al. (2019) on the neural activity correlated with this illusion found that the contents of subjective perception (i.e. the illusion) are neurally represented primarily in the PFC, as well as the temporo-parietal junction and other higher-order brain regions, while activity in the early visual cortices most reliably reflects genuine sensory input.

Not all neural representations of illusory percepts, however, lie in these higher-order areas. In the case of blind-spot phenomenology, genuine activation in the region of the primary visual cortex corresponding to the blind spot directly reflects the filling-in process (Komatsu et al., 2000). Another instance of perceptual filling-in, in which vertically-oriented moving gratings of low contrast that are separated by a large gap induce the appearance of the real grating extending across the gap, has likewise been shown to be supported by activation in the early visual cortices (Meng et al., 2005). As such, the mechanisms underlying illusory perception are diverse and cannot be classified unilaterally as either local or global. While direct perceptual filling-in may be a largely bottom-up process originating in the early visual cortices, more 'layered' percepts such as the illusion-induced afterimage phenomenon described above may be based more in higher-order neural processes.

Attention, Expectations, and Predictive Processing

Attention can have a powerful impact on visual phenomenology, affecting the filling-in process, influencing percept emergence, and even allowing for the deliberate modulation of visual experience. However, the precise role that attentional processes play in modulating perception interacts closely with current expectations and uncertainty (Meijs et al., 2018; Lasaponara et al., 2015). Divided attention increases the impact of expectation on subjective experience, enhancing the extent of the blind spot filling-in effect (Lou & Chen, 2002) and even inducing an illusory belief of having seen an absent stimulus directly preceding the time of report, if the stimulus was expected to be present (Aru & Bachmann, 2017). The predictive processing model of perception is based on findings that expectations directly modulate the contents of experience: for example, expecting to see a specific stimulus facilitates the emergence from suppression of visual stimuli masked using CFS (Pinto et al., 2015), and lowers the threshold of conscious perception while speeding the onset of neural signatures of stimulus awareness (Melloni et al., 2011).

In this way, attentional processes, expectations, and predictive processing are closely intertwined and, together, can have a profound impact on subjective phenomenology. Not only can a lack of attention or expectation lead to a failure to perceive a salient stimulus (Persuh & Melara, 2016), it can also do the inverse, leading to a false percept of an absent stimulus. In an experiment by Aru & Bachmann (2017), on each trial participants were presented with a display for 250 ms that consisted of four circles in the corners and a six-letter matrix in the center. On 90% of the trials, they were cued with a tone after stimulus offset to report whether one of the circles were different from the others, and on the other 10% they were cued to enter as many letters from the matrix as they could and rate their visibility. In two critical trials, however, they were cued to perform the circle task, but were instead immediately asked to rate the visibility of the letters as opposed to whether or not one of the circles was different. In one of the critical trials, the letter matrix was present, as in all non-critical trials; in the other trial, however, the matrix was wholly absent. Following the experiment, the participants were asked if they noticed that the letters were absent in any of the trials. Two-thirds of the

subjects reported having noticed the absence of the matrix on the critical trial, and all had given visibility ratings of 1 (“no experience”) on that trial. However, for the remaining third of subjects who failed to notice that the matrix was missing, their visibility ratings on the critical matrix-absent trial averaged 2.7, where 2 = “brief glimpse” and 3 = “almost clear experience” of the stimulus. None of these subjects gave a visibility rating of 1 on the critical trial (unlike the noticers), and their ratings did not differ significantly from the visibility ratings given by the same subset of participants on the matrix-present trials, which averaged 3.0.

The prompt to rate the subjective visibility of the letters came a mere 500 ms post-stimulus offset, but a full one-third of subjects failed to notice the absence of the matrix, and, furthermore, they reported their perception of the absent matrix as being, on average, an “almost clear experience” (Aru & Bachmann, 2017). Both attention and expectations are known to have a greater role in determining perceptual experience when stimulus salience is low or ambiguity is high (Meijs et al., 2018; Melloni et al., 2011; Botta et al., 2017). Importantly, however, in conditions when expectations cannot be formed, the ability to detect difficult-to-perceive stimuli is enhanced, emphasizing the extent to which expectations can ‘overwrite’ natural perception (Lasaponara et al., 2015).

Qualia Space

The identification of one-to-one NCCs that correspond directly to specific perceptual experiences is one of the most significant challenges in consciousness science, one that – if achieved – would represent a great advance for the field. At this time, the most successful and promising work in this area has focused on neural complexity and structural connectivity relating to specific percepts. In grapheme-color synesthesia, written letters and digits are accompanied by an associated color sensation, but not all grapheme-color synesthetes experience this color sensation in the same way: associator synesthetes have a strong, internal, generalized ‘color experience,’ while projector synesthetes perceive the graphemes as being of the synesthetically-associated color. Grapheme-color synesthesia is characterized by activation in the superior parietal lobe, visual area V4, and the letter shape area, and this network displays different patterns of functional connectivity depending upon the type of synesthesia experienced: associative

synesthesia is characterized by a top-down modulatory pathway, while projective synesthesia is characterized by a bottom-up pathway (van Leeuwen et al., 2011). Though the networks involved are the same, differences in functional connectivity fundamentally change the nature of the synesthetic experience.

The functional structure of the brain also corresponds to cognitive tasks, flexibly shifting in a dynamic manner depending on the context. When simultaneously performing two dissociated tasks, the brain can functionally 'split' into decoupled networks corresponding to each task. In a simulated driving paradigm, participants engaged in simulated driving while listening to either GPS instructions (integrated task) or an unrelated radio show (split task) (Sasai et al., 2016). In the integrated task, functional connectivity and integration between an identified 'driving network' and 'listening network' was high, but in the dual task, the two networks decoupled to the extent that the information integration between the two was zero (i.e. each network individually contained as much or more integrated information than the two together).

Changes in the complexity of percepts can also be accompanied by changes in network structure. Random dot stereograms are pairs of images of apparently random dots that, when viewed through a stereoscope, are perceived as 3-D images. Burgess et al. (2003) found that, when participants viewed a random dot stereogram, increases in neural complexity, integration, and differentiation accompanied the perceptual change from a 2-D to a 3-D image. Furthermore, during periods of perceptual transition, integration, complexity, and clustering were reduced, suggesting that brain states accompanying less unitary, clearly defined percepts are fundamentally less structured and informative than those that are stable, a hypothesis consistent with recent findings that subjective stimulus perception is accompanied by an increase in cortical stability (Schurger et al., 2015).

Neural differentiation has also been shown to reflect the subjective meaningfulness of audiovisual stimuli. Mensen et al. (2018) presented participants with 8 different types of audiovisual stimuli with different levels of meaningfulness while recording brain activity with EEG, and asked them to rate how interesting, meaningful, and understandable they found the stimuli to be. The audiovisual stimuli consisted of a novel, unaltered advertisement; five altered or scrambled versions of same advertisement; television noise; and an advertisement that the participants had been habituated to at the

beginning of the experiment. Neural differentiation was significantly lower for the meaningless-clip conditions (e.g. television noise or the scrambled clips) than for any of the others. Furthermore, subjective ratings of meaningfulness, interest, and, most significantly, the number of 'experiences' had, correlated strongly with levels of neural differentiation. In research based specifically on the predictions of IIT, Haun et al. (2017) used intracranial electrodes to assess the relationship between the structure of integrated information patterns in the brain and subjective experience, finding that the patterns corresponded to specific percepts. Elevated activity in the fusiform face area (FFA) is known to correlate with the subjective percept of a face when viewing ambiguous Rubin's vase-face pictures (Figure 7) (Hesselmann et al., 2008), and Haun et al. (2017) found that a specific pattern of integrated information in the FFA was associated with the perception of face stimuli.

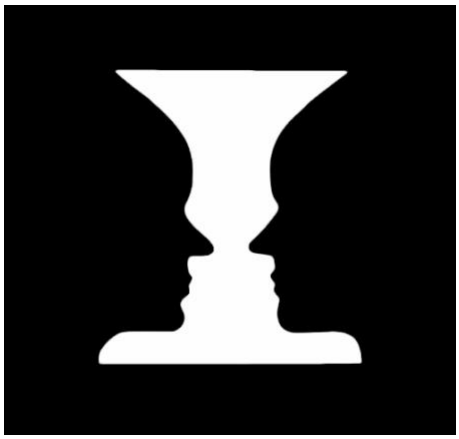

Figure 7: Rubin's vase-face image (Adapted from Wang et al., 2018.)

Discussion

Scientific understanding of consciousness has advanced enormously in the last 25 years. As more and more is known about the mechanisms of perception, the dynamic properties of neural networks, and the bases of phenomenology, myriad directions have been pursued by researchers throughout the field, and numerous theories have been developed centered around different categories of findings. Unfortunately, these theories are often considered in isolation or in opposition to other theories, and little has been done to evaluate empirical evidence for multiple theories, assess which aspects of each are robustly supported and which must be reconsidered, or identify common ground between them. A crucial aspect of making progress in any scientific field is the open and collaborative exchange of ideas among scientists engaged in the area of study, and a willingness among members of the community to adapt the field's paradigm in the face of new evidence – but consciousness science has yet to even adopt an agreed upon paradigm from which to work. Instead, theorists and research groups move forward with research programs primarily centered around testing the specific predictions of their preferred theory, and little effort is made to structure the accumulated wealth of data into a framework that can be shared, explored, expanded, and corrected. Doing so would foster efficient and cooperative advances in the field, but it requires the establishment of a structured, well-supported base of knowledge from which scientists can operate.

The present systematic review sought to address this fundamental issue in the field by surveying a wide array of relevant empirical evidence collected between 1995 and 2019 to evaluate support for three promising theories of consciousness: global neuronal workspace theory (GNWT), integrated information theory (IIT), and attention schema theory (AST). Criteria were established to ensure the quality of the studies included, appropriate search terms were developed, and search queries were run on both Web of Science and Scopus. From the 1618 returned articles that met the established inclusion criteria, 80 were included in the final review along with 26 collected from other sources, for a total of 106 included papers.

Broadly speaking, a key finding that emerged during the review process is that the three theories principally address distinct aspects of consciousness. This is crucial in considering how they may be compatible or conflicting: when a topic is approached from multiple angles, the different views may not align and might appear, at first glance, to be so distinct that they are incompatible; on the contrary, however, a diversity of perspectives can allow for different facets of the same subject to be considered. GNWT, IIT, and AST are all theories of consciousness with empirical grounding, but they each aim to explain a fundamentally different part of the umbrella term 'consciousness.' Recognizing this is critical and underscores the necessity of establishing an agreed-upon definition of consciousness that is inclusive yet specific. Consciousness is a multilayered, complex phenomenon, and the framework offered by one theory will never be accepted by the proponents of another if it seeks to address different questions, because the former will inevitably be inadequate to satisfy proponents of the latter, and vice versa. If the proposed theories are viewed as mutually exclusive competitors, a complete theory will never be developed because no single one provides an explanation for every aspect of consciousness. On the other hand, if they are considered as being potentially compatible, a comprehensive framework that accounts for the dynamic and multifaceted nature of consciousness may be possible to establish.

Assessing the Theories

Consciousness science tends towards three broad areas of investigation: understanding the structures and processes that support basic, primary consciousness; discovering the mechanistic workings behind perceptive and cognitive processes; and accounting for phenomenological experience. With overlap, the areas of interest and predictions of each of the theories pertain chiefly to one of these three areas. IIT addresses how primary consciousness is supported and the necessary requisites; GNWT offers a proposal for how the content of consciousness is determined and the neural events that take place during conscious processing; and AST seeks to provide an explanation for the subjective feel of consciousness. If the strengths and weaknesses of each can be identified and reconciled, the three together may offer a comprehensive

account of consciousness that can be paradigmatically established, facilitating cohesive and collaborative progress in the field.

Integrated Information Theory

The principal testable predictions of IIT are that a balance of differentiation and integration in neural networks constitute the basis of consciousness, and that the functional structure of the network is key to both supporting consciousness and determining the nature of experience. Overall, the evidence in support of these proposals is overwhelming. Loss of consciousness – whether in sleep, disorders of consciousness, or anesthesia – has consistently been observed to be characterized by a disruption in the functional integrity of neural networks. The impressive success of the perturbational complexity index (PCI) in distinguishing between different states of consciousness exemplifies the strength of IIT's mathematical formulation and its success in identifying the network properties crucial to consciousness. Healthy conscious states are characterized by a dynamic pattern of connections and frequent transitioning between patterns of neural activity, and IIT posits that the brain's functional structure reflects a given conscious state.

Specific percepts been found to correspond not only to the structure of integrated information in the brain, but also to specific cognitive states. The findings of Sasai et al. (2016) which demonstrated that in a dual task the functional connectivity of the brain can change such that the integrated information in it splits in two (Figure 8) further support the hypothesis that integrated information reflects states of consciousness and offers a concise explanation for the 'zombie driving' phenomenon. However, it also calls into question what is perhaps the most dubious claim of IIT: that the integrated information contained in the maximally irreducible conceptual structure (MICS) is itself conscious. If this were the case, then during this functional split there would be two consciousnesses in the brain, two MICS. In fact, the authors do tentatively speculate that perhaps there may be two separate 'streams of consciousness' in the brain when this type of functional split occurs, similar to what happens in split-brain patients whose corpus callosum is severed to treat intractable epilepsy. However, each half of the brain is an (approximate) mirror image of the other, so, in split-brain patients, both hemispheres have a full 'set' of parts

and can function normally. The two networks identified by Sasai et al. (2016), on the other hand, consisted of distinct brain regions. Inferring the presence of a second consciousness based on an observation of two transiently split structures of integrated information in the brain strays into the territory of taking IIT's unproven theoretical suppositions above scientific evidence, which has offered no support for this possibility.

A more prudent approach might be to take the presence of two separate integrated structures in the brain as an indicator that integrated information may not be intrinsically conscious itself – a conclusion which, in reality, does little to detract from the integrity of IIT. Ultimately, the claim that integrated information *is* consciousness is not significantly more meaningful than the statement that the firing of C-fibers *is* pain, but letting go of this identity-theory element of IIT in no way compromises the other proposals and predictions of the theory. It may even be the case that the structure of the MICS reflects 'consciousness' at any given point – under the right conditions.

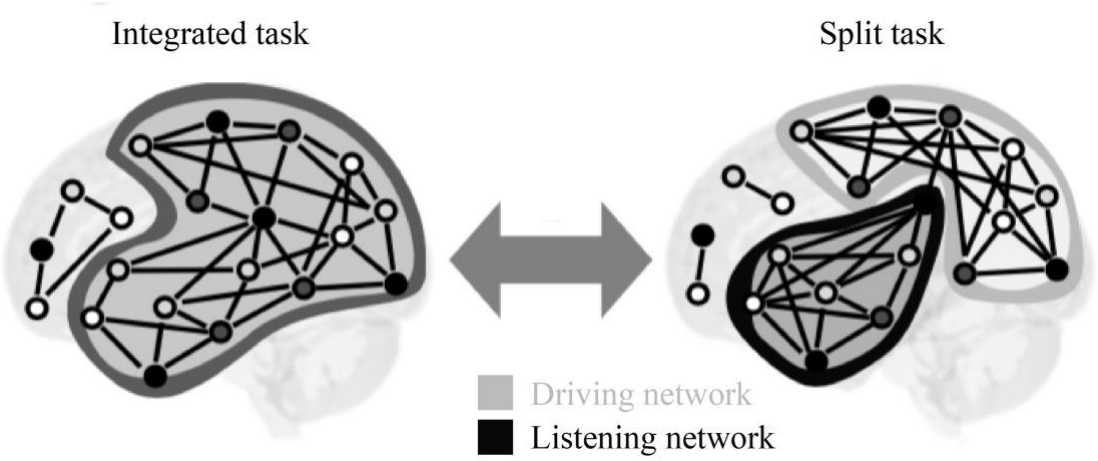

Figure 8: Functional split brain during dual-task performance

A representation of the split in integrated information observed by Sasai et al. (2016) in the driving/listening task. When participants were listening to GPS instructions while completing a driving simulation (integrated task; left), the integrated network remained unified, but when they listened to a radio show while driving (split task; right), the integrated information structure broke down and split into two, one associated with driving and the other with listening. (Adapted from Sasai et al, 2016.)

The study by Sasai et al. (2016) should by no means be taken as definitive in this regard, but it illustrates both a key weakness of IIT – its identity claim – and a key detail about this weakness – it is not actually critical to the applicable substance of the theory. Overall, its tenets are sound: consciousness depends on a balance of integration and differentiation in neural networks; the functional connectivity of the brain is critical to determining the quality of experience; and the structure of integrated information reflects conscious contents in any given moment.

Global Neuronal Workspace Theory

GNWT posits that conscious content is that which exists in a widespread 'global workspace' that enables communication between modular structures in the brain and allows for the maintenance and manipulation of content. Unlike both IIT and AST, GNWT is explicitly not a theory of phenomenal consciousness but rather one of so-called "access consciousness," referring to conscious content that is available for report and cognitive use (Mashour et al., 2020). Correspondingly, its proposals pertain primarily to the mechanisms of perception and the effects of awareness on content representation and manipulation. It makes some predictions regarding the neural conditions that support basic consciousness – namely the presence of long-distance neural connections that allow for widespread communication across the brain – but they are far less specific and accurate than the predictions of IIT, which emphasize both integration and differentiation, dynamic network properties, and network structure. Nevertheless, GNWT's hypotheses about the network properties that allow for the presence of consciousness are fully consistent with those of IIT, but the latter offers a much more comprehensive view of the factors that influence states of consciousness.

With regard to mechanisms of perception and how the content of consciousness is determined, on the other hand, GNWT's specific predictions are – for the most part – fully supported by empirical evidence, and its primary weakness lies in the rigidity of its framework. Neural ignition and the accompanying widespread wave of activation have been repeatedly shown to accompany full conscious perception, and the gateway hypothesis has stood up to scrutiny – with a few caveats. Growing evidence for piecemeal perception indicates that content that does not trigger widespread ignition can

be consciously perceived in a degraded form, which is not accommodated by the current GNWT framework. However, the two-step neural gateway mechanism observed by Levy et al. (2016) can be easily incorporated into the general GNWT model and allows for a more graded view of conscious perception: a first gateway modulates the content that may be available for partial or full perception, and a second gateway requires stimulus-induced activity to reach a threshold level of activation to trigger neural ignition and the corresponding full conscious perception. Importantly, stimuli that are only partially perceived are typically not 'bound;' that is, they are not perceived as a cohesive whole but instead in a fractionated manner, which allows the hypothesis that global ignition is necessary for binding to still stand.

Similarly, true multisensory integration does not appear to occur in the absence of awareness, as proposed by GNWT, but there is evidence that some multimodal interactions may take place in the absence of awareness. Even so, just as the gateway hypothesis and neural ignition model can be adapted to accommodate findings regarding partial perception, it may be possible to incorporate certain mechanisms that allow for unconscious (or, perhaps, partially conscious) multimodal interactions; this may be clarified if the neural processes underlying such interactions can be identified.

One significant challenge to the global workspace framework which may necessitate a more significant renovation of the theory lies in the growing evidence that neural representations of unseen stimuli may be maintained for significant periods of time, and that some manipulation of contents may be possible in the absence of awareness. However, a few considerations regarding these findings should be taken into account when contemplating them in the context of GNWT. First, the prolonged maintenance of information regarding stimuli that are not perceived does more to question GNWT's proposal that consciousness is necessary for the maintenance of information than anything in its actual mechanistic framework. Furthermore, stimuli do need to be consciously perceived to be maintained for more than 1 s at most, and – importantly – only neural representations of task-relevant stimuli are maintained in the absence of awareness, while activation corresponding to irrelevant stimuli fades out by 250 ms. GNWT proposes that selective mechanisms modulate the content that enters awareness, and these mechanisms may extend to the selection of information to be briefly

stored, even if it is not consciously perceived. Especially considering the two-step gateway that may be involved in this process, a similarly multi-tiered process may underlie the selection of unconscious information to be maintained.

With regard to the manipulation of unperceived content, the evidence is still inconclusive. It may be the case that some manipulation of unconscious content is possible; it is unclear whether the work by Trübutschek et al. (2019), for example, which demonstrated that representations of stimuli reported as unseen are neurally reinstated prior to attempted manipulation, indicates that stimuli must be conscious in order to be manipulated, as the authors propose, or that stimuli must be neurally re-represented in what the authors identify as the global workspace but need not be conscious. The former aligns with GNWT, but the latter possibility calls into question the core of the GNWT framework: that content in the global workspace is conscious. On the other hand, it should be considered that representation in the global workspace may not be sufficient for conscious awareness, though it might be necessary for full perception: consciousness is a dynamic and nuanced phenomena which is likely to be based in similarly dynamic and nuanced processes. Ultimately, the precise term 'global workspace' is likely one that should be discarded as a misnomer, a relic from Baars' original global workspace theory that falsely gives the impression of attempting localize consciousness in some 'workspace' region of the brain, a misconception that proponents of GNWT are quick to clarify is not the case (Mashour et al., 2020).

Where GNWT does appear to lack sufficient empirical support is in its emphasis on the frontal areas of the brain, specifically the PFC. The network properties of posterior regions are consistently more reliable than those in frontal areas at indexing states of consciousness, and the first correlates of conscious perception are typically observed around the occipital region. However, with regard to the first, GNWT presents itself as a theory of conscious access, not of conscious level or phenomenal consciousness. Second, considering the previously discussed findings around partial perception, it may be that occipital activity does reliably reflect some degree of conscious perception, while frontal activity corresponds to full, attended perception. As such, GNWT's focus on the PFC may be appropriate considering what it principally seeks to explain: full, explicit access to conscious information. It should also be recognized that GNWT does not posit that

only frontal regions are involved in consciousness and conscious processing, but only emphasizes their role.

The GNWT framework is well-supported, detailed, and offers significant insight into the processes of conscious perception, but it must evolve in a number of ways to accommodate the graded nature of conscious processing, specifically partial perception and a potentially expanded view of the limits of unconscious maintenance and manipulation of information. Furthermore, it remains unclear if representation in the global workspace is sufficient for full conscious perception, even if it is necessary. However, even the most robust, well-supported theories are rarely perfect, even after many iterations, and must change and adapt in the face of new evidence – but that by no means indicates that a theory should be discarded; GNWT is no exception.

Attention Schema Theory

Any theory that attempts to explain phenomenology – to answer the hard problem – is faced with the task of somehow relating neural processes and structures to the rich, subjective feeling of conscious experience. Doing so entails more than just linking patterns of brain activity to perceptual experiences or pinpointing the neural network properties associated with various states of consciousness: it demands the development of a conceptual framework that links what is subjectively experienced by an individual to what can be objectively observed in the brain. AST proposes a model in which cognitive and perceptual processes come together to construct a schematic representation of the brain's process of attention that allows for effective executive control but lacks details about the actual computational processes that are taking place, leading to a sense of the existence of a metaphysical entity – the 'self' – with no concrete connection to the physical brain. The nature of this sort of theoretical framework is such that a wide array of behavioral and neural evidence must first be considered as a whole and meticulously examined to ensure that the theory holds up in the face of scrutiny, after which its precise neurocomputational bases should be investigated, and methods by which the theory might be falsified should be pursued.

As a fairly young theory, AST is in the first phase of verification in which its predictions and explanans should be tested against existing evidence to assess its viability

– and its propositions are proving themselves to be robust. Attentional processes occur in both the presence and absence of consciousness, but awareness allows for effective, complex, and high-level control of attention. Symbolic and counterintuitive cues can be used to successfully direct attention only when they are consciously perceived; otherwise, their low-level and spatial characteristics dominate any influences on attentive processes they may have. Distractors can be ignored and relevant stimuli enhanced with greater efficacy and efficiency under conditions of awareness, allowing for the stable and concentrated direction of attention towards significant content.

Crucially, processes of attention play a critical role in shaping phenomenological experience, a point that has been under-emphasized in most discussions of AST. Attention modulates neural activity to directly impact the likelihood that a stimulus will be perceived, and a lack of attention can lead to a complete failure to notice a stimulus – or the construction of a fabricated percept based on what was expected to be present. The close link between attention and predictive processing further underscores the significance of attention on perceptual experience. The predictive processing model of perception posits that the brain generates a representation of the world based on its expectations, which is then continuously updated and adjusted in response to sensory input (Keller & Mrsic-Flogel, 2018). Attention and expectations interact to influence this model and modulate the contents – the experience – of consciousness in a process that is internally generated, inscrutable from any experiential perspective, and entirely counterintuitive – as is the proposed attention schema.

The foundational tenets of AST are sound and supported by empirical evidence, and the theory offers a grounded explanation of the experience of consciousness itself. Still, a concrete, neurally-based framework for the theory has yet to be proposed or tested; considering the way in which AST and its predictions align with GNWT and IIT may provide some insight as to what such a framework might look like.

A Comprehensive Perspective

None of the theories evaluated in this review offers a complete, satisfactory, and fully substantiated account of consciousness, but each addresses one of the three crucial

aspects of it that must be accounted for in a complete theory: IIT elucidates the network properties necessary consciousness to be present and how phenomenal experience is structured and represented in the brain; GNWT offers a detailed model of the mechanisms of perception and modulatory processes that determine the contents of consciousness; and AST proposes how these neural processes lead to the subjective feeling of being conscious. Outside of a few mostly reconcilable differences, the three are generally consistent with each other, offering complementary accounts of different elements of consciousness that enrich and refine one another.

The state of being conscious and aware of one's surroundings depends on a neural network structure characterized by a careful balance of modularity, global connections, differentiation, and integration. An excess of any one of these properties disrupts the dynamic patterns of activity necessary for the presence of healthy consciousness. A network that is overly integrated has impaired functional specificity and complexity, while one that is overly differentiated lacks the neural pathways crucial to widespread communication between different areas of the brain. Consciousness is characterized by diverse patterns of neural activity and network switching that reflect different modes of functioning and the specificity of conscious experience.

The importance of this process of network switching and dynamic activation becomes clear when the computational processes underlying not only perception but also cognitive states are considered. From the first moment of stimulus processing, highly specific representations of the incoming information are formed in the brain, some of which are selected for further processing and may be consciously perceived, either partially or fully. Not only must these representations reflect specific information regarding the stimulus, but the mechanisms of perception rely on complex patterns of activation throughout the brain. Beginning approximately 100-200 ms post-stimulus onset, neural activity in the early visual areas that is not reflective of the stimulus may be suppressed so that information contained in stimulus-driven activation can be accessed, a process especially important to facilitating conscious piecemeal or full perception under conditions of low visibility. If activity corresponding to the neural representation of a specific stimulus reaches a certain threshold level, a wave of widespread activation that surges forward through the brain is initiated and the stimulus enters awareness, allowing

for high-level, complex, and cross-modal processing. Not all stimuli induce enough activation for the activity to be amplified in this way, however, and their neural representations may fade away entirely, briefly reside in a non-conscious store, or become consciously perceived in a degraded form.

Critical to this process of selection are the constant attentional processes that act as primary gatekeeper modulating the content that is conscious in any given moment. For a stimulus to be fully perceived, it must induce a sufficient level of activation to trigger the neural ignition that allows for full conscious awareness, and ever-changing attentional influences enhance neural activation and sensitivity in specific areas, increasing the likelihood that a stimulus will trigger this wave of activation. Both bottom-up and top-down attention exert these effects specifically and dynamically depending on current goals and external input, and can directly affect the content of which one is aware. Without a network structure allowing for specific, detailed representations and precise control over these critical processes, the informative, unitary, and distinctive quality of experiencing breaks down, and the phenomenon known as consciousness fades away. The properties that allow for this intricate functional connectivity are not only graded in nature but also interact with one another such that different balances of integration, differentiation, modularity, and global connectivity lead to the distinct states of consciousness observed in disorders of consciousness, sleep, and wakefulness.

While significant changes in network structure can correspond to alterations in conscious level, most do not: the functional connectivity in the brain shifts constantly to reflect different percepts and cognitive states. Distinct patterns of integrated information are transiently constructed as representations of percepts, and networks couple and decouple according to the tasks being performed and – crucially – how attention is directed. When performing a dual task, the structure of integrated information can divide accordingly, but in an integrated task, it remains unified, demonstrating the close relationship between attentional processes and neural network properties.

Recognizing the significance of functional connectivity and dynamic brain activation to consciousness elucidates the powerful effects exerted by attention on phenomenology: by impacting the nature of neural activity and inter-neuronal interactions, attentive processes shape perception and cognition, giving rise to a sense of

subjectively experiencing an external (perceived) and internal (cognized) world. This ever-changing world feels rich and vivid, reflecting the complexity of the neurocomputational processes that fabricate it – but not the neurocomputational processes themselves. Accordingly, no amount of introspection can offer insight into the nature of this neurally-based, representational structure, which, accordingly, takes on the inexplicable, ephemeral, yet utterly palpable quality that characterizes the phenomenon known as consciousness.

Future Directions

The framework outlined above offers an empirically supported and internally consistent model for consciousness, but its proposals can only be confirmed, refuted, extended, or modified through focused, purposeful research aimed at understanding not only the properties supporting consciousness, the mechanisms behind perception, and the processes that generate subjective phenomenology, but also how these different aspects of consciousness interact and relate to one another.

The development of such a research program first requires the widespread collaboration of scientists across the field with diverse areas of expertise and an array of perspectives who are willing to openly exchange ideas, constructively disagree, and revise their hypotheses as new evidence emerges. Crucial to this is the establishment of a vocabulary by which to communicate about the topic: the term ‘consciousness’ cannot be used to mean different things by different people at different times. For a word to be used in scientific discourse, it must be unambiguously defined. If a terminology is established to distinguish between different aspects of consciousness (e.g. ‘primary consciousness’ to refer to the basic state of being conscious, ‘access consciousness’ to refer to conscious content, and ‘experiential consciousness’ to refer to the felt, subjective quality of consciousness), misunderstandings can be avoided and concepts clearly presented such that they can be correctly interpreted by those involved in the field without the need for constant clarification about what is meant by critical terms. With a mutually agreed-upon vocabulary in hand, research on consciousness can proceed with greater efficiency and

coordination, allowing for more rapid progress in the field and widespread communication regarding new discoveries and hypotheses.

Moving forwards, much of this progress is likely to come from the use of experimental paradigms that utilize both behavioral measures and neuroimaging methods to provide a more complete view of the neural processes related to conscious experience. While this practice is becoming more common, the behavioral designs of most neuroimaging studies tend to be far less sophisticated than those of studies utilizing behavioral measures exclusively. Due to the intricacies of neural activity and the complexity of analyzing neuroimaging data, the use of simple behavioral paradigms in neuroimaging studies can be a prudent approach that allows for the precise isolation of specific associations between neural and behavioral events. However, as behavioral experiments become more and more elaborate, the resultant findings display more and more nuance as well – but behavioral observations can only go so far in elucidating the nature of consciousness and conscious processing. Even when behavioral and neuroimaging data are collected simultaneously, establishing connections between the two is rarely simple; attempting to draw conclusions by relating multiple types of data from different experiments is far more speculative, especially when the designs differ radically.

As such, future research should employ sophisticated behavioral designs in tandem with brain imaging techniques to explore the neural bases of complex phenomena. An abundance of straightforward neuroimaging studies have been conducted, and just as many complex and informative behavioral studies – more than enough to offer a solid base of knowledge that can be used to guide interpretations of future data. The best approach initially will be to replicate previously conducted behavioral studies that have provided robust findings while simultaneously recording brain activity. Doing so will ensure the quality of the behavioral design and facilitate the accurate interpretation of the neuroimaging data. It is clear that neural activity related to consciousness is both highly complex and tightly linked to each and every subjective experience; the best, most reliable way to learn more about the relationship between the two is to directly compare behavioral and neural data collected concurrently.

A critical line of research to pursue is the further investigation into the ways in which different network structures reflect specific percepts and cognitive states, and the impact of attention on these structures. One of the core claims of IIT is that the structure of the integrated information comprising the purportedly conscious MICS specifies the quality of phenomenal experience at any given moment, but the theory fails to address the proven influence that attentional process have on neural activity and functional network connectivity: the flagship paper of the most recent iteration of IIT contains the word 'attention' just once – in the reference list (Oizumi et al., 2014). A handful of papers authored by Tononi and colleagues briefly mention that attention may exert some influences on the structure of the "physical substrate of consciousness" (i.e. the MICS) (Tononi, 2004; Tononi et al., 2016), but these offhand remarks are a far cry from actually considering the significant impact of attentional processes on phenomenal experience. GNWT and AST, on the other hand – in accordance with the significant empirical evidence – take attention as a primary player in modulating conscious experience.

While it should be acknowledged that IIT explicitly does not focus on cognitive processes, its blatant neglect of the role of attention in conscious experience is ill advised; even the theory's abstract claims regarding "qualia space" and "constellations of concepts" are centered around the functional structure of the brain (Oizumi et al., 2014), which is has been unambiguously demonstrated to be influenced by transient attentional processes. The evidence suggesting that patterns of integrated information in the brain reflect perceptual and cognitive experiences is promising, but a better understanding of the processes involved in forming these structures of integrated information may shed light on how conscious experiences – and, possibly, the experience of consciousness – are composed in the brain.

Another significant area of interest is that of determining the limits of unconscious or partially conscious processing. The challenges involved with investigating this are numerous, including the difficulty of differentiating between stimuli that are genuinely unseen and those that are very faintly perceived, since participants may not respond consistently and it is known that detection and identification tasks can still be performed at above-chance levels in the absence of conscious perception, making classic objective measures like forced-choice response paradigms of little use in distinguishing

null perception from partial perception. Nevertheless, understanding what content can be used and how in the absence of perception, in partial perception, or in full perception will provide key clues about the function of consciousness – or lack thereof. While it is clear that many perceptual and cognitive processes can take place in the absence of awareness, it is far less clear which, if any, require it. Mounting evidence suggests that many processes can take place unconsciously to some extent, including visual binding, multisensory integration, and content manipulation, though the scope of these processes does appear to be limited in the absence of awareness. However, whether consciousness is actually necessary for the high-level performance of these functions or if these limitations may be due to the weakness of neural representations of unnoticed stimuli, for example, remains ambiguous.

The strongest indicator that consciousness may serve a purpose comes from findings relating to attentional control. Consistently, it has been shown that attention is better controlled under conditions of awareness, and that the nuanced direction of attention (e.g. the use of counterintuitive cues) requires conscious awareness. Not only does this align directly with the foundational tenets of AST, but it may offer some insight into the evolution of consciousness and what consciousness might look like in other animals. It may be that few perceptual processes actually require awareness, but that the complex interpretation of content and high-level cognitive control depend on consciousness.

This hypothesis is also intuitively consistent with both GNWT and IIT. Interpreting counterpredictive cues, for example, relies on the ability to access information not only from sensory regions but also cognitive information regarding the significance of the cue, indicating that the widespread exchange of information may be a necessary requisite for this process. In addition, effective utilization of a counterpredictive cue necessarily entails that the information contained in that cue be interpreted in a qualitatively different manner (i.e. integrated), but this is only possible when its counterpredictive nature is explicitly known. If a blue arrow pointing to the left predicts that a target will appear on the left side, but a yellow arrow pointing to the left predicts that a target will appear on the right side, the meaning of those cues is entirely different as a result of the integration of those features. While IIT does not make this

specific prediction – as noted before, the theory omits most discussion of cognitive processes – it is neatly compatible with the general tenets of the theory.

Moving forward, research on consciousness should seek to verify claims made by specific theories, and explore possible connections between their various predictions; doing so will not only allow for the assessment of the validity of different theories, but also facilitate the development of a cohesive model of consciousness.

Conclusion

In the decades since neuroscientific research on consciousness began in earnest, great strides have been made towards understanding the neural basis of consciousness. Nevertheless, consciousness science has been a fractionated field swamped by stigmatized terminology and competitive theorizing, impeding its progression towards becoming a mature science. The copious array of collected data that is currently available, however, indicates that it is now possible for consciousness science to reach a paradigmatic stage and develop into a cohesive, mature scientific field. Integrated information theory, global neuronal workspace theory, and attention schema theory are three empirically supported, promising theories of consciousness that approach the subject from distinct angles. While these theories are often viewed as mutually exclusive competitors, they all demonstrate significant empirical support and, instead of conflicting with one another, offer generally complementary perspectives that fit together to form a multifaceted account of consciousness. A multitude of questions remain unanswered, but if a comprehensive, broadly applicable model of consciousness can be formed, true progress can be made towards solving both the hard problem and the hard question.

Appendix A: Eligibility Criteria - Details & Rationale

Inclusion Criteria

To be eligible for inclusion in the present systematic review, papers must have been:

- Available in English
- Presenting original empirical data: this review is concerned with the analysis of experimental evidence relevant to consciousness science conducted between 1995 and 2019. As such, papers must have reported original research.
- Published between 1995 and 2019: many of the primary theories of consciousness taken seriously in the field today began to take shape towards the end of the 20th century: GNWT was first described in 1998 (Dehaene, Kerszberg, & Changeux, 1998), and the dynamic core theory, the precursor of IIT, was introduced in the same year (Tononi & Edelman, 1998). To allow for the inclusion of research that these theories were based on, studies published up to three years earlier were included. While modern research on consciousness is typically considered to have been undertaken 1990 with the initiation of the search for the neural correlates of consciousness (Crick & Koch, 1990; Koch, Massimini, Boly, & Tononi, 2015), to maintain contemporality, studies between 1990 and 1995 were not included in this review. Since new research builds on previously conducted, replicable work, findings from those five years that have held up to scrutiny should naturally be represented in the results of research conducted in the following decades.
- Subject to peer review: peer-review plays a critical role in assuring the quality of the research presented.

Exclusion Criteria

Papers presenting research meeting any of the following criteria were excluded from this review:

- Pilot studies: pilot studies are performed as precursors to full studies in order to test the validity of their design and look for potential problems and changes to be made. While information derived from them is useful when considering future directions for research, the present systematic review of empirical evidence excluded such studies due to the greater likelihood of inaccuracies in the data collected.
- Retrospective studies: the validity of retrospective studies can be difficult to assess due to multiple risk factors (Tofthagen, 2012). Given the wealth of data available in the field, the inclusion of retrospective data would pose a greater potential threat to the soundness of the systematic review than excluding such data.
- Case reports: the value of case reports should not be underestimated and can provide critical information, but they lack sufficient statistical significance to be included in a systematic review.
- Studies including psychoactive drugs: the inclusion of studies using psychoactive drugs would add additional confounding variables whose effects would be difficult to assess. For this reason, they were excluded from the review.
- Proof-of-concept studies: proof-of-concept studies are often conducted with a desired outcome in mind; this additional risk of bias was considered to be sufficient grounds for exclusion.
- Entirely computer-based studies: entirely virtual models (such as the simulation of evolution) typically rely on pre-existing theories and frameworks, and, as such, can be manipulated to provide the desired results. Models of this sort must be assessed individually and in detail at the programming level to confidently confirm their validity; this was not feasible in the context of the current systematic review.

- Studies with participants with psychopathological conditions connected to consciousness: Studies centered around psychopathological conditions like schizophrenia can be highly informative with regard to certain aspects of consciousness, but they were excluded from the systematic review process due to potential discrepancies in reporting and the heightened possibility of additional unknown confounds including poor physical health and substance abuse (Hjorthøj et al., 2017). It is acknowledged that these problems are not unique to individuals with psychopathological conditions, but the greater probability of skewed results was considered sufficient grounds for exclusion in this review.
- Studies conducted on non-primate species: animal studies are advantageous because it is possible to conduct experiments that cannot be carried out on human participants and allow for the investigation of non-human consciousness, but significant differences in cognitive ability and the capability to self-report make combining data from human and animal studies difficult. Primates display a greater ability for self-report and show evidence of having higher cognitive abilities than most other animals, thus making it easier to relate data collected from studies on other primates to data from those conducted using human participants. In an effort to both maintain accuracy of matching and still take into account a broad range of research, primate studies were included, but research conducted on non-primate animals was not.
- Studies with non-adult participants: aside from the need to control for age, the incomplete brain development in children and adolescents makes it especially complicated to compare neurological research on different age groups.
- Studies which with a gerontological focus: the numerous neurological changes which take place related to ageing can be difficult to take into account when assessing the significance of a given data set; as such, studies focusing on geriatric neurology (i.e. including only elderly participants) were not included in the current review.

Appendix B: Full Index of Articles Included in the Systematic Review

Andersen, L. M., Pedersen, M. N., Sandberg, K., & Overgaard, M. (2016). Occipital MEG Activity in the Early Time Range (< 300 ms) Predicts Graded Changes in Perceptual Consciousness. *Cerebral Cortex*, *26*(6), 2677–2688.

Aru, J., & Bachmann, T. (2017). Expectation creates something out of nothing: The role of attention in iconic memory reconsidered. *Consciousness and Cognition*, *53*, 203–210.

Bahrami, B., Lavie, N., & Rees, G. (2007). Attentional load modulates responses of human primary visual cortex to invisible stimuli. *Current Biology*, *17*(6), 509–513.

Barttfeld, P., Uhrig, L., Sitt, J. D., Sigman, M., Jarraya, B., & Dehaene, S. (2015). Signature of consciousness in the dynamics of resting-state brain activity. *Proceedings of the National Academy of Sciences*, *112*(3), 887–892.

Boehler, C. N., Schoenfeld, M. A., Heinze, H.-J., & Hopf, J.-M. (2008). Rapid recurrent processing. Gates awareness in primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(25), 8742–8747.

Boly, M., Perlbarg, V., Marrelec, G., Schabus, M., Laureys, S., Doyon, J., Pélégrini-Issac, M., Maquet, P., & Benali, H. (2012). Hierarchical clustering of brain activity during human nonrapid eye movement sleep. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(15), 5856–5861.

Botta, F., Martin-Arevalo, E., Lupianez, J., & Bartolomeo, P. (2019). Does spatial attention modulate sensory memory? *Plos One*, *14*(7), e0219504.

Botta, F., Rodenas, E., & Chica, A. B. (2017). Target bottom-up strength determines the extent of attentional modulations on conscious perception. *Experimental Brain Research*, *235*(7), 2109–2124.

Buffalo, E. A., Fries, P., Landman, R., Liang, H., & Desimone, R. (2010). A backward progression of attentional effects in the ventral stream. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(1), 361–365.

Bullier, J., Hupé, J., James, A., & Girard, P. (1996). Functional interactions between areas V1 and V2 in the monkey. *Journal of Physiology-Paris*, *90*(3), 217–220.

Burgess, A. P., Rehman, J., & Williams, J. D. (2003). Changes in neural complexity during the perception of 3D images using random dot stereograms. *International Journal of Psychophysiology*, *48*(1), 35–42.

Casali, A. G., Gosseries, O., Rosanova, M., Boly, M., Sarasso, S., Casali, K. R., Casarotto, S., Bruno, M.-A., Laureys, S., Tononi, G., & Massimini, M. (2013). A theoretically based index of consciousness independent of sensory processing and behavior. *Science Translational Medicine*, *5*(198), 198ra105.

Chawla, D., Rees, G., & Friston, K. J. (1999). The physiological basis of attentional modulation in extrastriate visual areas. *Nature Neuroscience*, *2*(7), 671–676.

Chen, Y., Wang, X., Yu, Y., & Liu, Y. (2017). Dissociable electroencephalograph correlates of visual awareness and feature-based attention. *Frontiers in Neuroscience*, *11*(NOV).

Ching, A. S. M., Kim, J., & Davis, C. (2019). Auditory–visual integration during nonconscious perception. *Cortex*, *117*, 1–15.

Clavagnier, S., Falchier, A., & Kennedy, H. (2004). Long-distance feedback projections to area V1: Implications for multisensory integration, spatial awareness, and visual consciousness. *Cognitive Affective & Behavioral Neuroscience*, *4*(2), 117–126.

Cohen, M. R., & Maunsell, J. H. R. (2009). Attention improves performance primarily by reducing interneuronal correlations. *Nature Neuroscience*, *12*(12), 1594–1600.

Davoodi, R., Moradi, M. H., & Yoonessi, A. (2015). Dissociation Between Attention and Consciousness During a Novel Task: An ERP Study. *Neurophysiology*, *47*(2), 144–154.

de Lange, F. P., van Gaal, S., Lamme, V. A. F., & Dehaene, S. (2011). How awareness changes the relative weights of evidence during human decision-making. *PLoS Biology*, *9*(11).

Del Cul, A., Dehaene, S., Reyes, P., Bravo, E., & Slachevsky, A. (2009). Causal role of prefrontal cortex in the threshold for access to consciousness. *Brain: A Journal of Neurology*, *132*(Pt 9), 2531–2540.

Del Cul, Antoine, Baillet, S., & Dehaene, S. (2007). Brain dynamics underlying the nonlinear threshold for access to consciousness. *Plos Biology*, *5*(10), 2408–2423.

Demertzi, A., Tagliazucchi, E., Dehaene, S., Deco, G., Barttfeld, P., Raimondo, F., Martial, C., Fernández-Espejo, D., Rohaut, B., Voss, H. U., Schiff, N. D., Owen, A. M., Laureys, S., Naccache, L., & Sitt, J. D. (2019). Human consciousness is supported by dynamic complex patterns of brain signal coordination. *Science Advances*, *5*(2), eaat7603.

Faivre, N., & Koch, C. (2014a). Inferring the direction of implied motion depends on visual awareness. *Journal of Vision*, *14*(4).

Faivre, N., & Koch, C. (2014b). Temporal structure coding with and without awareness. *Cognition*, *131*(3), 404–414.

Gayet, S., Douw, I., van der Burg, V., Van der Stigchel, S., & Paffen, C. L. E. (2018). Hide and seek: Directing top-down attention is not sufficient for accelerating conscious access. *Cortex*.

Haun, A. M., Oizumi, M., Kovach, C. K., Kawasaki, H., Oya, H., Howard, M. A., Adolphs, R., & Tsuchiya, N. (2017). Conscious perception as integrated information patterns in human electrocorticography. *ENeuro*, *4*(5).

Herman, W. X., Smith, R. E., Kronemer, S. I., Watsky, R. E., Chen, W. C., Gober, L. M., Touloumes, G. J., Khosla, M., Raja, A., Horien, C. L., Morse, E. C., Botta, K. L., Hirsch, L. J., Alkawadri, R., Gerrard, J. L., Spencer, D. D., & Blumenfeld, H. (2019). A Switch and Wave of Neuronal Activity in the Cerebral Cortex During the First Second of Conscious Perception. *Cerebral Cortex*, *29*(2), 461–474.

Hesselmann, G., Darcy, N., Rothkirch, M., & Sterzer, P. (2018). Investigating masked priming along the "Vision-for-Perception" and "Vision-for-Action" dimensions of unconscious processing. *Journal of Experimental Psychology: General*, *147*(11), 1641–1659.

Hesselmann, G., & Malach, R. (2011). The link between fMRI-BOLD activation and perceptual awareness is stream-invariant in the human visual system. *Cerebral Cortex*, *21*(12), 2829–2837.

Hesselmann, Guido, Kell, C. A., Eger, E., & Kleinschmidt, A. (2008). Spontaneous local variations in ongoing neural activity bias perceptual decisions. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(31), 10984–10989.

Horovitz, S. G., Braun, A. R., Carr, W. S., Picchioni, D., Balkin, T. J., Fukunaga, M., & Duyn, J. H. (2009). Decoupling of the brain's default mode network during deep sleep. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(27), 11376–11381.

Hsu, S.-M., George, N., Wyart, V., & Tallon-Baudry, C. (2011). Voluntary and involuntary spatial attentions interact differently with awareness. *Neuropsychologia*, *49*(9), 2465–2474.

Hurme, M., Koivisto, M., Revonsuo, A., & Railo, H. (2017). Early processing in primary visual cortex is necessary for conscious and unconscious vision while late processing is necessary only for conscious vision in neurologically healthy humans. *NeuroImage*, *150*, 230–238.

Keizer, A. W., Hommel, B., & Lamme, V. A. F. (2015). Consciousness is not necessary for visual feature binding. *Psychonomic Bulletin & Review*, *22*(2), 453–460.

Kentridge, R. W., Nijboer, T. C. W., & Heywood, C. A. (2008). Attended but unseen: Visual attention is not sufficient for visual awareness. *Neuropsychologia*, *46*(3), 864–869.

King, Jean-Remi, Sitt, J. D., Faugeras, F., Rohaut, B., El Karoui, I., Cohen, L., Naccache, L., & Dehaene, S. (2013). Information Sharing in the Brain Indexes

Consciousness in Noncommunicative Patients. *Current Biology*, *23*(19), 1914–1919.

King, J.-R., & Dehaene, S. (2014). A model of subjective report and objective discrimination as categorical decisions in a vast representational space. *Philosophical Transactions of the Royal Society B-Biological Sciences*, *369*(1641), 20130204.

King, J.-R., Pescetelli, N., & Dehaene, S. (2016). Brain Mechanisms Underlying the Brief Maintenance of Seen and Unseen Sensory Information. *Neuron*, *92*(5), 1122–1134.

Koivisto, M., Revonsuo, A., & Salminen, N. (2005). Independence of visual awareness from attention at early processing stages. *NeuroReport*, *16*(8), 817–821.

Koivisto, M., Ruohola, M., Vahtera, A., Lehmusvuo, T., & Intaite, M. (2018). The effects of working memory load on visual awareness and its electrophysiological correlates. *Neuropsychologia*, *120*, 86–96.

Koivisto, Mika, & Grassini, S. (2016). Neural processing around 200ms after stimulus-onset correlates with subjective visual awareness. *Neuropsychologia*, *84*, 235–243.

Koivisto, Mika, Mäntylä, T., & Silvanto, J. (2010). The role of early visual cortex (V1/V2) in conscious and unconscious visual perception. *NeuroImage*, *51*(2), 828–834.

Koivisto, Mika, & Revonsuo, A. (2008). The role of unattended distractors in sustained inattentional blindness. *Psychological Research-Psychologische Forschung*, *72*(1), 39–48.

Komatsu, H., Kinoshita, M., & Murakami, I. (2000). Neural Responses in the Retinotopic Representation of the Blind Spot in the Macaque V1 to Stimuli for Perceptual Filling-In. *Journal of Neuroscience*, *20*(24), 9310–9319.

Lambert, A. J., Wilkie, J., Greenwood, A., Ryckman, N., Sciberras-Lim, E., Booker, L.-J., & Tahara-Eckl, L. (2018). Towards a unified model of vision and attention: Effects of visual landmarks and identity cues on covert and overt attention

movements. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(3), 412–432.

Lasaponara, S., Dragone, A., Lecce, F., Di Russo, F., & Doricchi, F. (2015). The "serendipitous brain": Low expectancy and timing uncertainty of conscious events improve awareness of unconscious ones (evidence from the Attentional Blink). *Cortex*, *71*, 15–33.

Laureys, S., Goldman, S., Phillips, C., Van Bogaert, P., Aerts, J., Luxen, A., Franck, G., & Maquet, P. (1999). Impaired effective cortical connectivity in vegetative state: Preliminary investigation using PET. *NeuroImage*, *9*(4), 377–382.

Lee, M., Baird, B., Gosseries, O., Nieminen, J. O., Boly, M., Postle, B. R., Tononi, G., & Lee, S.-W. (2019). Connectivity differences between consciousness and unconsciousness in non-rapid eye movement sleep: A TMS-EEG study. *Scientific Reports*, *9*, 5175.

Levy, J., Vidal, J. R., Fries, P., Demonet, J.-F., & Goldstein, A. (2016). Selective Neural Synchrony Suppression as a Forward Gatekeeper to Piecemeal Conscious Perception. *Cerebral Cortex*, *26*(7), 3010–3022.

Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences*, *99*(14), 9596–9601.

Liu, S., Yu, Q., Tse, P. U., & Cavanagh, P. (2019). Neural correlates of the conscious perception of visual location lie outside visual cortex. *Current Biology*, 660597.

Lou, L., & Chen, J. (2003). Attention and blind-spot phenomenology. *Psyche*, *9*.

Ludwig, K., Sterzer, P., Kathmann, N., & Hesselmann, G. (2016). Differential modulation of visual object processing in dorsal and ventral stream by stimulus visibility. *Cortex*, *83*, 113–123.

Mäki-Marttunen, V., Castro, M., Olmos, L., Leiguarda, R., & Villarreal, M. (2016). Modulation of the default-mode network and the attentional network by self-referential processes in patients with disorder of consciousness. *Neuropsychologia*, *82*, 149–160.

Marti, S., & Dehaene, S. (2017). Discrete and continuous mechanisms of temporal selection in rapid visual streams. *Nature Communications*, *8*.

Meijs, E. L., Slagter, H. A., de Lange, F. P., & van Gaal, S. (2018). Dynamic interactions between top–down expectations and conscious awareness. *Journal of Neuroscience*, *38*(9), 2318–2327.

Melloni, L., Schwiedrzik, C. M., Müller, N., Rodriguez, E., & Singer, W. (2011). Expectations change the signatures and timing of electrophysiological correlates of perceptual awareness. *Journal of Neuroscience*, *31*(4), 1386–1396.

Meng, M., Remus, D. A., & Tong, F. (2005). Filling-in of visual phantoms in the human brain. *Nature Neuroscience*, *8*(9), 1248–1254.

Mensen, A., Marshall, W., Sasai, S., & Tononi, G. (2018). Differentiation analysis of continuous electroencephalographic activity triggered by video clip contents. *Journal of Cognitive Neuroscience*, *30*(8), 1108–1118.

Monti, M. M., Lutkenhoff, E. S., Rubinov, M., Boveroux, P., Vanhaudenhuyse, A., Gosseries, O., Bruno, M.-A., Noirhomme, Q., Boly, M., & Laureys, S. (2013). Dynamic Change of Global and Local Information Processing in Propofol-Induced Loss and Recovery of Consciousness. *PLoS Computational Biology*, *9*(10).

Mudrik, L., Breska, A., Lamy, D., & Deouell, L. Y. (2011). Integration without awareness: Expanding the limits of unconscious processing. *Psychological Science*, *22*(6), 764–770.

Mudrik, L., Faivre, N., & Koch, C. (2014). Information integration without awareness. *Trends in Cognitive Sciences*, *18*(9), 488–496.

Noel, J.-P., Ishizawa, Y., Patel, S. R., Eskandar, E. N., & Wallace, M. T. (2019). Leveraging Nonhuman Primate Multisensory Neurons and Circuits in Assessing Consciousness Theory. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *39*(38), 7485–7500.

Otten, M., Pinto, Y., Paffen, C. L. E., Seth, A. K., & Kanai, R. (2017). The Uniformity Illusion: Central Stimuli Can Determine Peripheral Perception. *Psychological Science*, *28*(1), 56–68.

Overgaard, M., Nielsen, J. F., & Fuglsang-Frederiksen, A. (2004). A TMS study of the ventral projections from V1 with implications for the finding of neural correlates of consciousness. *Brain and Cognition*, *54*(1), 58–64.

Persuh, M., & Melara, R. D. (2016). Barack Obama Blindness (BOB): Absence of Visual Awareness to a Single Object. *Frontiers in Human Neuroscience*, *10*, 118.

Pins, D., & Ffytche, D. (2003). The neural correlates of conscious vision. *Cerebral Cortex (New York, N.Y.: 1991)*, *13*(5), 461–474.

Pinto, Y., van Gaal, S., de Lange, F. P., Lamme, V. A. F., & Seth, A. K. (2015). Expectations accelerate entry of visual stimuli into awareness. *Journal of Vision*, *15*(8).

Pinto, Y., Vandenbroucke, A. R., Otten, M., Sligte, I. G., Seth, A. K., & Lamme, V. A. F. (2017). Conscious visual memory with minimal attention. *Journal of Experimental Psychology: General*, *146*(2), 214–226.

Pinto, Yaïr, Sligte, I. G., Shapiro, K. L., & Lamme, V. A. F. (2013). Fragile visual short-term memory is an object-based and location-specific store. *Psychonomic Bulletin & Review*, *20*(4), 732–739.

Pitts, M.A., Padwal, J., Fennelly, D., Martínez, A., & Hillyard, S. A. (2014). Gamma band activity and the P3 reflect post-perceptual processes, not visual awareness. *NeuroImage*, *101*, 337–350.

Pitts, Michael A., Martínez, A., & Hillyard, S. A. (2012). Visual Processing of Contour Patterns under Conditions of Inattentional Blindness. *Journal of Cognitive Neuroscience*, *24*(2), 287–303.

Rensink, R., O'Regan, J., & Clark, J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, *8*, 368–373.

Sanchez, G., Hartmann, T., Fuscà, M., Demarchi, G., & Weisz, N. (2019). Decoding across sensory modalities reveals common supramodal signatures of conscious perception. *BioRxiv*, 115535.

Sasai, S., Boly, M., Mensen, A., & Tononi, G. (2016). Functional split brain in a driving/listening paradigm. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(50), 14444–14449.

Schartner, M., Seth, A., Noirhomme, Q., Boly, M., Bruno, M.-A., Laureys, S., & Barrett, A. (2015). Complexity of Multi-Dimensional Spontaneous EEG Decreases during Propofol Induced General Anaesthesia. *Plos One*, *10*(8), e0133532.

Schmidt, F., & Schmidt, T. (2010). Feature-based attention to unconscious shapes and colors. *Attention, Perception & Psychophysics*, *72*(6), 1480–1494.

Schrouff, J., Perlbarg, V., Boly, M., Marrelec, G., Boveroux, P., Vanhaudenhuyse, A., Bruno, M.-A., Laureys, S., Phillips, C., Pélégrini-Issac, M., Maquet, P., & Benali, H. (2011). Brain functional integration decreases during propofol-induced loss of consciousness. *NeuroImage*, *57*(1), 198–205.

Schurger, A., Sarigiannidis, I., Naccache, L., Sitt, J. D., & Dehaene, S. (2015). Cortical activity is more stable when sensory stimuli are consciously perceived. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(16), E2083–E2092.

Scott, R. B., Samaha, J., Chrisley, R., & Dienes, Z. (2018). Prevailing theories of consciousness are challenged by novel cross-modal associations acquired between subliminal stimuli. *Cognition*, *175*, 169–185.

Sergent, C., Baillet, S., & Dehaene, S. (2005). Timing of the brain events underlying access to consciousness during the attentional blink. *Nature Neuroscience*, *8*(10), 1391–1400.

Sergent, Claire, Ruff, C. C., Barbot, A., Driver, J., & Rees, G. (2011). Top-Down Modulation of Human Early Visual Cortex after Stimulus Offset Supports Successful Postcued Report. *Journal of Cognitive Neuroscience*, *23*(8), 1921–1934.

Shimojo, S., Kamitani, Y., & Nishida, S. (2001). Afterimage of Perceptually Filled-in Surface. *Science*, *293*(5535), 1677–1680.

Silva, S., Alacoque, X., Fourcade, O., Samii, K., Marque, P., Woods, R., Mazziotta, J., Chollet, F., & Loubinoux, I. (2010). Wakefulness and loss of awareness: Brain and brainstem interaction in the vegetative state. *Neurology*, *74*(4), 313–320.

Simons, D. J., & Chabris, C. F. (1999). Gorillas in Our Midst: Sustained Inattentional Blindness for Dynamic Events. *Perception*, *28*(9), 1059–1074.

Sligte, I. G., Scholte, H. S., & Lamme, V. A. F. (2008). Are There Multiple Visual Short-Term Memory Stores? *PLOS ONE*, *3*(2), e1699.

Soto-Faraco, S., & Alsius, A. (2007). Conscious access to the unisensory components of a cross-modal illusion. *NeuroReport*, *18*(4), 347–350.

Sun, S. Z., Cant, J. S., & Ferber, S. (2016). A global attentional scope setting prioritizes faces for conscious detection. *Journal of Vision*, *16*(6).

Tacikowski, P., Berger, C. C., & Ehrsson, H. H. (2017). Dissociating the Neural Basis of Conceptual Self-Awareness from Perceptual Awareness and Unaware Self-Processing. *Cerebral Cortex*, *27*(7), 3768–3781.

Tagliazucchi, E., von Wegner, F., Morzelewski, A., Brodbeck, V., Jahnke, K., & Laufs, H. (2013). Breakdown of long-range temporal dependence in default mode and attention networks during deep sleep. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(38), 15419–15424.

Tapal, A., Yeshurun, Y., & Eitam, B. (2019). Relevance-based processing: Little role for task-relevant expectations. *Psychonomic Bulletin & Review*, *26*(4), 1426–1432.

Taylor, P. C. J., Walsh, V., & Eimer, M. (2010). The neural signature of phosphene perception. *Human Brain Mapping*, *31*(9), 1408–1417.

Thakral, P. P. (2011). The neural substrates associated with inattentional blindness. *Consciousness and Cognition*, *20*(4), 1768–1775.

Travis, S. L., Dux, P. E., & Mattingley, J. B. (2019). Neural correlates of goal-directed enhancement and suppression of visual stimuli in the absence of conscious perception. *Attention, Perception, and Psychophysics*, *81*(5), 1346–1364.

Trübutschek, D., Marti, S., Ueberschar, H., & Dehaene, S. (2019). Probing the limits of activity-silent non-conscious working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(28), 14358–14367.

Uehara, T., Yamasaki, T., Okamoto, T., Koike, T., Kan, S., Miyauchi, S., Kira, J.-I., & Tobimatsu, S. (2014). Efficiency of a small-world brain network depends on consciousness level: A resting-state fMRI study. *Cerebral Cortex*, *24*(6), 1529–1539.

van Leeuwen, T. M., den Ouden, H. E. M., & Hagoort, P. (2011). Effective connectivity determines the nature of subjective experience in grapheme-color synesthesia. *Journal of Neuroscience*, *31*(27), 9879–9884.

van Vugt, B., Dagnino, B., Vartak, D., Safaai, H., Panzeri, S., Dehaene, S., & Roelfsema, P. R. (2018). The threshold for conscious report: Signal loss and response bias in visual and frontal cortex. *Science (New York, N.Y.)*, *360*(6388), 537–542.

Vandenbroucke, A. R. E., Sligte, I. G., & Lamme, V. A. F. (2011). Manipulations of attention dissociate fragile visual short-term memory from visual working memory. *Neuropsychologia*, *49*(6), 1559–1568.

Vanhaudenhuyse, A., Demertzi, A., Schabus, M., Noirhomme, Q., Bredart, S., Boly, M., Phillips, C., Soddu, A., Luxen, A., Moonen, G., & Laureys, S. (2011). Two Distinct Neuronal Networks Mediate the Awareness of Environment and of Self. *Journal of Cognitive Neuroscience*, *23*(3), 570–578.

Webb, T. W., Kean, H. H., & Graziano, M. S. A. (2016). Effects of Awareness on the Control of Attention. *Journal of Cognitive Neuroscience*, *28*(6), 842–851.

Weisz, N., Wuehle, A., Monittola, G., Demarchi, G., Frey, J., Popov, T., & Braun, C. (2014). Prestimulus oscillatory power and connectivity patterns predispose conscious somatosensory perception. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(4), E417–E425.

Ye, M., Lyu, Y., Sclodnick, B., & Sun, H.-J. (2019). The P3 reflects awareness and can be modulated by confidence. *Frontiers in Neuroscience, 13*.

Zadbood, A., Lee, S.-H., & Blake, R. (2011). Stimulus fractionation by interocular suppression. *Frontiers in Human Neuroscience*, *5*, 135.

Zhan, M., Goebel, R., & de Gelder, B. (2018). Ventral and dorsal pathways relate differently to visual awareness of body postures under continuous flash suppression. *ENeuro*, *5*(1).

Zhang, W., & Luck, S. J. (2009). Feature-based attention modulates feedforward visual processing. *Nature Neuroscience*, *12*(1), 24–25.

Zhou, J., Liu, X., Song, W., Yang, Y., Zhao, Z., Ling, F., Hudetz, A. G., & Li, S.-J. (2011). Specific and nonspecific thalamocortical functional connectivity in normal and vegetative states. *Consciousness and Cognition*, *20*(2), 257–268.

References

Andersen, L. M., Pedersen, M. N., Sandberg, K., & Overgaard, M. (2016). Occipital MEG Activity in the Early Time Range (< 300 ms) Predicts Graded Changes in Perceptual Consciousness. *Cerebral Cortex*, *26*(6), 2677–2688.

Aru, J., & Bachmann, T. (2017). Expectation creates something out of nothing: The role of attention in iconic memory reconsidered. *Consciousness and Cognition*, *53*, 203–210.

Baars, B. J. (1988). *A cognitive theory of consciousness*. Cambridge [England]; New York: Cambridge University Press.

Baars, B. J. (1997). *In the theater of consciousness: The workspace of the mind*.

Baars, B. J. (2002). The conscious access hypothesis: Origins and recent evidence. *Trends in Cognitive Sciences*, *6*(1), 47–52.

Baars, B. J. (2005). Global workspace theory of consciousness: Toward a cognitive neuroscience of human experience. In S. Laureys (Ed.), *Progress in Brain Research*, 45–53.

Baars, B. J. (2017). The Global Workspace Theory of Consciousness. In *The Blackwell Companion to Consciousness*, 227–242.

Baars, B. J., Franklin, S., & Ramsøy, T. Z. (2013). Global Workspace Dynamics: Cortical "Binding and Propagation" Enables Conscious Contents. *Frontiers in Psychology*, *4*.

Bahrami, B., Lavie, N., & Rees, G. (2007). Attentional load modulates responses of human primary visual cortex to invisible stimuli. *Current Biology*, *17*(6), 509–513.

Barttfeld, P., Uhrig, L., Sitt, J. D., Sigman, M., Jarraya, B., & Dehaene, S. (2015). Signature of consciousness in the dynamics of resting-state brain activity. *Proceedings of the National Academy of Sciences*, *112*(3), 887–892.

Beck, D. M., & Kastner, S. (2009). Top-down and bottom-up mechanisms in biasing competition in the human brain. *Vision Research*, *49*(10), 1154–1165.

Bernstein, R. J. (2000). "The challenge of scientific materialism." In Rosenthal, D. M. (Ed.). *Materialism and the mind-body problem* (2nd ed.). Indianapolis: Hackett PubCo. (Original work published 1968).

Blackmore, S. J. (1999). *The meme machine*. Oxford [England] ; New York: Oxford University Press.

Blackmore, S., & Troscianko, E.. (2018). *Consciousness* (3rd ed.). Abingdon, Oxon ; New York, NY: Routledge.

Blake, R., & Lee, S.-H. (2005). The Role of Temporal Structure in Human Vision. *Behavioral and Cognitive Neuroscience Reviews*, *4*(1), 21–42.

Boehler, C. N., Schoenfeld, M. A., Heinze, H.-J., & Hopf, J.-M. (2008). Rapid recurrent processing. Gates awareness in primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(25), 8742–8747.

Boly, M., Massimini, M., Tsuchiya, N., Postle, B. R., Koch, C., & Tononi, G. (2017). Are the Neural Correlates of Consciousness in the Front or in the Back of the Cerebral Cortex? Clinical and Neuroimaging Evidence. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *37*(40), 9603–9613.

Boly, M., Perlbarg, V., Marrelec, G., Schabus, M., Laureys, S., Doyon, J., Pélégrini-Issac, M., Maquet, P., & Benali, H. (2012). Hierarchical clustering of brain activity during human nonrapid eye movement sleep. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(15), 5856–5861.

Boly, M., Seth, A. K., Wilke, M., Ingmundson, P., Baars, B., Laureys, S., Edelman, D., & Tsuchiya, N. (2013). Consciousness in humans and non-human animals: Recent advances and future directions. *Frontiers in Psychology*, *4*.

Botta, F., Martin-Arevalo, E., Lupianez, J., & Bartolomeo, P. (2019). Does spatial attention modulate sensory memory? *Plos One*, *14*(7), e0219504.

Botta, F., Rodenas, E., & Chica, A. B. (2017). Target bottom-up strength determines the extent of attentional modulations on conscious perception. *Experimental Brain Research*, *235*(7), 2109–2124.

Buffalo, E. A., Fries, P., Landman, R., Liang, H., & Desimone, R. (2010). A backward progression of attentional effects in the ventral stream. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(1), 361–365.

Bullier, J., Hupé, J., James, A., & Girard, P. (1996). Functional interactions between areas V1 and V2 in the monkey. *Journal of Physiology-Paris*, *90*(3), 217–220.

Bullock, T. H., McClune, M. C., Achimowicz, J. Z., Iragui-Madoz, V. J., Duckrow, R. B., & Spencer, S. S. (1995). Temporal fluctuations in coherence of brain waves. *Proceedings of the National Academy of Sciences*, *92*(25), 11568–11572.

Burgess, A. P., Rehman, J., & Williams, J. D. (2003). Changes in neural complexity during the perception of 3D images using random dot stereograms. *International Journal of Psychophysiology*, *48*(1), 35–42.

Casali, A. G., Gosseries, O., Rosanova, M., Boly, M., Sarasso, S., Casali, K. R., Casarotto, S., Bruno, M.-A., Laureys, S., Tononi, G., & Massimini, M. (2013). A theoretically based index of consciousness independent of sensory processing and behavior. *Science Translational Medicine*, *5*(198), 198ra105.

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, *2*(3), 200–19.

Chalmers, D. J. (2018). The Meta-Problem of Consciousness. *Journal of Consciousness Studies*, *25*(9–10), 6-61.

Chawla, D., Rees, G., & Friston, K. J. (1999). The physiological basis of attentional modulation in extrastriate visual areas. *Nature Neuroscience*, *2*(7), 671–676.

Chen, Y., Wang, X., Yu, Y., & Liu, Y. (2017). Dissociable electroencephalograph correlates of visual awareness and feature-based attention. *Frontiers in Neuroscience*, *11*(NOV).

Ching, A. S. M., Kim, J., & Davis, C. (2019). Auditory–visual integration during nonconscious perception. *Cortex*, *117*, 1–15.

Clavagnier, S., Falchier, A., & Kennedy, H. (2004). Long-distance feedback projections to area V1: Implications for multisensory integration, spatial awareness, and visual consciousness. *Cognitive Affective & Behavioral Neuroscience*, *4*(2), 117–126.

Cohen, M. R., & Maunsell, J. H. R. (2009). Attention improves performance primarily by reducing interneuronal correlations. *Nature Neuroscience*, *12*(12), 1594–1600.

Crick, F., & Koch, C. (1990). Towards a neurobiological theory of consciousness. *Seminars in the Neurosciences*, *2*, 263–275.

Crick, F., & Koch, C. (1992). The Problem of Consciousness. *Scientific American*, *267*(3), 152–159.

Crick, F., & Koch, C. (1998). Consciousness and neuroscience. *Cerebral Cortex*, *8*(2), 97–107.

Crick, F., & Koch, C. (2003). A framework for consciousness. *Nature Neuroscience*, *6*(2), 119–126.

Damasio, A. (1999). *The feeling of what happens: Body and emotion in the making of consciousness* (First edition.). New York: Harcourt Brace, Harcourt Inc.

Damasio, A. (2010). *Self comes to mind: Constructing the conscious brain* (1st ed.). New York: Pantheon Books.

Damasio, A. (2018). *The strange order of things: Life, feeling, and the making of cultures* (First edition.). New York: Pantheon Books.

Davoodi, R., Moradi, M. H., & Yoonessi, A. (2015). Dissociation Between Attention and Consciousness During a Novel Task: An ERP Study. *Neurophysiology*, *47*(2), 144–154.

de Lange, F. P., van Gaal, S., Lamme, V. A. F., & Dehaene, S. (2011). How awareness changes the relative weights of evidence during human decision-making. *PLoS Biology*, *9*(11).

Deco, G., Jirsa, V. K., & McIntosh, A. R. (2011). Emerging concepts for the dynamical organization of resting-state activity in the brain. *Nature Reviews Neuroscience*, *12*(1), 43–56.

Dehaene, S. (2014). *Consciousness and the brain: Deciphering how the brain codes our thoughts*. New York, New York: Viking.

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

Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition*, *79*(1), 1–37.

Dehaene, S., Kerszberg, M., & Changeux, J.-P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(24), 14529–14534.

Dehaene, S., Lau, H., & Kouider, S. (2017). What is consciousness, and could machines have it? *Science*, *358*(6362), 486–492.

Dehaene, S., Naccache, L., Cohen, L., Bihan, D. L., Mangin, J.-F., Poline, J.-B., & Rivière, D. (2001). Cerebral mechanisms of word masking and unconscious repetition priming. *Nature Neuroscience*, *4*(7), 752–758.

Del Cul, A., Baillet, S., & Dehaene, S. (2007). Brain dynamics underlying the nonlinear threshold for access to consciousness. *PLoS Biology*, *5*(10), e260.

Del Cul, A., Dehaene, S., Reyes, P., Bravo, E., & Slachevsky, A. (2009). Causal role of prefrontal cortex in the threshold for access to consciousness. *Brain: A Journal of Neurology*, *132*(Pt 9), 2531–2540.

Demertzi, A., Tagliazucchi, E., Dehaene, S., Deco, G., Barttfeld, P., Raimondo, F., Martial, C., Fernández-Espejo, D., Rohaut, B., Voss, H. U., Schiff, N. D., Owen, A. M., Laureys, S., Naccache, L., & Sitt, J. D. (2019). Human consciousness is supported by dynamic complex patterns of brain signal coordination. *Science Advances*, *5*(2), eaat7603.

Dennett, D. C. (2016). Illusionism as the Obvious Default Theory of Consciousness. *Journal of Consciousness Studies*, *23*(11–12), 64–72.

Dennett, D. C. (2018a). Facing up to the hard question of consciousness. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1755), 20170342.

Dennett, D. C. (2018b). *From bacteria to Bach and back: The evolution of minds*. New York: W W Norton & Company.

Dennett, D. C. D. C. (1991). *Consciousness explained* (1st ed.). Boston: Little, Brown and Co.

Descartes, R. (1998). *Discourse on method; and: Meditations on first philosophy* (Fourth edition.). Indianapolis: Hackett Pub. (Original work published 1641).

Desimone, R., & Duncan, J. (1995). Neural Mechanisms of Selective Visual Attention. *Annual Review of Neuroscience*, *18*(1), 193–222.

Edelman, G. M., & Tononi, G. (2000). *A universe of consciousness: How matter becomes imagination* (1st ed.). New York, NY: Basic Books.

Faivre, N., & Koch, C. (2014a). Inferring the direction of implied motion depends on visual awareness. *Journal of Vision*, *14*(4).

Faivre, N., & Koch, C. (2014b). Temporal structure coding with and without awareness. *Cognition*, *131*(3), 404–414.

Feinberg, T. E., & Mallatt, J. (2018). *Consciousness demystified*. Cambridge, Massachusetts: The MIT Press.

Feinberg, T. E., & Mallatt, J. (2019). Subjectivity "demystified": Neurobiology, evolution, and the explanatory gap. *Frontiers in Psychology*, *10*, 1686.

Feldman, J. (2013). The neural binding problem(s). *Cognitive Neurodynamics*, *7*(1), 1–11.

Fodor, J. A. (1994). "The mind-body problem." In Warner, R., & Szubka, T. (Eds.). *The mind-body problem: A guide to the current debate*. Oxford, UK; Cambridge, USA: Blackwell, 24-40.

Gauch, H. G. (2012). *Scientific method in brief*. New York: Cambridge University Press.

Gayet, S., Douw, I., van der Burg, V., Van der Stigchel, S., & Paffen, C. L. E. (2018). Hide and seek: Directing top-down attention is not sufficient for accelerating conscious access. *Cortex*.

Gennaro, R. J. (2012). *The consciousness paradox: Consciousness, concepts, and higher-order thoughts*. Cambridge, Mass.: MIT Press.

Giacino, J. T., Fins, J. J., Laureys, S., & Schiff, N. D. (2014). Disorders of consciousness after acquired brain injury: The state of the science. *Nature Reviews Neurology*, *10*(2), 99–114.

Goldman, A. I. (1997). Consciousness, folk psychology, and cognitive science. In Block, N. J., Flanagan, O. J., & Güzeldere, G. (Eds.), *The nature of consciousness: Philosophical debates* (111-125). Cambridge, Mass.: MIT Press.

Graziano, M. S. A. (2010). *God soul mind brain: A neuroscientist's reflections on the spirit world* (1st ed.). Teaticket, Mass.: Leapfrog Press.

Graziano, M. S. A. (2013). *Consciousness and the social brain*. Oxford; New York: Oxford University Press.

Graziano, M. S. A. (2016). Consciousness engineered. *Journal of Consciousness Studies*, *23*(11–12), 98–115.

Graziano, M. S. A. (2017). The attention schema theory: A foundation for engineering artificial consciousness. *Frontiers in Robotics and AI*, *4*.

Graziano, M. S. A. (2018a). The attention schema theory of consciousness. In R. J. Gennaro (Ed.), *The Routledge Handbook Of Consciousness* (1st ed., 174–187; By R. J. Gennaro).

Graziano, M. S. A. (2018b). *The spaces between us: A story of neuroscience, evolution, and human nature*. Oxford University Press.

Graziano, M. S. A. (2019). *Rethinking consciousness: A scientific theory of subjective experience*. New York, NY: W.W. Norton & Company, Inc.

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

Graziano, M. S. A., & Webb, T. W. (2015). The attention schema theory: A mechanistic account of subjective awareness. *Frontiers in Psychology*, *6*.

Graziano, M. S. A., Guterstam, A., Bio, B. J., & Wilterson, A. I. (2019). Toward a standard model of consciousness: Reconciling the attention schema, global workspace, higher-order thought, and illusionist theories. *Cognitive Neuropsychology*, 1–18.

Hameroff, S., & Penrose, R. (2014). Consciousness in the universe: A review of the 'Orch OR' theory. *Physics of Life Reviews*, *11*(1), 39–78.

Haun, A. M., Oizumi, M., Kovach, C. K., Kawasaki, H., Oya, H., Howard, M. A., Adolphs, R., & Tsuchiya, N. (2017). Conscious perception as integrated information patterns in human electrocorticography. *ENeuro*, *4*(5).

Hentschel, K. (2018). *Photons: The History and Mental Models of Light Quanta* (1st ed. 2018.). Cham: Springer International Publishing: Imprint: Springer.

Herman, W. X., Smith, R. E., Kronemer, S. I., Watsky, R. E., Chen, W. C., Gober, L. M., Touloumes, G. J., Khosla, M., Raja, A., Horien, C. L., Morse, E. C., Botta, K. L., Hirsch, L. J., Alkawadri, R., Gerrard, J. L., Spencer, D. D., & Blumenfeld, H. (2019). A Switch and Wave of Neuronal Activity in the Cerebral Cortex During the First Second of Conscious Perception. *Cerebral Cortex*, *29*(2), 461–474.

Hesselmann, G., & Malach, R. (2011). The link between fMRI-BOLD activation and perceptual awareness is stream-invariant in the human visual system. *Cerebral Cortex*, *21*(12), 2829–2837.

Hesselmann, G., Darcy, N., Rothkirch, M., & Sterzer, P. (2018). Investigating masked priming along the "Vision-for-Perception" and "Vision-for-Action" dimensions of unconscious processing. *Journal of Experimental Psychology: General*, *147*(11), 1641–1659.

Hesselmann, G., Kell, C. A., Eger, E., & Kleinschmidt, A. (2008). Spontaneous local variations in ongoing neural activity bias perceptual decisions. *Proceedings of the*

National Academy of Sciences of the United States of America, *105*(31), 10984–10989.

Hjorthøj, C., Stürup, A. E., McGrath, J. J., & Nordentoft, M. (2017). Years of potential life lost and life expectancy in schizophrenia: A systematic review and meta-analysis. *The Lancet. Psychiatry*, *4*(4), 295–301.

Holland, O., & Goodman, R. (2003). Robots with internal models: A route to machine consciousness? *Journal of Consciousness Studies*, *10*, 77–109.

Horovitz, S. G., Braun, A. R., Carr, W. S., Picchioni, D., Balkin, T. J., Fukunaga, M., & Duyn, J. H. (2009). Decoupling of the brain's default mode network during deep sleep. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(27), 11376–11381.

Hsu, S.-M., George, N., Wyart, V., & Tallon-Baudry, C. (2011). Voluntary and involuntary spatial attentions interact differently with awareness. *Neuropsychologia*, *49*(9), 2465–2474.

Hurme, M., Koivisto, M., Revonsuo, A., & Railo, H. (2017). Early processing in primary visual cortex is necessary for conscious and unconscious vision while late processing is necessary only for conscious vision in neurologically healthy humans. *NeuroImage*, *150*, 230–238.

Irani, K. D. (1980). Conceptual changes in the problem of the mind-body relation. In R. W. Rieber (Ed)., *Body and mind: Past, present, and future* (57-77). New York: Academic Press.

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, *2*(3), 194–203.

James, W. (1890). *The principles of psychology* (Vol. 1). New York: H. Holt and Company.

Keizer, A. W., Hommel, B., & Lamme, V. A. F. (2015). Consciousness is not necessary for visual feature binding. *Psychonomic Bulletin & Review*, *22*(2), 453–460.

Keller, G. B., & Mrsic-Flogel, T. D. (2018). Predictive Processing: A Canonical Cortical Computation. *Neuron*, *100*(2), 424–435.

Kentridge, R. W., Nijboer, T. C. W., & Heywood, C. A. (2008). Attended but unseen: Visual attention is not sufficient for visual awareness. *Neuropsychologia*, *46*(3), 864–869.

King, J.-R., & Dehaene, S. (2014). A model of subjective report and objective discrimination as categorical decisions in a vast representational space. *Philosophical Transactions of the Royal Society B-Biological Sciences*, *369*(1641), 20130204.

King, J.-R., Pescetelli, N., & Dehaene, S. (2016). Brain Mechanisms Underlying the Brief Maintenance of Seen and Unseen Sensory Information. *Neuron*, *92*(5), 1122–1134.

King, Jean-Remi, Sitt, J. D., Faugeras, F., Rohaut, B., El Karoui, I., Cohen, L., Naccache, L., & Dehaene, S. (2013). Information Sharing in the Brain Indexes Consciousness in Noncommunicative Patients. *Current Biology*, *23*(19), 1914–1919.

Koch, C. (1996). "Towards the neuronal substrate of visual consciousness." In Hameroff, S. R., Kaszniak, A. W., & Scott, A. (Eds.). *Toward a science of consciousness: The first Tucson discussions and debates*. Cambridge, Mass.: MIT Press.

Koch, C. (2019). *The Feeling of Life Itself*. London, England: The MIT Press.

Koch, C., & Crick, F. (1994). Some further ideas regarding the neuronal basis of awareness. In C. Koch & J. Davis (Eds.), *Large-scale neuronal theories of the brain* (93–109). Cambridge, Mass.: MIT Press.

Koch, C., & Hepp, K. (2006). Quantum mechanics in the brain. *Nature*, *440*(7084), 611–611.

Koch, C., Massimini, M., Boly, M., & Tononi, G. (2016). Neural correlates of consciousness: Progress and problems. *Nature Reviews Neuroscience*, *17*(5), 307–321.

Koivisto, M., Revonsuo, A., & Salminen, N. (2005). Independence of visual awareness from attention at early processing stages. *NeuroReport*, *16*(8), 817–821.

Koivisto, M., Ruohola, M., Vahtera, A., Lehmusvuo, T., & Intaite, M. (2018). The effects of working memory load on visual awareness and its electrophysiological correlates. *Neuropsychologia*, *120*, 86–96.

Koivisto, Mika, & Grassini, S. (2016). Neural processing around 200ms after stimulus-onset correlates with subjective visual awareness. *Neuropsychologia*, *84*, 235–243.

Koivisto, Mika, & Revonsuo, A. (2008). The role of unattended distractors in sustained inattentional blindness. *Psychological Research-Psychologische Forschung*, *72*(1), 39–48.

Koivisto, Mika, Mäntylä, T., & Silvanto, J. (2010). The role of early visual cortex (V1/V2) in conscious and unconscious visual perception. *NeuroImage*, *51*(2), 828–834.

Komatsu, H., Kinoshita, M., & Murakami, I. (2000). Neural Responses in the Retinotopic Representation of the Blind Spot in the Macaque V1 to Stimuli for Perceptual Filling-In. *Journal of Neuroscience*, *20*(24), 9310–9319.

Kouider, S., de Gardelle, V., Sackur, J., & Dupoux, E. (2010). How rich is consciousness? The partial awareness hypothesis. *Trends in Cognitive Sciences*, *14*(7), 301–307. I

Kwisthout, J., Bekkering, H., & van Rooij, I. (2017). To be precise, the details don't matter: On predictive processing, precision, and level of detail of predictions. *Brain and Cognition*, *112*, 84–91.

Lambert, A. J., Wilkie, J., Greenwood, A., Ryckman, N., Sciberras-Lim, E., Booker, L.-J., & Tahara-Eckl, L. (2018). Towards a unified model of vision and attention: Effects of visual landmarks and identity cues on covert and overt attention movements. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(3), 412–432.

Lasaponara, S., Dragone, A., Lecce, F., Di Russo, F., & Doricchi, F. (2015). The "serendipitous brain": Low expectancy and timing uncertainty of conscious events improve awareness of unconscious ones (evidence from the Attentional Blink). *Cortex*, *71*, 15–33.

Lau, H. C., & Passingham, R. E. (2006). Relative blindsight in normal observers and the neural correlate of visual consciousness. *Proceedings of the National Academy of Sciences*, *103*(49), 18763–18768.

Lau, H., & Rosenthal, D. (2011a). Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Sciences*, *15*(8), 365–373.

Lau, H., & Rosenthal, D. (2011b). The higher-order view does not require consciously self-directed introspection: Response to Malach. *Trends in Cognitive Sciences*, *15*(11), 508–509.

Laureys, S., Celesia, G. G., Cohadon, F., Lavrijsen, J., León-Carrión, J., Sannita, W. G., Sazbon, L., Schmutzhard, E., von Wild, K. R., Zeman, A., Dolce, G., & the European Task Force on Disorders of Consciousness. (2010). Unresponsive wakefulness syndrome: A new name for the vegetative state or apallic syndrome. *BMC Medicine*, *8*(1), 68.

Laureys, S., Goldman, S., Phillips, C., Van Bogaert, P., Aerts, J., Luxen, A., Franck, G., & Maquet, P. (1999). Impaired effective cortical connectivity in vegetative state: Preliminary investigation using PET. *NeuroImage*, *9*(4), 377–382.

LeDoux, J. E., & Brown, R. (2017). A higher-order theory of emotional consciousness. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(10), E2016–E2025.

Lee, M., Baird, B., Gosseries, O., Nieminen, J. O., Boly, M., Postle, B. R., Tononi, G., & Lee, S.-W. (2019). Connectivity differences between consciousness and unconsciousness in non-rapid eye movement sleep: A TMS-EEG study. *Scientific Reports*, *9*, 5175.

Lemon, R. N., & Edgley, S. A. (2010). Life without a cerebellum. *Brain*, *133*(3), 652–654.

Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly*, *64*(4), 354–361.

Levy, J., Vidal, J. R., Fries, P., Demonet, J.-F., & Goldstein, A. (2016). Selective Neural Synchrony Suppression as a Forward Gatekeeper to Piecemeal Conscious Perception. *Cerebral Cortex*, *26*(7), 3010–3022.

Lewis, D. K. (1966). An Argument for the Identity Theory. *The Journal of Philosophy*, *63*(1), 17–25.

Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences*, *99*(14), 9596–9601.

Libet, B. (2004). *Mind time: The temporal factor in consciousness*. Harvard University Press.

Liu, S., Yu, Q., Tse, P. U., & Cavanagh, P. (2019). Neural correlates of the conscious perception of visual location lie outside visual cortex. *BioRxiv*, 660597.

Lou, L., & Chen, J. (2003). Attention and blind-spot phenomenology. *Psyche*, *9*.

Luck, S. J. (2014). *An introduction to the event-related potential technique* (2nd edition.). The MIT Press.

Ludwig, K., Sterzer, P., Kathmann, N., & Hesselmann, G. (2016). Differential modulation of visual object processing in dorsal and ventral stream by stimulus visibility. *Cortex*, *83*, 113–123.

Mack, A., & Rock, I. (1998). *Inattentional blindness*. MIT Press.

Mäki-Marttunen, V., Castro, M., Olmos, L., Leiguarda, R., & Villarreal, M. (2016). Modulation of the default-mode network and the attentional network by self-referential processes in patients with disorder of consciousness. *Neuropsychologia*, *82*, 149–160.

Marti, S., & Dehaene, S. (2017). Discrete and continuous mechanisms of temporal selection in rapid visual streams. *Nature Communications*, *8*.

Massimini, M. & Tononi, G. (2018). *Sizing up consciousness: Towards an objective measure of the capacity for experience*. Oxford, United Kingdom: Oxford University Press.

Meijs, E. L., Slagter, H. A., de Lange, F. P., & van Gaal, S. (2018). Dynamic interactions between top–down expectations and conscious awareness. *Journal of Neuroscience*, *38*(9), 2318–2327.

Melloni, L., Schwiedrzik, C. M., Müller, N., Rodriguez, E., & Singer, W. (2011). Expectations change the signatures and timing of electrophysiological correlates of perceptual awareness. *Journal of Neuroscience*, *31*(4), 1386–1396.

Meng, M., Remus, D. A., & Tong, F. (2005). Filling-in of visual phantoms in the human brain. *Nature Neuroscience*, *8*(9), 1248–1254.

Mensen, A., Marshall, W., Sasai, S., & Tononi, G. (2018). Differentiation analysis of continuous electroencephalographic activity triggered by video clip contents. *Journal of Cognitive Neuroscience*, *30*(8), 1108–1118.

Metzinger, T. (2003). *Being no one: The self-model theory of subjectivity*. Cambridge, Mass.: MIT Press.

Mole, C., Smithies, D., & Wu, W. (2011). *Attention: Philosophical and psychological essays*. New York: Oxford University Press.

Monti, M. M., Lutkenhoff, E. S., Rubinov, M., Boveroux, P., Vanhaudenhuyse, A., Gosseries, O., Bruno, M.-A., Noirhomme, Q., Boly, M., & Laureys, S. (2013). Dynamic Change of Global and Local Information Processing in Propofol-Induced Loss and Recovery of Consciousness. *PLoS Computational Biology*, *9*(10).

Mudrik, L., Breska, A., Lamy, D., & Deouell, L. Y. (2011). Integration without awareness: Expanding the limits of unconscious processing. *Psychological Science*, *22*(6), 764–770.

Nagel, T. (1974). What Is It Like to Be a Bat? *The Philosophical Review*, *83*(4), 435–450.

Nagel, T. (1994). “Conciousness and objective reality.” In Warner, R., & Szubka, T. (1994). *The mind-body problem: A guide to the current debate*. Oxford, UK ; Cambridge, USA: Blackwell, 63-68.

Newman, J., Baars, B. J., & Cho, S.-B. (1997). A Neural Global Workspace Model for Conscious Attention. *Neural Networks*, *10*(7), 1195–1206.

Noel, J.-P., Ishizawa, Y., Patel, S. R., Eskandar, E. N., & Wallace, M. T. (2019). Leveraging Nonhuman Primate Multisensory Neurons and Circuits in Assessing Consciousness Theory. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *39*(38), 7485–7500.

Nunez, P. L. (2016). *The new science of consciousness: Exploring the complexity of brain, mind, and self*. Amherst, New York: Prometheus Books.

Odegaard, B., Knight, R. T., & Lau, H. (2017). Should a few null findings falsify prefrontal theories of conscious perception? *Journal of Neuroscience*, *37*(40), 9593–9602.

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).

Otten, M., Pinto, Y., Paffen, C. L. E., Seth, A. K., & Kanai, R. (2017). The Uniformity Illusion: Central Stimuli Can Determine Peripheral Perception. *Psychological Science*, *28*(1), 56–68.

Overgaard, M., Nielsen, J. F., & Fuglsang-Frederiksen, A. (2004). A TMS study of the ventral projections from V1 with implications for the finding of neural correlates of consciousness. *Brain and Cognition*, *54*(1), 58–64.

Persuh, M., & Melara, R. D. (2016). Barack Obama Blindness (BOB): Absence of Visual Awareness to a Single Object. *Frontiers in Human Neuroscience*, *10*, 118.

Pins, D., & Ffytche, D. (2003). The neural correlates of conscious vision. *Cerebral Cortex (New York, N.Y.: 1991)*, *13*(5), 461–474.

Pinto, Y., van Gaal, S., de Lange, F. P., Lamme, V. A. F., & Seth, A. K. (2015). Expectations accelerate entry of visual stimuli into awareness. *Journal of Vision*, *15*(8).

Pinto, Y., Vandenbroucke, A. R., Otten, M., Sligte, I. G., Seth, A. K., & Lamme, V. A. F. (2017). Conscious visual memory with minimal attention. *Journal of Experimental Psychology: General*, *146*(2), 214–226.

Pinto, Y., Sligte, I. G., Shapiro, K. L., & Lamme, V. A. F. (2013). Fragile visual short-term memory is an object-based and location-specific store. *Psychonomic Bulletin & Review*, *20*(4), 732–739.

Pitts, M.A., Padwal, J., Fennelly, D., Martínez, A., & Hillyard, S. A. (2014). Gamma band activity and the P3 reflect post-perceptual processes, not visual awareness. *NeuroImage*, *101*, 337–350.

Pitts, M. A., Martínez, A., & Hillyard, S. A. (2012). Visual Processing of Contour Patterns under Conditions of Inattentional Blindness. *Journal of Cognitive Neuroscience*, *24*(2), 287–303.

Popper, K. R. (1968). *Conjectures and refutations: The growth of scientific knowledge*. New York: Harper & Row.

Rensink, R., O'Regan, J., & Clark, J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, *8*, 368–373.

Revonsuo, A. (1999). Binding and the Phenomenal Unity of Consciousness. *Consciousness and Cognition*, *8*(2), 173–185.

Revonsuo, A. (2010). *Consciousness: The science of subjectivity*. New York: Psychology Press.

Rosenthal, D. M. (Ed.) (2000). *Materialism and the mind-body problem* (2nd ed.). Indianapolis: Hackett PubCo.

Sanchez, G., Hartmann, T., Fuscà, M., Demarchi, G., & Weisz, N. (2019). Decoding across sensory modalities reveals common supramodal signatures of conscious perception. *BioRxiv*, 115535.

Sasai, S., Boly, M., Mensen, A., & Tononi, G. (2016). Functional split brain in a driving/listening paradigm. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(50), 14444–14449.

Schartner, M., Seth, A., Noirhomme, Q., Boly, M., Bruno, M.-A., Laureys, S., & Barrett, A. (2015). Complexity of Multi-Dimensional Spontaneous EEG Decreases during Propofol Induced General Anaesthesia. *Plos One*, *10*(8), e0133532.

Schmidt, F., & Schmidt, T. (2010). Feature-based attention to unconscious shapes and colors. *Attention, Perception & Psychophysics*, *72*(6), 1480–1494.

Schrouff, J., Perlbarg, V., Boly, M., Marrelec, G., Boveroux, P., Vanhaudenhuyse, A., Bruno, M.-A., Laureys, S., Phillips, C., Pélégrini-Issac, M., Maquet, P., & Benali, H. (2011). Brain functional integration decreases during propofol-induced loss of consciousness. *NeuroImage*, *57*(1), 198–205.

Schurger, A., Sarigiannidis, I., Naccache, L., Sitt, J. D., & Dehaene, S. (2015). Cortical activity is more stable when sensory stimuli are consciously perceived. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(16), E2083–E2092.

Scott, R. B., Samaha, J., Chrisley, R., & Dienes, Z. (2018). Prevailing theories of consciousness are challenged by novel cross-modal associations acquired between subliminal stimuli. *Cognition*, *175*, 169–185.

Seager, W. (1999). *Theories of consciousness: An introduction and assessment*. London ; New York: Routledge.

Searle, J. R. (1997). *The mystery of consciousness* (1st ed.). New York: Review of Books.

Searle, J. R. (2000). Consciousness. *Annual Review of Neuroscience*, *23*(1), 557–578.

Searle, J. R. (2004). *Mind: A brief introduction*. Oxford ; New York: Oxford University Press.

Sergent, C., Baillet, S., & Dehaene, S. (2005). Timing of the brain events underlying access to consciousness during the attentional blink. *Nature Neuroscience*, *8*(10), 1391–1400.

Sergent, C., Ruff, C. C., Barbot, A., Driver, J., & Rees, G. (2011). Top-Down Modulation of Human Early Visual Cortex after Stimulus Offset Supports Successful Postcued Report. *Journal of Cognitive Neuroscience*, *23*(8), 1921–1934.

Seth, A. K. (2015). Presence, objecthood, and the phenomenology of predictive perception. *Cognitive Neuroscience*, *6*(2–3), 111–117.

Seth, A. K., Dienes, Z., Cleeremans, A., Overgaard, M., & Pessoa, L. (2008). Measuring consciousness: Relating behavioural and neurophysiological approaches. *Trends in Cognitive Sciences*, *12*(8), 314–321.

Shannon, C. E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*, *27*.

Shimojo, S., Kamitani, Y., & Nishida, S. (2001). Afterimage of Perceptually Filled-in Surface. *Science*, *293*(5535), 1677–1680.

Shoemaker, S. (1994). "The mind-body problem." In Warner, R., & Szubka, T. (1994). *The mind-body problem: A guide to the current debate*. Oxford, UK ; Cambridge, USA: Blackwell, 63-68.

Silva, S., Alacoque, X., Fourcade, O., Samii, K., Marque, P., Woods, R., Mazziotta, J., Chollet, F., & Loubinoux, I. (2010). Wakefulness and loss of awareness: Brain and brainstem interaction in the vegetative state. *Neurology*, *74*(4), 313–320.

Simons, D. J., & Chabris, C. F. (1999). Gorillas in Our Midst: Sustained Inattentional Blindness for Dynamic Events. *Perception*, *28*(9), 1059–1074.

Sligte, I. G., Scholte, H. S., & Lamme, V. A. F. (2008). Are There Multiple Visual Short-Term Memory Stores? *PLOS ONE*, *3*(2), e1699.

Soto-Faraco, S., & Alsius, A. (2007). Conscious access to the unisensory components of a cross-modal illusion. *NeuroReport*, *18*(4), 347–350.

Sporns, O. (2013). Network attributes for segregation and integration in the human brain. *Current Opinion in Neurobiology*, *23*(2), 162–171.

Sporns, O., Chialvo, D. R., Kaiser, M., & Hilgetag, C. C. (2004). Organization, development and function of complex brain networks. *Trends in Cognitive Sciences*, *8*(9), 418–425.

Squire, L. R. (2013). *Fundamental neuroscience* (4th ed.). Amsterdam ; Boston: Elsevier/Academic Press.

Sun, S. Z., Cant, J. S., & Ferber, S. (2016). A global attentional scope setting prioritizes faces for conscious detection. *Journal of Vision*, *16*(6).

Tacikowski, P., Berger, C. C., & Ehrsson, H. H. (2017). Dissociating the Neural Basis of Conceptual Self-Awareness from Perceptual Awareness and Unaware Self-Processing. *Cerebral Cortex*, *27*(7), 3768–3781.

Tagliazucchi, E., von Wegner, F., Morzelewski, A., Brodbeck, V., Jahnke, K., & Laufs, H. (2013). Breakdown of long-range temporal dependence in default mode and attention networks during deep sleep. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(38), 15419–15424.

Tapal, A., Yeshurun, Y., & Eitam, B. (2019). Relevance-based processing: Little role for task-relevant expectations. *Psychonomic Bulletin & Review*, *26*(4), 1426–1432.

Taylor, P. C. J., Walsh, V., & Eimer, M. (2010). The neural signature of phosphene perception. *Human Brain Mapping*, *31*(9), 1408–1417.

Thakral, P. P. (2011). The neural substrates associated with inattentional blindness. *Consciousness and Cognition*, *20*(4), 1768–1775.

Tofthagen, C. (2012). Threats to Validity in Retrospective Studies. *Journal of the Advanced Practitioner in Oncology*, *3*(3), 181–183.

Tononi, G. (2004). An information integration theory of consciousness. *BMC Neuroscience*, *5*, 42.

Tononi, G. (2008). Consciousness as Integrated Information: A Provisional Manifesto. *The Biological Bulletin*, *215*(3), 216–242.

Tononi, G. (2012a). Integrated information theory of consciousness: An updated account. *Archives Italiennes De Biologie*, *150*(2–3), 56–90.

Tononi, G. (2012b). *Phi: A voyage from the brain to the soul* (1st ed.). Pantheon.

Tononi, G., & Edelman, G. M. (1998). Consciousness and Complexity. *Science*, *282*(5395), 1846–1851.

Tononi, G., & Koch, C. (2015). Consciousness: Here, there and everywhere? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1668).

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.

Tranel, D., & Damasio, A. R. (1993). The covert learning of affective valence does not require structures in hippocampal system or amygdala. *Journal of Cognitive Neuroscience*, *5*(1), 79–88.

Travis, S. L., Dux, P. E., & Mattingley, J. B. (2019). Neural correlates of goal-directed enhancement and suppression of visual stimuli in the absence of conscious perception. *Attention, Perception, and Psychophysics*, *81*(5), 1346–1364.

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., … Straus, S. E. (2018). PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Annals of Internal Medicine*, *169*(7), 467.

Trübutschek, D., Marti, S., Ueberschar, H., & Dehaene, S. (2019). Probing the limits of activity-silent non-conscious working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(28), 14358–14367.

Tsushima, Y., Sasaki, Y., & Watanabe, T. (2006). Greater Disruption Due to Failure of Inhibitory Control on an Ambiguous Distractor. *Science*, *314*(5806), 1786–1788.

Uehara, T., Yamasaki, T., Okamoto, T., Koike, T., Kan, S., Miyauchi, S., Kira, J.-I., & Tobimatsu, S. (2014). Efficiency of a small-world brain network depends on consciousness level: A resting-state fMRI study. *Cerebral Cortex*, *24*(6), 1529–1539.

van Leeuwen, T. M., den Ouden, H. E. M., & Hagoort, P. (2011). Effective connectivity determines the nature of subjective experience in grapheme-color synesthesia. *Journal of Neuroscience*, *31*(27), 9879–9884.

van Vugt, B., Dagnino, B., Vartak, D., Safaai, H., Panzeri, S., Dehaene, S., & Roelfsema, P. R. (2018). The threshold for conscious report: Signal loss and response bias in visual and frontal cortex. *Science (New York, N.Y.)*, *360*(6388), 537–542.

Vandenbroucke, A. R. E., Sligte, I. G., & Lamme, V. A. F. (2011). Manipulations of attention dissociate fragile visual short-term memory from visual working memory. *Neuropsychologia*, *49*(6), 1559–1568.

Vanderwolf, C. H. (2000). Are neocortical gamma waves related to consciousness? *Brain Research*, *855*(2), 217–224.

Vanhaudenhuyse, A., Demertzi, A., Schabus, M., Noirhomme, Q., Bredart, S., Boly, M., Phillips, C., Soddu, A., Luxen, A., Moonen, G., & Laureys, S. (2011). Two Distinct Neuronal Networks Mediate the Awareness of Environment and of Self. *Journal of Cognitive Neuroscience*, *23*(3), 570–578.

von Bartheld, C. S., Bahney, J., & Herculano-Houzel, S. (2016). The search for true numbers of neurons and glial cells in the human brain: A review of 150 years of cell counting. *The Journal of Comparative Neurology*, *524*(18), 3865–3895.

Wang, X., Sang, N., Hao, L., Zhang, Y., Bi, T., & Qiu, J. (2017). Category Selectivity of Human Visual Cortex in Perception of Rubin Face–Vase Illusion. *Frontiers in Psychology*, *8*.

Warner, R., & Szubka, T. (1994). *The mind-body problem: A guide to the current debate*. Oxford, UK ; Cambridge, USA: Blackwell.

Webb, T. W., Kean, H. H., & Graziano, M. S. A. (2016). Effects of Awareness on the Control of Attention. *Journal of Cognitive Neuroscience*, *28*(6), 842–851.

Weisz, N., Wuehle, A., Monittola, G., Demarchi, G., Frey, J., Popov, T., & Braun, C. (2014). Prestimulus oscillatory power and connectivity patterns predispose conscious somatosensory perception. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(4), E417–E425.

Wiese, W. (2018). Toward a mature science of consciousness. *Frontiers in Psychology*, *9*.

Ye, M., Lyu, Y., Sclodnick, B., & Sun, H.-J. (2019). The P3 reflects awareness and can be modulated by confidence. *Frontiers in Neuroscience*, *13*.

Yu, F., Jiang, Q., Sun, X., & Zhang, R. (2015). A new case of complete primary cerebellar agenesis: Clinical and imaging findings in a living patient. *Brain*, *138*(6), e35.

Zadbood, A., Lee, S.-H., & Blake, R. (2011). Stimulus fractionation by interocular suppression. *Frontiers in Human Neuroscience*, *5*, 135.

Zhan, M., Goebel, R., & de Gelder, B. (2018). Ventral and dorsal pathways relate differently to visual awareness of body postures under continuous flash suppression. *ENeuro*, *5*(1).

Zhang, W., & Luck, S. J. (2009). Feature-based attention modulates feedforward visual processing. *Nature Neuroscience*, *12*(1), 24–25.

Zhou, J., Liu, X., Song, W., Yang, Y., Zhao, Z., Ling, F., Hudetz, A. G., & Li, S.-J. (2011). Specific and nonspecific thalamocortical functional connectivity in normal and vegetative states. *Consciousness and Cognition*, *20*(2), 257–268.