

CONSCIOUSNESS VOL 4

Science is close to understanding where consciousness resides in the brain

Science may be one step closer to understanding where consciousness resides in the brain, a new study shows.



[[University of Tokyo](#)]

Where in the brain consciousness resides has been one of the biggest questions in science.

Science may be one step closer to understanding where consciousness resides in the brain. A new study shows the importance of certain types of neural connections in identifying consciousness.

The research, published in [Cerebral Cortex](#), was led by Jun Kitazono, a corresponding author and a project researcher in the Department of General Systems Studies at the University of Tokyo.

"Where in the brain consciousness resides has been one of the biggest questions in science," said Associate Professor Masafumi Oizumi, corresponding author and head of the lab conducting the study. "Although we have not reached a conclusive answer, much empirical evidence has been accumulated in the course of searching for the minimal mechanisms sufficient for conscious experience, or the neural correlates of consciousness."

For this study, the team took a step toward identifying the minimally sufficient subnetworks in the brain that support conscious experience.

To identify the areas of the brain where consciousness resides, the researchers looked for one specific hallmark of consciousness within the neural networks of the brain: bidirectional pathways. When we see something or experience a sensation, our brains take in information. This is called a feed-forward signal, but receiving such feed-forward signals is not enough for consciousness.

Our brains also need to send information back, in what is called feedback. Not every part of the brain can both receive feed-forward and return feedback information. Researchers hypothesized that these bidirectional connections are an essential hallmark of the parts of the brain responsible for consciousness.

"Feed-forward processing alone is insufficient for subjects to consciously perceive stimuli; rather, feedback is also necessary, indicating the need for bidirectional processing. The feedback component disappears not only during the loss of specific contents of consciousness in awake states, but also during unconscious states where conscious experiences are generally lost, such as general anesthesia, sleep and vegetative states," said Kitazono.

It has been suggested that the part of the brain network supporting consciousness brain regions should be bidirectionally connected because both feed-forward and feedback processes are necessary for conscious experience. For example, previous studies examining visual perception have shown that conscious perception does not arise when there is only feed-forward processing, whereas it arises when there is feedback as well as feed-forward processing. (CREDIT: Jun Kitazono)

He also explained that it does not matter if you are looking at a human, monkey, mouse, bird or fly; the bidirectionality of processing remains essential.

Researchers used a mouse connectome and computational techniques to test their idea. A connectome is a detailed map of the connections in the brain. First, they developed an efficient algorithm to extract the parts of the brain with strong bidirectional connections, called complexes. Then, they applied the algorithm to the mouse connectome.

The proposed algorithm can decompose the entire network hierarchically, into the network part with the strongest bidirectional connections, the part with the second strongest, and so on down the line. (CREDIT: Jun Kitazono)

"We found that the extracted complexes with the most bidirectionality were not evenly distributed among all major regions, but rather are concentrated in the cortical regions and thalamic regions," said Kitazono. "On the other hand, regions in the other major regions have low bidirectionality. In particular, regions in the cerebellum have much lower bidirectionality."

These findings align with where scientists have long thought consciousness resides in the brain. The cerebral cortex, located on the surface of the brain, contains sensory areas, motor areas and association areas that are thought to be essential to consciousness experience.

Conventional neuroscience has studied the relationship of what kind of brain activity r occurs in response to an external stimulus s (e.g., an image of an apple). If we write this relationship using the function f as $r=f(s)$, we can say that clarifying the function f is the main research that conventional neuroscience has been doing. Such research has revealed much about the mechanism of information processing, that is, how the brain processes information from external stimuli. On the other hand, our brain not only processes information from the external world, but also produces the subjective experience of "seeing an apple." The ultimate goal of the Oizumi Lab is to theoretically understand the subjective experience and consciousness produced by the brain: that is, to clarify the function g that connects brain activity r and consciousness C , where C is the consciousness produced from brain activity r ($C=g(r)$). (CREDIT: Jun Kitazono)

The thalamus, located in the middle of the brain, has likewise been thought to be related to consciousness, and in particular, the interaction between the thalamus and cortical regions, called the thalamo-cortical loop, is considered important for consciousness. These results support the idea that the bidirectionality in the brain network is a key to identifying the place of consciousness.

Researchers emphasized that they are still working toward identifying the place of consciousness.

"This study focuses only on 'static' anatomical connections between neurons or brain areas. However, consciousness is 'dynamic,' changing from moment to moment depending on neural activity," said Oizumi. "Although anatomical connections tell us how neural activity would propagate and how brain areas would interact, we need to directly investigate the dynamics of neural activity to identify the place of consciousness at any given moment."

As a next step, he said the team is currently analyzing activity-based networks of the brain in various types of neural recordings.

"The ultimate goal of our lab is to find the mathematical relationship between consciousness and the brain," said Oizumi. "In this study, we have attempted to relate the network properties of the brain to the place of consciousness. We will further investigate the relationship between consciousness and the brain, toward what is our ultimate goal."

Journal article: Jun Kitazono, Yuma Aoki, Masafumi Oizumi, "Bidirectionally connected cores in a mouse connectome: towards extracting the brain subnetworks essential for consciousness," [Cerebral Cortex: DOI: 10.1093/cercor/bhac143](https://doi.org/10.1093/cercor/bhac143)

Scientists identify the part of the brain tied to experiencing religion and spirituality

A new study has identified a brain circuit that seems to mediate aspects of our personality responsible for religiosity and spirituality

Scientists have long suspected that religiosity and spirituality could be mapped to specific brain circuits, but the location of those circuits remains unknown. Now, [a new study](#) using novel technology and the human connectome, a map of neural connections, has identified a brain circuit that seems to mediate that aspect of our personality. The work appears in [Biological Psychiatry](#), published by Elsevier.

The study, led by Michael Ferguson, PhD, an investigator in the Center for Brain Circuit Therapeutics at Brigham and Women's Hospital, Boston, MA, USA, shows that the brain circuitry of spirituality centers around a brainstem area called the periaqueductal gray (PAG).

"Patients routinely report that spiritual and religious experiences are among their most meaningful life events. Yet medical science historically has been shy about empirically studying the impacts of spirituality or its physiological mechanisms, Dr. Ferguson said.

"Now, we're at a moment of sufficient cultural openness in science coupled with threshold technological proficiency in brain mapping that we don't need to be so shy anymore about examining spirituality using our best scientific methods."

Some people with neurological diseases such as schizophrenia experience hallucinations or ideations that are religious or spiritual in nature. And previous studies have shown that people with a certain type of epilepsy or with damage in specific brain areas could experience alterations in their spiritual identification. Earlier research has also indicated

that religious experiences are not the product of one particular brain region – that there is no so-called “god spot” – but that spirituality likely arises from dynamic activity across multiple brain regions.

In order to map the newly found brain circuit, Dr. Ferguson and colleagues examined previously collected data about self-reported spirituality from 88 patients before and after undergoing surgery for a brain tumor. The researchers then used a technique called lesion network mapping to analyze how the site of each patient’s lesion, or tumor, interacts with the rest of the brain using a connectome dataset. That analysis revealed the PAG as a critical hub to changes in spiritual identification. Dr. Ferguson and the team then confirmed the importance of the PAG by analyzing previously collected data from 105 veterans of the Vietnam War who experienced head trauma.

The PAG is an evolutionarily old brain structure known for its roles in fear response behaviors and autonomic functions such as regulating heart rate. The PAG is perhaps best known (and most studied) for its role in dampening pain by triggering the release of endogenous opioids, the brain’s own painkillers.

Dr. Ferguson and his colleagues were surprised to find the spirituality circuitry centered on the PAG, rather than “higher” brain regions such as the cortex, which is usually associated with cognitive function and abstract thoughts.

Dr. Ferguson added: “The fact that our results in this study point to such an evolutionarily ancient brainstem structure to define a circuit for spirituality is potentially significant for a litany of reasons. Among the more immediately material reasons is that the [PAG] is well-known to mediate pain inhibition. This leads to curiosity about ways that spirituality might be clinically relevant for helping manage physical and emotional pain.

“The fact that the [PAG] is also involved in attachment and bonding may also hint at mechanistic explanations for the emerging observation that spirituality can be effectively integrated into psychotherapy. These are of course early speculations about possible clinical relevance for the neuroscience of spirituality; the fact, though, that there is so much more work to do in this area is exhilarating and motivating!”

John Krystal, MD, Editor of *Biological Psychiatry*, said of the work: “It is important to understand that this study does not suggest that religion or spirituality in healthy people are in any way abnormal. Rather, this study identifies circuits in the brain that enable us to have religious or spiritual experiences. Researchers have long believed these circuits to exist, because some mental and neurologic illnesses can be associated with alterations in religious experiences, such as religious hallucinations or delusions. However, the study by Dr. Ferguson and colleagues now maps the brain circuits involved in religious experience and spirituality with new precision.”

Exploring consciousness with 'eureka' moments



Credit: Pixabay/CC0 Public Domain© Provided by Medical Xpress

For generations, researchers have been pondering the question of how and where consciousness is formed in the brain. Professor Ekrem Dere from Ruhr

University Bochum, Germany, proposes a new approach to researching conscious cognitive information processing. He advocates defining phases of conscious cognitive processes on the basis of behavioral observations and learning curves.

"Learning is often not a gradual process, but takes place in leaps and bounds; you could say that humans and animals experience sudden epiphanies every now and then," he says. "It's likely that these experiences are preceded by conscious processes."

Dere outlines his new approach, which might apply to both humans and animals, in an article [published](#) in the journal *Frontiers in Behavioral Neuroscience*.

Different levels of consciousness

Consciousness is not an all-or-nothing process. "There are different levels of consciousness, depending for example on whether we're sleeping or writing an email," says Ekrem Dere from the Mental Health Research and Treatment Center in Bochum, who is also a member of the Sorbonne Université in Paris. "At the upper end of this gradation, so to speak, we find conscious cognitive information processing that is required to deal with a complicated problem."

In order to study the neurobiological correlates of these processes using scientific methods, a human or animal must be presented with an experimental task that can only be solved with conscious cognitive information processing—it's crucial that there isn't a preconceived solution.

"In the long history of cognitive behavioral psychology, a great many such tasks have been developed," says Dere. "However, the main difficulty is that a human or animal may not use conscious cognitive information processing throughout the entire processing time."

Eureka moment is the time stamp

The researcher therefore suggests using learning curves to narrow down the phases of conscious information processing with regard to their timing. In these

curves, performance in a specific task is plotted over time. "Learning performance often doesn't improve continuously, but rather by leaps or in stages," explains Dere. This so-called discontinuous learning after insight can serve as a time stamp.

"Conscious cognitive information processing must have taken place at this point and presumably also in the seconds leading up to it," says the psychologist.

"Armed with this knowledge, we can use imaging or electrophysiological methods to observe the brain during conscious cognitive information processing by comparing the time periods immediately before the sudden increase in learning with earlier or later points in time during the processing of the task."

This would enable researchers to find out which mechanisms the brain uses in which region for conscious information processing.

More information: Ekrem Dere, Insights into conscious cognitive information processing, *Frontiers in Behavioral Neuroscience* (2024). [DOI: 10.3389/fnbeh.2024.1443161](https://doi.org/10.3389/fnbeh.2024.1443161)

Provided by Ruhr-Universitaet-Bochum

‘Mysticism’ seeks clarity about a fuzzy but profound concept

‘Mysticism’ seeks clarity about a fuzzy but profound concept© Historia/Shutterstock/Historia/Shutterstock

Mysticism is a concept that famously, perhaps even by definition, refuses to be pinned down. In 1911, the English writer Evelyn Underhill called it "a certain spiritual and intangible quest: the finding of a 'way out' or a 'way back' to some desirable state in which alone [the mystic] can satisfy their craving for absolute truth." In a slightly earlier formulation, William James described mystical experiences as those in which "the sense of a tremendous *muchness* [is] suddenly revealed." To the skeptic, "mysticism" is just another word for mystification. To the believer, it is mystification's opposite: not an obfuscation of reality but a clarification of it, a union with God, or the cosmos, that empties out and remakes the self.

In his ambitious new book, "," the philosopher Simon Critchley engages with Underhill and James — along with the medieval visionary Julian of Norwich, the

post-punk singer-songwriter Julian Cope and T.S. Eliot, among many others — in an attempt “to take mysticism seriously.” He’s tired of his professional field’s blinkered vision, its understanding of the philosopher “as the heroic public defender in the court of reason, or perhaps the humble janitor in the crystal palace of the sciences.”

‘Mysticism’ seeks clarity about a fuzzy but profound concept© NYRB/NYRB

By contrast, Critchley describes himself as “temperamentally a mystic.” In “Four Quartets,” Eliot wrote, “We had the experience but missed the meaning.” Critchley wants both. Though not a religious believer, he longs to give himself over “to that which is larger than the self, outside the self: a pliable ground of love that is prior



to the will.” He also desires to know what this giving over of the self signifies: where the desire comes from and how, if met, it might change one’s life.

Taking mysticism seriously means approaching it from many angles. First, Critchley historicizes: Julian of Norwich wouldn't have understood herself as a mystic, he writes, since "mysticism is a modern category projected back over a largely invented tradition." Then, he considers specific writers on the subject, always alert for how they might, in different ways, teach us to "strip away our attachments to things and to persons and allow us to peer outwards and upwards."

The mystic approaches the ecstatic via the ascetic, through what Simone Weil called "decreation," or, in Critchley's summary, "an undoing of the creature in us" to "open us to what precedes the self." Julian of Norwich had a vision of Christ's crucifixion and spent years unfolding its meanings. Eliot demonstrated how language, precisely and paradoxically through failure, might point toward that which is beyond language. Punk music, through "its exploration of aesthetic *restriction*," gives a strange feeling of plenitude: "Suddenly the world and your place in it shifts and everything opens up in a different way, forever."

Critchley's argument is looping and recursive. He'll make a claim ("Mysticism is a way of describing an existential ecstasy that is outside and more than the conscious self"), then shift ground (mysticism is "a distillation of existing devotional practice, textual interpretation, theological teaching, and conceptual, philosophical reflection") and then shift again (mysticism is "a singular form of floating attention"). This method appropriately mirrors his subject. In trying to articulate the inarticulable, mystics regularly employ repetition and negation. "Mysticism is a learning of the practice of unsaying, an unsaying which must continuously and excessively be said," Critchley writes. In trying to capture the ineffable, "language fails, has to fail, should fail, and should go on failing, loquaciously failing."

Most often, Critchley's loquaciousness is well used, his repetitions effective: Weil's "decreation" helps us to understand Julian of Norwich; Julian illuminates Eliot's work; Eliot enriches our reading of Annie Dillard. Sometimes, though, the book's repetitions feel simply undisciplined. Critchley claims that "mysticism lives on as aesthetic experience," that "mystical experience continues, for us, in aesthetic experience" and that "mysticism lives on in the modern world as aesthetic experience." "It's impossible to be an atheist when listening to the music one loves," he proclaims early on. "When we listen to music that we love," he proclaims later, "it is impossible to be an atheist."

Precision matters, especially when talking about things that frustrate precision. Meister Eckhart sought a different kind of ecstasy than the Stooges did. Critchley knows this, of course, but his wide angle of vision risks dissolving into mistiness. He writes of Julian of Norwich: "The ambition of her writing is to show how the self can somehow lift itself from woe to wellness, from sin to salvation." Yet as a Christian, Julian believed that God's grace, not her own self, did the lifting. Critchley writes that mysticism "might permit us to push back against the violent pressure of reality and allow a richness of life and a possible transfiguration of self and the world." But mysticism doesn't claim to push back against reality. It claims to *reveal* reality, to allow us to see how things are if only we had eyes to see.

"Mysticism" occasionally reads like (deeply, engagingly learned) self-help: "Ecstasy is what it feels like to be alive when we push away the sadness that clings to us"; "why mysticism? To give us some relief from misery." At its strongest, though, and it's often very strong, the book offers a sense not just of the relief that mysticism offers but also the cost it exacts. As Eliot put it, "To apprehend/ The point of intersection of the timeless/ With time, is an occupation for the saint," and the saint's life is a difficult one: "No occupation either, but something given/ And taken, in a lifetime's death in love,/ Ardour and selflessness and self-surrender." If feeling the world's muchness is a gift, it's also a discipline, and "Mysticism" shows just how hard a discipline it can be.

Your Brain Is Not a Computer. It Is a Transducer

A new theory of how the brain works — neural transduction theory — might upend everything we know about consciousness and the universe itself.

By [Robert Epstein](#)

Aug 25, 2021 9:00 AM Aug 25, 2021 1:24 PM



Let's start with my 95-year-old mom. Her memory is unreliable, but she's still lucid, churns out sarcasm like a pro, and plays a lightning-fast game of double solitaire. Today I finally quit after she won seven games in a row, and, yes, I was trying my best.

She also hears music continuously, and it's not the kind of music that drives us nuts when we can't get a tune out of our head. She mainly hears *original* music, and she will sometimes try to hum or sing what she's hearing. She says it's coming from "the neighbors downstairs," and it doesn't bother her, she says, because some of it isn't bad and because it helps her fall asleep. The fact that other people can't hear it doesn't bother her either. She simply smiles slyly and says, "Maybe you should get your hearing checked."

Am I concerned? Well, just a bit — not about the music but about its source. As I told my mom the other day, I would be more comfortable if the inconsiderate neighbors lived *upstairs*. She laughed and said, "I see what you're getting at, but don't worry. I'm *not* going to hell." Very determined, my mom. Was she planning to negotiate the issue with a hand of double solitaire?

Where is all this original music coming from? My mother has never composed music, and she insists she would be incapable of doing so "no matter how much you paid me." That's mom-speak for case closed.

In case you haven't noticed, we are surrounded by mysteries like this. Some, in my view, are highly suspect, such as demonic possession and communication with spirits. Others are undeniably real: dreams, daydreams, hallucinations, the *déjà vu* experience, and so on. My staff recently came up with a list of 58 such phenomena. You don't have to look far to find them.

Are we doomed to remain in the dark about these mysteries, or is there a way to peel back the veil? What if spirits, dreams, and my mom's music could all be accounted for by a relatively simple idea about how the brain works — an idea that might even be testable?

The idea, which is quite simple on its face, is that the brain is a bidirectional transducer.

Transduction Is All Around Us

A few inches to the left of the laptop computer on which I am now typing stands an imposing Sennheiser microphone that I use mainly when someone is interviewing me

for a radio or TV show. A thick black cable protrudes from the back of the microphone and snakes its way to a boxy analogue-to-digital converter, which links to my computer via a bright red USB cable.

My computer links wirelessly to a router in the next room, and the router connects to AT&T, my internet service provider, through a telephone cable that runs from the router to a wall socket. That cable leads to dozens of other transition points through which crude representations of my voice pass before some semblance of it is finally heard by a talk-show host in, say, London.

If I'm speaking to a host of a BBC radio program, representations of his or her voice are also traveling in the other direction from his or her microphone through dozens of transition points until they finally activate a tiny speaker in my ear bud, through which I hear a semblance of his or her voice. If we're also using video to communicate, cameras are sending images to screens, again through many transition points, and again in both directions.

Representations of these images and sounds might be passing not only through multiple transition points but also through thousands of miles of copper or fiber optic cables, or are perhaps being transmitted to satellites hundreds of miles above the earth and then re-transmitted to receivers on the ground. Long ago, pathways like this could accommodate only one conversation at a time, with communications running in only one direction at a time, but now such pathways are bidirectional and are often shared simultaneously by thousands of different conversations.

Remarkably, when everything is working smoothly, my conversation with the London host is as seamless as it would be if we were in the same room. Even though I'm 5,500 miles away in San Diego, I can't detect any time gaps between my ramblings and the host's reply. Those gaps exist, but they're so short that neither I nor the host can perceive them.

What's happening here? Is my voice actually traveling 5,500 miles? Definitely not. If you were the proprietor of a 19th century mansion, you might have been able to shout into a speaking tube that snaked its way through your home and carried your voice all the way to the servants' quarters. In that case, the sound of your voice was literally carried to its destination by the air in the tube.

But when I speak into my microphone, the pattern of sound waves produced by my voice — a distinctive, non-random pattern of air pressure waves — is being *converted* by the microphone into a similar pattern of electrical activity. The

better the microphone, the more accurately it duplicates the original pattern, and the more I sound like me at the other end.

That conversion process — the shifting of a meaningful, non-random pattern of activity — from one medium (say, the air in front of the microphone) to another medium (say, the wire at the back of a microphone) is called *transduction*.

And transduction is all around us, even in organic processes. Our bodies are completely encased by transducers. Our sense organs — eyes, ears, nose, tongue, and skin — transduce distinctive properties of electromagnetic radiation, air pressure waves, airborne chemicals, liquid-borne chemicals, textures, pressure, and temperature into distinctive patterns of electrical and chemical activity in the brain. Organic compounds can even be used these days to create new kinds of transducers, such as OECTs: [organic electrochemical transistors](#).

Evolution didn't just create millions of new species of organisms, it also created millions of new types of transducers, and engineers are now using both organic and inorganic materials to create thousands more.

To repeat (because this is important): Transduction is all around us — forms of transduction that have evolved over eons and new forms of transduction that humans are inventing right now.

The Ultimate Transducer

What if evolution, at some point, produced a special kind of transducer that could shift signals from the physical world as we know it to a very different kind of world?

Nearly all religions teach that immaterial realms exist that transcend the reality we know. For Christians and Muslims, those realms are Heaven and Hell. One of the simplest and clearest statements of such a concept comes from ancient Greek mythology: As long as the deceased had the required toll in hand — well, actually in mouth — he or she would be transported by the ferryman Charon across the river Styx to Hades, the land of the dead — quite literally, to the *Other Side*. (I'll call it the OS from this point on.) Unfortunately, not everyone was eligible to make the crossing. If no one thought to bury you or to put that coin in your mouth, you were doomed to roam *this* side of the river as a ghost.

The idea of a realm transcending the one we experience directly has taken on many forms over the centuries. George Griffith, England's most prominent and prolific

science fiction writer of the late 1800s, published a prescient novel about this realm in 1906: [*The Mummy and Miss Nitocris: A Phantasy of the Fourth Dimension*](#). The book's protagonist, professor Franklin Marmion, is a distinguished mathematician and physicist who anticipates discoveries and concepts that real quantum physicists would eventually propose decades later. Over the course of the story, Marmion not only reluctantly accepts the fact that a higher dimension must exist, he also acquires the power to shift his body there, learning, among other counterintuitive things, that multiple objects can occupy the same space at the same time.

Griffith might have been aware of a contemporary of his, William James, a prominent Harvard philosopher and also arguably America's first psychologist. In 1898, James published a short book entitled [*Human Immortality: Two Supposed Objections to the Doctrine*](#), in which he praised his contemporaries for boldly using scientific methods to investigate "providential leadings [sic] in answer to prayer, instantaneous healings, premonitions, apparitions at time of death, clairvoyant visions or impressions, and the whole range of mediumistic capacities."

James asserted that a universe-wide consciousness beamed human consciousness into our brains "as so many finite rays," just as the sun beams rays of light onto our planet. Our brains, being limited in their capabilities, he said, generally filter and suppress real consciousness, while sometimes allowing 'glows of feeling, glimpses of insight, and streams of knowledge' to shine through. He called this idea 'transmission-theory.'

Ideas like James's have been around for thousands of years. In his 2006 book, *Life After Death: The Burden of Proof*, alternative medicine guru Deepak Chopra says that ancient Hindu texts teach that the material world we know is nothing but a projection from the universal consciousness that fills all space. From this perspective, death is not an end; it is a merging of a relatively pathetic human consciousness with that of the dazzling universal one. To add gravitas to this idea, Chopra does what many recent authors have done: he suggests that modern formulations of quantum physics are consistent with his belief in a universal consciousness.

The connection between physics and modern theories of mind and consciousness is [*tenuous at best*](#), but modern physicists do take the idea of parallel universes seriously. They debate the details, but they can hardly ignore the fact that the mathematics of [*at least three*](#) of the grand theories at the core of modern physics — inflation theory, quantum theory, and string theory — predict the existence of alternate universes. Some physicists even believe that signals can leak between the universes and that the existence of parallel universes [*can be confirmed*](#) through measurements or experiments. In a recent essay, physicist [*A. A. Antonov*](#) argues that our inability to

detect the vast amount of dark energy that almost certainly exists in our own universe is evidence of the existence of parallel universes, six of which, he speculates, are directly adjacent to our own.

Again, setting the details aside, physicists agree that the three-dimensional space we experience is simply not the whole picture. As theoretical physicist Lee Smolin put it recently, "[Space is dead](#)."

Evidence for Transduction?

Hard evidence that supports a neural transduction theory is lacking at the moment, but we are surrounded by odd phenomena that are at least consistent with such a theory. And, no, I'm not talking about the claims that best-selling authors have made over the decades about proof that telepathy, out-of-body experiences, and communication with the dead are real. No such proof exists, in my view, but other well documented phenomena are difficult to brush aside.

When I was a graduate student at Harvard, I noticed a stranger roaming the hallway near my office and offered to assist her. Doris, it turns out, had heard voices for years, and she was hoping she could find someone in the psychology building — William James Hall — to help her eliminate them because they "caused trouble." I didn't have the heart to tell her that Harvard, at the time, had no clinical psychology program and that I was doing behavioral research with pigeons. If her voices had sent her there, they were troublemakers indeed.

When internal perception goes awry, people can be overwhelmed by hallucinations, visions, or distortions of reality so extreme they have to be hospitalized, and Doris had been hospitalized at times.

But is Doris that different from the rest of us? After all, even the healthiest among us hallucinate several times each night — we call it dreaming. And we all have at least two highly disorienting experiences each day called "hypnogogic states" — those eerie, sometimes creative interludes between sleeping and waking.

I have sometimes dreamt intricate full-length movies that seemed as good as any Hollywood film. Alas, most of the time, no matter how hard I try, I can't seem to hold on to even a shred of a dream during the few seconds when I'm staggering from my bed to the bathroom.

Where does all this content come from, and why do we have so little control over it?

In recent years, researchers have explored what they have clumsily labeled "paradoxical lucidity," or, even worse, "terminal lucidity." These labels refer to what some of us know as "the last hurrah" — the burst of mental clarity that sometimes occurs shortly before people die, even people for whom such clarity should be impossible.

For more than two centuries, medical journals have published credible reports of highly impaired, uncommunicative people who suddenly became lucid for a few minutes before they died. There are documented cases in which people with dementia, advanced Alzheimer's, schizophrenia, and even severe brain damage — people who have not been able to speak or to recognize their closest relatives for years — suddenly recognized their loved ones and spoke normally.

[A 2020 study](#) summarizing the observations of 124 caregivers of dementia patients, concluded that in "more than 80 percent of these cases, complete remission with return of memory, orientation, and responsive verbal ability was reported by observers of the lucid episode" and that "[the] majority of patients died within hours to days after the episode." The periods of lucidity typically lasted 30 to 60 minutes.

Some of the historical reports of lucid episodes are truly extraordinary.

Here is one of many cases [reported by the German biologist Michael Nahm](#) and his colleagues in 2012:

In a case published in 1822, a boy at the age of 6 had fallen on a nail that penetrated his forehead. He slowly developed increasing headaches and mental disturbances. At the age of 17, he was in constant pain, extremely melancholic, and starting to lose his memory. He fantasized, blinked continuously, and looked for hours at particular objects.... He remained in the hospital in this state for 18 days. On the morning of the 19th day, he suddenly left his bed and appeared very bright, claiming he was free of all pain and feelings of sickness.... A quarter of an hour after the attending physician left him, he fell unconscious and died within a few minutes. The front part of his brain contained two pus-filled tissue bags the size of a hen's egg (Pfeuffer, 1822)....

And another:

Haig (2007) reported the case of a young man dying of lung cancer that had spread to his brain. Toward the end of his life, a brain scan showed little brain tissue left, the metastasized tumors having not simply pushed aside normal brain tissue but actually destroyed and replaced it. In the days before his death, he lost all ability to speak or

move. According to a nurse and his wife, however, an hour before he died, he woke up and said good-bye to his family, speaking with them for about five minutes before losing consciousness again and dying.

If the brain is a self-contained information processor, how can we explain the sudden return of lucidity when the brain is severely damaged? For that matter, think about the variability that occurs in your own lucidity over the course of 24 hours, during which you are, at various times, completely unconscious, partially conscious, or fully conscious. If you add drugs and alcohol to the picture, the variability is even greater, and it can be quite bizarre.

The variability problem is addressed in an [intriguing paper](#) published by Jorge Palop and his colleagues in *Nature* in 2006, who note that patients suffering from a variety of neurodegenerative disorders often fluctuate over the course of a single day between states of extreme confusion and relatively normal mental states. Such radical changes, they note, "cannot be caused by sudden loss or gain of nerve cells." They speculate about changes in neural networks, but that doesn't solve the problem.

What if the variability is not caused by changes in processing power in the brain but rather by transduction effects? By changes occurring not in our local universe but in the OS? Or by minor changes occurring at the point of connection? Or by changes occurring in brain structures that are essential to signal transfers?

I've also been intrigued by what appear to be credible reports about visual experiences that some congenitally-blind people have had when they were near death.

Experiences of this sort were first summarized in a [1997 paper](#) by Kenneth Ring and Sharon Cooper, later expanded into a book called *Mindsight* (1999). The paper and book describe the experiences of 14 people who were blind from birth and who had near-death experiences (NDEs), some of which included content that appeared to be visual in nature. Soon after Vicki U. was in a near-fatal car accident at age 22, she remembered "seeing" a male physician and a woman from above in the emergency room, and she "saw" them working on a body. Said Vicki:

I knew it was me.... I was quite tall and thin at that point. And I recognized at first that it was a body, but I didn't even know that it was mine initially. Then I perceived that I was up on the ceiling, and I thought, "Well, that's kind of weird. What am I doing up here?" I thought, "Well, this must be me. Am I dead?"

Vicki had never had a visual experience before her NDE, and, according to the researchers, did not even "understand the nature of light." While near death, she also claimed to have been flooded with information about math and science. Said Vicki:

I all of a sudden understood intuitively almost [all] things about calculus, and about the way planets were made. And I don't know anything about that.... I felt there was nothing I didn't know.

Several aspects of Vicki's recollections are intriguing, but the most interesting to me are the visual experiences. How can someone who has never had such an experience "No light, no shadows, no nothing, ever," according to Vicki — suddenly have rich and detailed experiences of this sort? Ring and Cooper found others like Vicki — congenitally blind people who not only had visual experiences when near death but whose NDEs were remarkably similar to some common NDEs of sighted people.

Just recently, an Australian woman [made the news](#) worldwide when, post surgery, she woke up with an Irish accent. Her strong Australian accent was completely gone. Called 'the foreign accent syndrome', this sudden switch in accents is rare but real. The shift doesn't make sense given the framework of reasoning we usually apply to the world, but what if it's a transduction error?

And why can't we remember pain? We can remember facts and figures and images, and we can even get choked up remembering strong emotions we've felt in the past — but *we can't remember pain*. Are sensations of pain getting filtered out by transduction pathways? Could that be why our dreams are pain free? That begs a question that is both eerie and obvious: Is the OS a kind of pain-free Heaven?

And have you ever met a stranger who made you feel, almost immediately, that you had known him or her your entire life? And sometimes this stranger has the same feeling about *you*. It's a strong feeling, almost overwhelming. We can try to explain such feelings with speculations about how a voice or physical characteristics might remind us of someone from our past, but there is another possibility — that in some sense *you had actually known this person your whole life*. If the brain is a bidirectional transducer, that is not a strange idea at all.

In fact, when viewed through the lens of transduction theory, *none* of these odd phenomena — dreams, hallucinations, lucidity that comes and goes, blind vision, and so on — looks mysterious.

And All That Jazz

This brings me, reluctantly, to the recent rise of “postmaterialist” science, or at least postmaterialist psychology. The latter is marked by — or perhaps blemished by — the founding of the journal *Spirituality in Clinical Practice* by the American Psychological Association (APA) in 2014 and the founding of the Association for the Advancement of Postmaterialistic Sciences in 2017. (Disclosure: I have been a full member of the APA since 1983.)

Postmaterialism is all about controlled experiments that have supposedly proved, or at least supported, claims that mediums can communicate with the dead, that ghosts will happily comply when we ask them to climb into little boxes in a laboratory, that people can send their thoughts to strangers in another room telepathically, and that future events can somehow travel backward in time to impact people’s current behavior.

I am so tempted here to start naming names and tearing down reputations, but my musical mom raised me better than that. I will say this: One of the early papers published in that new APA journal — a paper that was accepted without peer review — demonstrated thinking so shoddy it startled me.

Fortunately, I don’t *need* to tear apart shoddy thinking or flawed experiments to advocate for transduction theory. In fact, if this theory proves to be valid, every fantasy of the postmaterialists will be fulfilled — every fantasy except one, that is, and that is *the postmaterialist claim itself*. That’s because parallel universes are not wispy, physics-free spiritual entities; according to many mainstream physicists, they are just non-obvious companions of the material universe in which we happen to live.

A Better Brain Theory

Let’s set aside both the mundane and the exotic reasons we should take transduction theory seriously and get to the heart of the matter: The main reason we should give serious thought to such a theory has nothing to do with ghosts. It has to do with the sorry state of brain science and its reliance on the computer metaphor. One of my research assistants recently calculated that Beethoven’s thirty-two piano sonatas contain a total of 307,756 notes, and that doesn’t take into account the hundreds of sections marked with repeat symbols. Beethoven’s scores also include more than 100,000 symbols that guide the pianist’s hands and feet: time signatures, pedal notations, accent marks, slur and trill marks, key signatures, rests, clefs, dynamic notations, tempo marks, and so on.

Why am I telling you about Beethoven? Because piano virtuoso and conductor Daniel Barenboim memorized *all thirty-two of Beethoven’s sonatas by the time he was 17*,

and he has since memorized hundreds of other major piano works, as well as dozens of entire symphony scores — *tens of millions* of notes and symbols.

Do you think all this content is somehow *stored* in Barenboim's [ever-changing, ever-shrinking, ever-decaying brain](#)? Sorry, but if you study his brain for a hundred years, you will never find a single note, a single musical score, a single instruction for how to move his fingers — not even a “representation” of any of those things. The brain is simply *not* a storage device. It is an extraordinary entity for sure, but not because it stores or processes information. (See my Aeon essay, “[The Empty Brain](#),” for more of my thinking on this issue, and for a truly great thrill, watch Barenboim play the third movement from Beethoven’s 14th piano sonata [here](#).)

Over the centuries — completely baffled by where human intelligence comes from — people have used one metaphor after another to ‘explain’ our extraordinary abilities, beginning, of course, with the divine metaphor millennia ago and progressing – and I use that word hesitatingly — to the current information-processing metaphor. I am proposing now that we abandon the metaphors and begin to consider substantive ideas we can test.

To be clear: I am *not* offering transduction theory as yet another metaphor. I am suggesting that the brain is truly a bidirectional transducer and that, over time, we will find empirical support for this theory.

Recall that Einstein’s Special Theory of Relativity, published in 1905, and then his General Theory of Relativity, published in 1915, received no direct and convincing empirical support for years — first regarding predictions his equations made about the perihelion precession of Mercury’s orbit, then about the bending of light around the sun (observed by Sir Arthur Eddington in 1919), and then about the gravitational redshift of light. It took a full century before his predictions about gravity waves were confirmed.

If we can cast some aspects of transduction theory into formal, predictive terms (I’m working on that now and am looking for collaborators), we might be able to make specific predictions about transduction — about subtle variations in reaction times, for example, or about how transduction errors might help us explain schizophrenia. We might also be able to predict quantitative aspects of dreams, daydreams, hallucinations, and more.

Ignore It at Your Peril

If transduction theory has merit, let's think about what happens if we ignore it. If we transported a 17th century scientist to the present day and showed him or her how well we can converse with someone using a cell phone, he or she would almost certainly want to look inside the phone. The remote voice must be *in* the phone, after all. To put this another way, a Renaissance scientist would naively view the phone as a self-contained processing unit, much as today's brain scientists naively view the brain.

But that scientist will never find the remote voice inside the phone, because it is not there to be found.

If we explain to the scientist that the phone is a transducer, however, he or she will now examine the phone in a different way, searching for evidence of transduction, which he or she — aided by appropriate instruments and knowledge — will eventually find.

And here is the problem: If you *never* teach that scientist about transduction, he or she might *never* unravel the mysteries of that phone.

This brings me to the claustrum, a small structure just below the cerebral cortex that is [poorly understood](#), although [recent research](#) is beginning to shed some light. Many areas of the brain connect to the claustrum, but what does it do? If the claustrum turns out to be the place where signals are transduced by the brain, *you will probably never discover this remarkable fact if transduction is not on your list of possibilities*. (If you're a history buff, you might also be aware of another small brain structure — the pineal gland — that could conceivably be a transduction site. In his first book, *Treatise of man*, written in the early 1600s, French philosopher René Descartes identified this gland as the seat of the soul. Remarkably, in the late 1900s, [scientists discovered](#) that tissue in the pineal gland responds to electromagnetic radiation.)

If modern brain scientists begin to look for evidence that the brain is a transducer, they might find it directly through a new understanding of neural pathways, structures, electro-chemical activity, or brain waves. Or they might find such evidence indirectly by simulating aspects of brain function that appear to be capable of transducing signals. They might even be able to create devices that send signals to a parallel universe, or, of greater interest, that *receive* signals from that universe. Comparative studies of animal brains, which could conceivably have limited connections to the OS, might help move the research along.

Efficient and clear transduction might also prove the key to understanding the emergence of human language and consciousness; here is a possible explanation for

what might have been the relatively sudden appearance of such abilities in humans (see Julian Jaynes's 1976 book, *The Origin of Consciousness in the Breakdown of the Bimodal Mind*). Neural transduction might also prove to be the mechanism underlying Carl Jung's concept of the "collective unconscious." Even Noam Chomsky's theory of universal grammar could get a boost from transduction theory; it would hardly be surprising that most or all human languages share certain grammatical rules if languages are all constrained by signals emanating from a common source. And then there's that 'flow' state my friend Mihaly Csikszentmihalyi taught everyone about. When I'm in a hyper-creative mode – now, for example, as I'm writing – I have almost no awareness of this world or of the passage of time. Is the OS the source of our creativity?

It might take decades for us to see significant advances in transduction research, but with vast resources already devoted to the brain sciences, we could conceivably move much faster. And if you're worried that transduction theory is just another one of those inherently untestable theories — like string theory or theories about parallel universes — think again. With neural transduction theory, we have an enormous advantage: *The transduction device is available for immediate in-depth study.*

Implications and Final Notes

Will transduction theory finally clear up the old consciousness problem? That I doubt, because I don't think there *is* a consciousness problem. Consciousness is just the experience we have when we're observing ourselves or the world. It seems grand simply because we're part of the system we're observing. It's a classic example of how difficult it can be to study a system of which one is an integral part; think of this problem as a kind of Gödel's theorem of the behavioral sciences. (For my whole spiel on this issue, see my 2017 essay, "[Decapitating Consciousness](#).")

If transduction theory proves to be correct, our understanding of the universe and of our place in it will change profoundly. We might not only be able to make sense of dozens of odd aspects of human experience, we might also begin to unravel some of the greatest mysteries in the universe: where our universe came from, what else and who else is out there — even whether there is, in some sense, a God.

If you are as skeptical about flimsy theories as I am, by now you might be thinking: Has Epstein lost his mind (and, if so, where did it go)? Let me assure you that I am as hard-headed as ever. I won't believe in ghosts until Casper himself materializes in front of an audience and pushes me off the stage. But I am also acutely aware of how little we actually know, both about ourselves and our universe. If one simple idea — brain as

transducer — might stimulate new kinds of research and might also bring order to what seem to be scores of unrelated, bizarre, and highly persistent human beliefs, I'm all for it.

Robert Epstein is senior research psychologist at the American Institute for Behavioral Research and Technology in California. He holds a doctoral degree from Harvard University and is the former editor-in-chief of Psychology Today magazine. He's authored 15 books and more than 300 articles on various topics in the behavioral sciences. His mathematical model of the Carrier Separation Plan, a strategy for quickly eradicating SARS-CoV-2 and comparable pathogens, can be accessed at <https://doi.org/10.3389/fpubh.2021.640009>.

Consciousness Beyond Neural Fields: Expanding the Possibilities of What Has Not Yet Happened

[Birgitta Dresch-Langley](#)^{1,*}

- Author information
- Article notes
- Copyright and License information

PMCID: PMC8784399 PMID: [35082717](#)

Abstract

In the field theories in physics, any particular region of the presumed space-time continuum and all interactions between elementary objects therein can be objectively measured and/or accounted for mathematically. Since this does not apply to any of the field theories, or any other neural theory, of consciousness, their explanatory power is limited. As discussed in detail herein, the matter is complicated further by the facts that any scientifically operational definition of consciousness is inevitably partial, and that the phenomenon has no spatial dimensionality. Under the light of insights from research on meditation and expanded consciousness, chronic pain syndrome, healthy aging, and eudaimonic well-being, we may conceive consciousness as a source of potential energy that has no clearly defined spatial dimensionality, but can produce significant changes in others and in the world, observable in terms of changes in time. It is argued that consciousness may have evolved to enable the human species to generate such changes in order to cope with unprecedented and/or unpredictable adversity. Such coping could, ultimately, include the conscious planning of our own extinction when survival on the planet is no longer an acceptable option.

Keywords: field theory, physics, consciousness, evolution, deep meditation, aging and wellbeing people, *eudaimonia*

Introduction

In field theories of consciousness (e.g., [Köhler, 1940](#); [Cacha and Poznanski, 2014](#) among others), the latter is conceived in terms of a field in the sense in which it is used in quantum or particle physics, where the notion of “field” applies to all fundamental forces and relationships between elementary particles within a unifying theoretical framework where the forces lead to energy fields that occupy space-time and mediate interactions between elementary particles. In field theories of consciousness the latter is, similarly, seen as having duration and extension in space. In field theories in physics, however, each point of a particular region of the presumed space-time continuum, as well as all interactions between elementary objects, are objectively measurable and accounted for mathematically. This cannot be claimed by any current

theory of consciousness, including the field theories (e.g., [Köhler, 1940](#); [Lashley et al., 1951](#); [Fingelkurts and Fingelkurts, 2002](#); [John, 2002](#); [McFadden, 2002](#); [Cacha and Poznanski, 2014](#)). These will not be reviewed in detail again here, as an excellent review has been provided earlier by [Pockett \(2013\)](#) earlier. Reasons why Libet's Mind Field Theory of consciousness, carved out in his book *Mind Time* (2004), may be discussed outside rather than within the realm of the other field theories are clarified herein. Neural field theories of consciousness, whether they relate to representational fields, where *Gestalten* or *qualia* are seen as reflecting the very nature of consciousness, occupying a presumed spatio-temporal brain field generating electrical brain states ([Köhler, 1940](#)), or to the functionally specific spatio-temporal structure of an electromagnetic field in the brain ([Lashley et al., 1951](#); [Fingelkurts and Fingelkurts, 2002](#); [John, 2002](#); [McFadden, 2002](#)), all account for specific aspects of brain-behavior function while humans are in a conscious or non-conscious state. Yet, consciousness is a far more complex product of brain evolution, both at the phylogenic ([Cabanac et al., 2009](#)) and at the ontogenetic ([Jaynes, 1990](#); [Feinberg and Mallatt, 2013](#)) scales, that reaches well beyond brain-behavior function (e.g., [Pockett, 2004](#)). Since the theory of evolution, the problem of the origin of mind or, more specifically, the origin of consciousness has been posed. Where and how in evolution has it begun to emerge, and how can it be measured. How can we derive mindfulness out of mere biophysical matter? If we want to define consciousness adequately, it would need to be in terms of the capability of the human Self to know and analyze its own condition and existence in space and time, and to project this knowledge into a future that has not yet happened. Why it may not be possible to render the whole complexity of the phenomenon of consciousness scientifically operational is discussed further in the following chapters. Ontological links between mind, time, and the Self as a window to understanding a specific aspect of human consciousness, the ability to project one's own existence into the future, are brought forward. Finally, the question whether we need neural field theories of consciousness as an explanation of the latter at all is justified under the light of clear argumentation. It is concluded that investigating functional links between eudaimonic well-being and consciousness could give us answers more important to the future of humankind.

Limitations to a Scientifically Operational Definition of Consciousness

None of the field theories of consciousness has succeeded in providing a definition that would be both scientifically operational and, at the same time, capture the complex nature of this phenomenon. This was already pointed out some time ago by [Block \(2007\)](#) as a clear limitation to not only field theories, but any theory of consciousness. He argued for an "abstract solution" to the "problem of consciousness" given that phenomenal consciousness by far exceeds perception, cognitive accessibility and performance, or any of their directly measurable brain correlates. What others have referred to as the "hard problem of consciousness" (e.g., [Chalmers, 1996](#); [Searle, 1998](#)) relates to the difficulty of finding brain measures of a highly complex phenomenon, the conscious *Self*, experienced in terms of *I do, I think, I feel, I was, am, and will be*, independently of any particular conscious perception, memory, decision, or action

(behavior). If a representational or neural field of consciousness occupying a presumed space-time continuum inside the brain, or outside the brain, as suggested by [Sheldrake \(2013\)](#) and others, exists, it would have to be independent of the neural activities underlying any particular perceptual or cognitive process operating at the same time. While a specific conscious perception or conscious memory recall, can be measurably correlated with specific brain activities (e.g., [Nani et al., 2019](#)), interpreted adequately as the neural correlates of the particular conscious behavior being measured, it is not a neural correlate of consciousness. The contents of phenomenal experience are associated with neural activities in multiple networks ([Rees et al., 2002](#)) of the temporal-parietal-occipital brain areas, but this does not account for how and why, or through which mechanisms, our brain has evolved consciousness, while most other mammal brains have not, or at least not to the same extent. To get around this problem, the concept of a “conscious brain state,” with an abstract and scientifically operational definition of consciousness in terms of a “continuous process with limited duration” was introduced ([Tononi and Edelman, 1998](#), and subsequently within a larger theoretical framework of consciousness, [Edelman, 2003](#)) based on a definition proposed earlier ([von der Malsburg, 1997](#)). Such a parsimonious definition of consciousness may allow looking for invariant properties of a brain state during conscious wakefulness and its equivalents. [LaBerge et al. \(1986\)](#) and [LaBerge \(1990\)](#) argued that states of lucid dreaming, for example, are equivalent to wakeful conscious states. The invariant properties of what John called “the conscious ground state,” which in his theory corresponds to a specific EEG wave pattern that is observed when patients recover from deep anesthesia, could then be told apart from the subjective phenomenal contents of any particular conscious experience as reflected in a particular overt behavior. In other words, the conscious brain state would have universal properties that can be consistently identified and measured whether we are consciously daydreaming ([Singer, 1975](#); [Carver and Humphries, 1981](#)), engaged in abstract analytic thinking ([Gilead et al., 2014](#)), in a state of lucid dreaming ([LaBerge, 1990](#)), or whether we consciously perceive and remember objects, as in conscious three-dimensional perception and selective memory recall ([Nani et al., 2019](#)). Invariant or universal properties of a generic conscious brain state were, however, never found elsewhere than in brain patterns measured after recovery from deep anesthesia ([John, 2002](#)), which is, yet again, only a particular conscious state to which consciousness is not reducible. Fifteen years ago, the nuclear physicist Jean Durup and myself ([Dresp-Langley and Durup, 2009](#)) suggested a biophysical brain code independent of particular cognitive processes operating while we are conscious (perception, spatial awareness, conscious motor planning and execution etc.), but providing a generic neural mechanism that would trigger, maintain, and terminate an individual conscious state in similar ways in which electronic bar codes may activate, maintain, and disable the electronic locks of a safe. The functional assumptions underlying this somewhat wild concept were inspired by earlier work on an arbitrary, and possibly genetically prewired, selection of dedicated neural circuitry for consciousness ([Helekar, 1999](#)), temporal coincidence coding in the brain (e.g., [Ainsworth et al., 2012](#)), and adaptive resonance theory ([Grossberg, 1999, 2000](#)). Our idea then was to carve out a computational hypothesis for conscious in terms of a purely temporal (time bin) resonance of memory signals in long-range neural circuits well beyond functionally identified sensory areas.

Such circuitry would first be arbitrarily selected and short-term potentiated by Hebbian synaptic learning, then subsequently consolidated during ontogenesis, long-term potentiated and dedicated to the generation of temporal signatures (firing patterns) of conscious states. Since consciousness is fully operational in the absence of spatially defined stimulus input from the outside world (when we have our eyes closed and daydream, for example, or in lucid dreaming), we daringly proposed that the temporal signatures of consciousness generated in dedicated neural circuits could become progressively de-correlated from spatial signals during ontogenesis. This hypothesis negates the concept of a spatially defined activity field within the brain as a potential correlate of consciousness, and it makes sense under the light of the fact that the phenomenon itself has no measurable spatial dimensionality. Brain activity patterns are representations, as adequately and parsimoniously defined earlier by [Churchland \(2002\)](#), in terms of patterns of activity across groups of neurons which carry information. Interpretations of such patterns in terms of functionally specific correlates (correlate = co-occurring with) of consciousness under specific conditions of testing (for a recent review, see for example, [Koch et al., 2016](#)) all carefully avoid the term causality, as correlation does not necessarily imply causality. Thus, when it comes to an *explanation* of consciousness, we are still found wanting, consistently faced with the same old problem, over and over again. It has up to now not been possible to confirm that specific brain activity patterns, or synchronization thereof, recorded during a specific conscious experience, explains consciousness, or even leads to an understanding of the phenomenon. This is so, because we do not know beyond reasonable doubt whether the brain activity patterns demonstrated in any of the relevant studies in the field are neural signatures of consciousness, or nothing more (or less) than the traces of different levels of integrated brain activity (see also the review by [Pockett, 2013](#)), representing ongoing sensations, memories, or mental images during conscious experience.

Absence of a Spatial Dimensionality of Consciousness

Consciousness has no observable spatial dimensionality (e.g., [Libet, 2004](#); [Buzsáki, 2007](#); [Marchetti, 2014](#)). It corresponds to a specific state of being in time independent of spatial location. It is a phenomenon that can be observed and accounted for scientifically across changes with time in ontogenetic development. The ability of human beings to both consciously relive past events and conceive future events ([Callender, 2010](#); [Carroll, 2010](#)) entails an active process of construction of consciousness in time that underpins many other important aspects of conscious human life. During early childhood, the brain learns about the temporal order of the physical world, well before we become phenomenally conscious of a *Self* and its immediate or distant environments ([Piaget, 1967](#)). The temporal rhythms or order of stimuli is the first way through which humans acquire knowledge of physical reality, structure, and continuity, as illustrated by results from experiments where the responses of newborns exposed to speech stream inputs have been systematically quantified ([Bulf et al., 2011](#)). Human awareness of temporality and the projection of the *Self* and its most abstract concepts and reasoning toward a future that has not yet happened occurred relatively late in evolution as a mental capacity allowing to push the technological and cultural

development of our species further and further. It develops ontogenetically over the first years of an individual lifespan ([Piaget, 1967](#); [Jaynes, 1990](#); [Edelman, 1993](#)). The idea of ontological identity link between consciousness and awareness of temporality harks back to [Hume \(1740\)](#), [Hegel \(1807\)](#), and [Heidegger \(1927\)](#) concepts of *Sein* (Being) and *Zeit* (Time), where consciousness is hardly more than a succession of psychological moments where we realize that we exist in, and are part of, moments in time. This places all other perceptual or sensorial processes which may characterize any particular phenomenal experience at a different level of analysis. The idea of a fundamental identity link between awareness of Self (*das Ich*) and awareness of what Heidegger termed *Ursprüngliche Zeit* (original time) implies that human consciousness may have progressively evolved from the primitive ability to be aware of, to remember, and to predict temporal order and change in nature found in other species such as rodents (e.g., [Fouquet et al., 2010](#)). The limits of this capacity are directly determined by brain capacity and extent of interaction between the brain and the world. The activity of individual neurons in the brain only has a small quantifiable relationship to sensory representations and motor outputs. The most recent evidence ([Ainsworth et al., 2012](#)) confirms that coactivation of a few 10s–100s of neurons can code sensory inputs and behavioral task performance within clear psychophysical limits. However, in a sea of sensory inputs with memory representations linked to complex motor output, the temporal activity of neurons has to be functionally organized independent of spatial codes relative to sensory data and representations. In the brain, this could be made possible either through spike rates (rate coding), or the selective reinforcement of spike coincidences active neurons and networks (temporal coding). Both have computational advantages and are not mutually exclusive. There is recent evidence ([Ainsworth et al., 2012](#)) for a bias in neuronal circuits toward temporal coding. Just as the temporal signal sequence or activity pattern of any single coding cell is determined by its firing activity across a certain length of time, the temporal signature of a conscious state is also linked to duration, with variations in the limited dynamic range of a few hundreds of milliseconds. Most perceptual and cognitive contents are processed implicitly by the brain (i.e., non-consciously) and truly conscious states seem to be reflected by short oscillatory activity periods of not more than a few hundreds of milliseconds each (e.g., [Buzsáki, 2007](#); [del Cul et al., 2009](#); [Nyberg et al., 2010](#)). The clockworks of consciousness have thus been conceived in terms of rapid temporal successions of microscopic brain states ([Edelman, 2003](#)). Spike time-based models of processing relating to such have been suggested (e.g., [Singer, 2000](#); [Thorpe et al., 2001](#)). The temporal characteristics of resonant brain networks, under the hypothesis of a functional separation from spatial mechanisms, can explain the temporal stability of representations as defined by [Churchland \(2002\)](#); see here above) despite the highly plastic and diffuse functional anatomy of the brain ([Wall et al., 2002](#)). Such stability is ensured by the dynamics of bottom-up patterns triggering neocortical excitatory activity and matched in time with top-down memory representations or expectation signals ([Grossberg, 1999](#); [Dresp-Langley and Durup, 2009](#); [Cacha and Poznanski, 2014](#)). The temporal dynamics of these brain activity patterns are driven by environmental pressure (relevance) toward functional interaction (e.g., [Grossberg, 1999](#); [Hari, 2017](#)). Consciousness can also be understood in terms of changes that have occurred in time during phylogenesis. All vertebrates appear to be phenomenally and affectively

conscious, function according to circadian cycles, and experience various states of being with dynamic transitions between different states of consciousness ([Birch et al., 2020](#)). In humans, however, the variety of higher-order states of consciousness described in the literature is not only larger ([Fabbro et al., 2015](#)), but also qualitatively different. So-called primary consciousness related to representations coding for perception, affect, and action ([Edelman, 2003](#)) are likely to be shared by all mammals, while higher-order consciousness linked to the interpretation of contents of primary consciousness, self-related awareness of past and future, and symbolic activities relating to language ([Kotchoubey, 2018](#)) is what makes humans unique. There is no doubt that conscious ability results from neural network development and processing resources within the physical boundaries of a brain. However, human consciousness transcends these underlying brain mechanisms. By changing and developing with experience, in interaction with other beings, society, and the physical world, it creates potential for mobilizing new resources beyond currently existing or known physical boundaries.

Consciousness Transcends Brain, Space, and Time

The capability of consciously shaping our lives within the world and of projecting them into a distant future, imagined but not yet real, is a critical aspect of fully evolved human consciousness ([Natsoulas, 1999](#)), and drives human creativity and the technological and cultural development of human societies (e.g., [Fabbro et al., 2015](#)). As pointed out earlier ([Logan, 2007](#)), when humans started living in complex communities where cooperation was a key to survival, through the maintenance of the camp fire among other things, consciousness may have emerged from a new complexity of human interaction. Consciousness may thus be seen as a specific form of energy, with a creative potential that can lead to directly observable changes, in other beings and in physical environments. Consciousness thereby has, or can have, the power to determine and/or change future non-physical (mental) states, in the *Self* and in others, and/or future physical states in the outside world. To clarify how we may link this form of energy to the brain on the one hand, and to human society and the physical world on the other, we may consider the following general definition:

“Energy is the capacity for doing work. It may exist in potential, kinetic, thermal, electrical, chemical, nuclear, or other forms” ([Encyclopedia Britannica, 2021](#)).

The origin of conscious energy is definitely the brain, its form potential, and the work it does when operational is the work of producing change, in the *Self*, in other people, and in the world. Consciousness is thereby defined as the energy source of all change and creativity. The latter can be as diverse as the things we may observe, in ourselves, in others, and in the world. Consciousness enabled human creativity breaks currently known mental and physical barriers every day in art and science, finding new solutions to problems that previously appeared insurmountable. Phenomena such as consciously guided brain-to-brain communication ([Grau et al., 2014](#)) are now being investigated in respectable research laboratories. New particles that do not seem to obey conventional

laws of physics have been discovered ([Bazavov et al., 2014](#)). Yet, our knowledge about the processes underlying this individual and collective power of consciousness is still very limited. Physical theory entails that the functioning of organisms or environments (living systems) is continuously challenged by the laws of thermodynamics in its attempt to maintain energy. Living organisms interact with each other and their environments in a continuously evolving process aimed at precisely that goal. With the evolution of the species, such interactions become increasingly sophisticated and effective. It is possible that consciousness, well beyond mere biological adaptive function (e.g., [Jordan and Ghin, 2006](#); [Kotchoubey, 2018](#)) has evolved for the purpose of extending human capacity to, not only maintain, but create new energy. Interactionism thus becomes a key to understanding consciousness in terms of a product of the “transcendental human brain” ([Cacha and Poznanski, 2014](#)) or, in other words, a specific kind of temporally defined energy that results from interaction with space, but is not spatially defined or limited. As pointed out elsewhere ([Pepperell, 2018](#)), if we want to explain consciousness as a physical process we must acknowledge that, as in all physical processes, energetic activity is fundamental also to the processes that drive evolution and have produced consciousness. The nature of energy itself can take many different forms in physics. In the case of brain function, energy may be conceived in terms of the forces that produce specific electric activity dynamics (observable as brain waves) while we are conscious. By observing physical systems, we can understand how energy produces temporal change in systemic states on the basis of the observed differences between these systemic states. Consciousness as a transcendental form of energy, produced in the brain but not limited to the latter, produces temporal change in states of the *Self*, as suggested by current insights from studies on deep meditation. Meditative practice detunes the brain processes of self-awareness and blocks the instantiation of self-referential conscious states ([Nair et al., 2017](#); [Keppler, 2018](#)). This leads to what has been called the “dissolution of the ego,” which is not to be confounded with dissolution of consciousness ([Keppler, 2018](#)). Instead, deep meditative states tap into a wider spectrum of functional brain modes, opening the gates to extended phenomenal experience and expanded consciousness. The subject/object relationship, or relationship of the *Self* to its immediate environment during such a transcendental experience ([Nair et al., 2017](#)) is characterized by absence of a conscious perception of time passing, space, or body sense, in other words anything that would give meaning to conscious waking experience. Transcendental consciousness is described as full self-awareness isolated from the processes and objects of experience but characterized by the absence of sense of time, space, the physical body or the contents of perception that define waking experiences ([Vieta et al., 2018](#); [Paoletti and Ben-Soussan, 2020](#)). The integration of transcendental experience with waking, dreaming, and sleeping has been labeled by distinct subjective and objective markers. It is subjectively marked by full self-awareness during waking, sleeping, or dreaming, greater emotional stability, decreased anxiety during challenging tasks and, physiologically, by the coexistence of specific brain wave patterns ([Travis, 2014](#)). Transcendental experience may be an engine that fuels human development and creates potential for the evolution of higher forms of human “intelligence,” to be understood here in terms of mental capability. Insights from the neuroscience of meditation and its effects on human well-being and the development of higher forms of consciousness ([Muehsam et al., 2017](#); [Mahone et](#)

[al., 2018](#); [Vieten et al., 2018](#); [Brandmeyer et al., 2019](#); [Vivot et al., 2020](#)) points toward states of enhanced consciousness in deep meditation as a crucial driver of human psychological development ([Wahbeh et al., 2018](#)). Deep meditation also has proven therapeutic effects, as in chronic pain management ([Hilton et al., 2017](#)) in individuals where consciousness is often reduced to little else than overwhelming sensations of pain, limiting the full, health expression of conscious capability in their everyday lives. One of the most controversial issues in vegetative state or a minimally conscious state syndromes concerns the potential capacity of such patients to continue to experience pain in the absence of any measurable self- or environmental awareness. Some of such patients might continue to experience elementary emotions or feelings, as suggested by results from neuroimaging studies showing activation of specific cerebral areas in response to situations which commonly generate empathy ([Pistoia et al., 2013](#)). In short, there are still more questions to be answered in consciousness research, well beyond what we call biological adaptation in the Darwinian tradition. For example, the intimate link between transcendental states of mind and a phenomenon called *eudaimonia* needs to be unraveled in well-targeted research across the human lifespan. This novel, largely uncharted terrain of scientific investigation into human consciousness may provide deeper insights into its function far more important to our species than biophysical explanations in terms of neural correlates, or biophysical fields. As pointed out by [Churchland \(2005\)](#), we have to, ultimately, ask ourselves what we want from a science of consciousness.

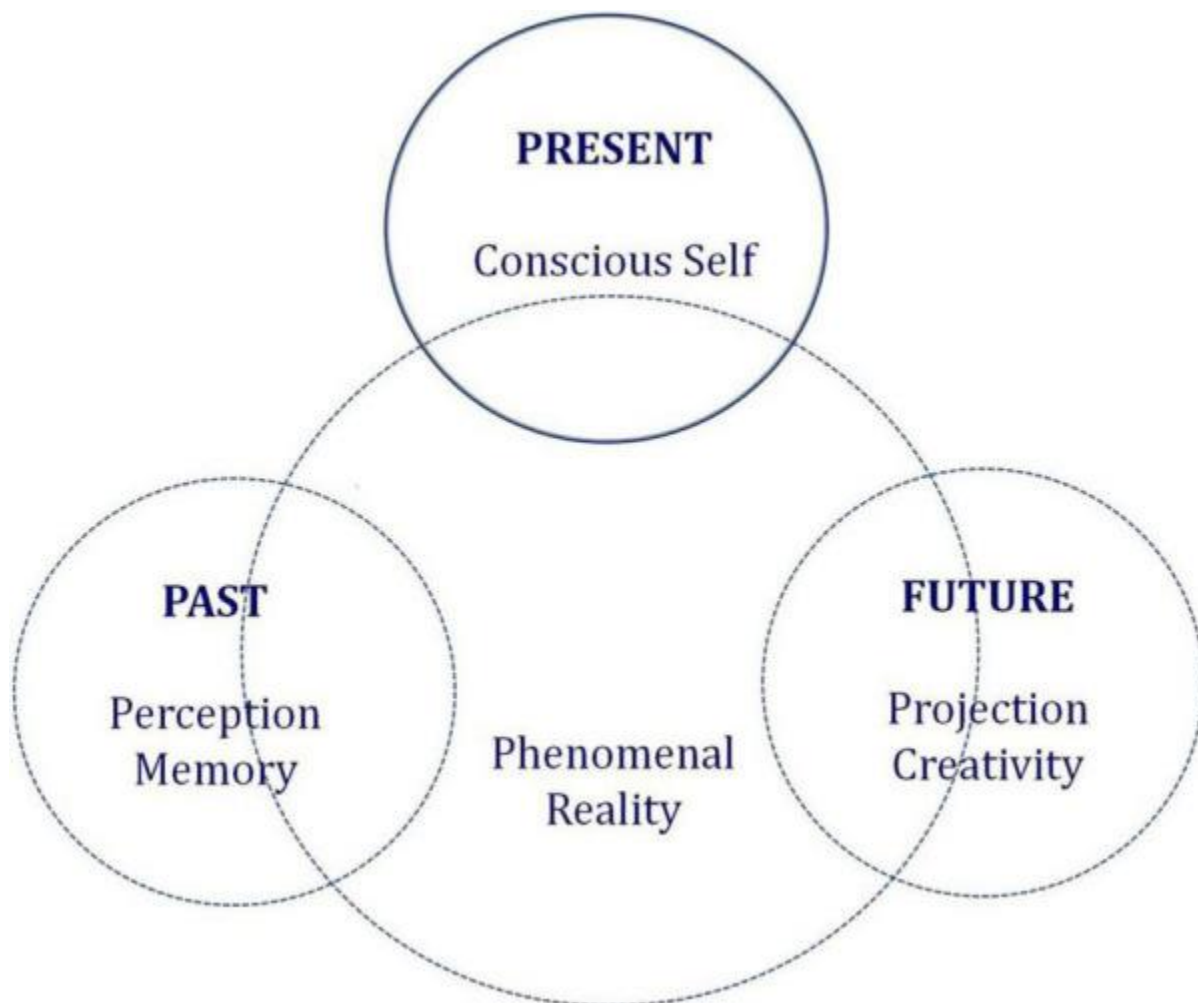
Eudaimonic Well-Being as a Window to the Purpose of Consciousness

Eudaimonia is a central concept in the Aristotelian philosophy of ethics. It is related to other concepts such as virtue, human excellence, and *phronesis*, which is an ethically grounded form of wisdom (cf. [Walsh, 2015](#)). The viewpoint on consciousness delivered here suggests potential for generating creative change, in non-physical (mental) states of the *Self*, of others, or in physical states as a neglected possibility for scientific investigation. The newly emerging field of research on eudaimonic development (e.g., [Rosenfeld, 2019](#); [Alexander et al., 2021](#)) can be linked to the clinical neuroscience of expanded consciousness ([Paulson et al., 2017a,b](#); [Vieten et al., 2018](#); [Vivot et al., 2020](#)) to generate new insights on how we can foster acceptance of what cannot be changed, in ourselves and in others. It could help us understand how expanding our consciousness can help us adjust our individual expectations, and find purpose and fulfillment despite adversity. A deeper understanding of consciousness in terms of practice (mindfulness, deep meditation) could help us understand the intuitive processes that enable us to think “outside the box,” and to transcend the limitations of preconceived knowledge. Unconscious mechanism play an essential role in these intuitive processes (e.g., [Grobstein, 2005](#); [Dresp-Langley, 2011](#); [Paulson et al., 2017a,b](#)). Eudaimonia most certainly fuels on consciousness as energy potential, as a fundamental driving force toward a better life, referring to the subjective experiences associated with living a life of virtue and purpose in the pursuit of human excellence. The clinical neurosciences have only just begun to explore the underlying psychological and physiological mechanisms and processes. The phenomenological experiences

derived from eudaimonic living include self-realization, self-actualization, personal expressiveness, optimism, vitality and, as experienced in states of deep meditation, capability of transcending the Self ([Di Fabio and Palazzeschi, 2015](#)). While hedonic well-being focuses on happiness defined in terms of pleasure and the avoidance of pain, the eudaimonic approach, on the other hand, relates to meaning and self-realization where well-being is seen as the full functioning of a person at a higher of consciousness than the one that seeks pleasure to attain happiness. A conscious Self in a state of eudaimonic well-being is focused on inner resources, resilience, higher meaning, authenticity, and purposefulness ([Di Fabio and Palazzeschi, 2015](#); [Cosco et al., 2017](#)). Consciousness changes as our brains age, with changes in challenges to meet, and in perspectives for the future. Although younger individuals may be deemed more resilient than aging ones, this is a preconception that needs to be reconsidered in the current society context under the light of new challenges and problems related to current society contexts, the internet, and social media (e.g., [Garland and Howard, 2018](#); [Dresp-Langley, 2020](#); [Liu et al., 2021](#)). In aging individuals, on the other hand, changes in embodiment occur with aging neuronal mechanisms and their associated sensorimotor and cognitive deficits ([Costello and Bloesch, 2017](#)). Investigating the affective and cognitive mechanisms of deep meditation, mindfulness, and expanded consciousness at the behavioral, metabolic, and neurobiological levels, with research into the mechanisms of action underlying expanded consciousness, will no doubt contribute to the development of treatment options beyond age-related changes in neuronal function. A wider range of comprehensive approaches needs to be developed to address the multiple mechanisms underlying the many different conditions of breakdown of human resilience (e.g., [Franklin et al., 2012](#)). Such may be, but not always, related to brain aging. While embodiment changes as we age ([Kuehn et al., 2018](#)), these changes do not inevitably produce lesser well-being ([Ryff et al., 2016](#)). Yet, bodily awareness is a central component of human consciousness ([Treves et al., 2019](#)), and clearly an important aspect of well-being while we are young ([Ginzburg et al., 2014](#)). Resilience theory (e.g., [Maltby et al., 2019](#)) reflects the general idea that managing to navigate adversity and maintaining high levels of functioning across a large number of various domains demonstrates resilience, i.e., the capacity to cope with adversity. Similarly, traditional models of healthy aging (cf. [Wong, 2018](#), for a review) suggest that having a high level of functioning across a large number of various domains would be a requirement for well-being and resilience. Yet, this may not necessarily be so. For example, health benefits have been identified among older adults who maintain a purposeful life engagement and thereby exhibit a high level of eudaimonic well-being ([Cosco et al., 2017](#)). These benefits would include extended longevity, reduced risks in various disease outcomes, reduced physiological deregulation, and gene expression linked to better inflammatory profiles ([Ryff et al., 2016](#)). Similarly, the brains of mindful or meditating individuals may be less affected by aging processes under the light of research suggesting that meditation and similar forms of mindfulness practice, or the states of expanded consciousness such practice generates, could contribute to preserving healthy brain tissue, cognitive and emotional resilience, and diminish the risk of dementia and other age-related neurodegenerative diseases ([Kurth et al., 2017](#); [Lutz et al., 2018](#)). However, as already pointed out herein, the capability of projecting its own existence into a future that has not happened yet is

the sole unique property of a fully conscious Self. Only the human primate has evolved this capability and, as postulated here, this capability is intimately linked to eudaimonic well-being as defined here above. This working assumption stems from the, previously discussed ([Dresp-Langley, 2011, 2018](#); [Dresp-Langley and Durup, 2012](#)), ontological link between consciousness and time itself, which is summarized again in the context of this paper here, for illustration, in [Figure 1](#). In the current societal context, with the many challenges we have to face and to anticipate, expanding our consciousness and exploiting it mindfully could be the key to our future as well as that of the whole planet. One of the most important functions of expanded consciousness may be its power to generate holistic resilience, to increase our potential to cope with new forms of adversity, whether we are young or old. A science of consciousness that leads to novel forms of well-being in the face of increasing adversity could have a significant impact, for individuals and for society as a whole.

FIGURE 1.



[Open in a new tab](#)

All phenomenal reality originates from the ontological link between time and consciousness (after [Dresp-Langley and Durup, 2012](#)). This leads to consider consciousness as a form of creative energy beyond space and time, where specific

cognitive abilities such as perception, memory or projective thinking and reasoning, although they may exploit conscious energy, need to be placed at a separate ontological level. An optimally expanded level of consciousness (by deep meditation or other mindfulness practice) would be equivalent to an expanded Self in a state of deep of sense of present, past, and future at one and the same moment in time. Such deep states of consciousness may not be attainable by our species.

Conclusion and Perspectives

Field theories of consciousness where the latter is seen as having duration and extension in space are limited by the fact that, unlike in the field theories in physics, particular regions of the presumed space-time continuum and interactions between elementary objects herein cannot be objectively measured, or accounted for mathematically. [Libet \(1993\)](#) was well aware of this fundamental problem by recognizing that “the mind field of consciousness” does not correspond to any category of known physical fields and cannot be observed directly by known physical means. [Pockett \(2013\)](#) wrote that “a field that is not observable directly by known physical means is in some danger of remaining confined to the realms of philosophy,” leaving it to her readers to decide whether this statement is to be considered as outrageous, or as a sign of a healthy sense of humor. Indeed, what strikes in all of the existing neural theories of consciousness including the field theories is the preconception that there has to be a “scientific” (as opposed to “philosophical”) account for consciousness. This is even more astonishing than the “astonishing hypothesis” ([Crick, 1994](#)) of a neural correlate for consciousness itself, because it means dismissing the fact that all contemporary science, including mathematics and physics, stems from nothing else but philosophy. Thus, quite ironically, the contemporary science of consciousness is based on the preconception that all reality has to be material in the sense of measurable by the known tools of physics, and that consciousness must be a direct product of physical activity in the brain. Yet, the biggest nut to crack for this kind of materialism has remained the existence of consciousness. It is the latter which has allowed to conceive all our methods and tools for scientific measurement, yet, it looks like these methods currently fail to fully explain what has made their conception possible in the first place. The full recognition that our conscious minds are not confined to our brains, or to what is currently measurable inside the brain or beyond, may eventually free contemporary neuroscience. Why look for a biophysical explanation of consciousness? Pressing society needs call for a science of consciousness in terms of its power to generate and foster eudaimonic well-being, of individuals as well as nations. Consciousness is above anything else an internal, dynamic process that governs intentionality for the purpose of creativity and social interaction and their complex relationships with the first-person perspective (e.g., [Metzinger and Gallese, 2003](#)). Thus, consciousness may have evolved to form a “mind field” that reaches beyond space and time to enable us, ultimately, to plan for both our survival, and our extinction (two essential and complementary concepts in Darwin’s theory of evolution) when survival on the planet is no longer an acceptable option for our species.

Data Availability Statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

Author Contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Conflict of Interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Ainsworth M., Lee S., Cunningham M. O., Traub R. D., Kopell N. J., Whittington M. A. (2012). Rates and rhythms: a synergistic view of frequency and temporal coding in neuronal networks. *Neuron* 75 572–583.
10.1016/j.neuron.2012.08.004 [DOI] [PubMed] [Google Scholar]
2. Alexander R., Aragón O. R., Bookwala J., Cherbuin N., Gatt J. M., Kahrilas I. J., et al. (2021). The neuroscience of positive emotions and affect: implications for cultivating happiness and wellbeing. *Neurosci. Biobehav. Rev.* 121 220–249.
10.1016/j.neubiorev.2020.12.002 [DOI] [PubMed] [Google Scholar]
3. Bazavov A., Ding H. T., Hegde P., Kaczmarek O., Karsch F., Laermann E., et al. (2014). Additional strange hadrons from QCD thermodynamics and strangeness freezeout in heavy ion collisions. *Phys. Rev. Lett.* 113:072001.
10.1103/PhysRevLett.113.072001 [DOI] [PubMed] [Google Scholar]
4. Birch J., Schnell A. K., Clayton N. S. (2020). Dimensions of animal consciousness. *Trends Cogn. Sci.* 24 789–801.
10.1016/j.tics.2020.07.007 [DOI] [PMC free article] [PubMed] [Google Scholar]
5. Block N. (2007). Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behav. Brain Sci.* 30 481–499.
10.1017/S0140525X07002786 [DOI] [PubMed] [Google Scholar]

6. Brandmeyer T., Delorme A., Wahbeh H. (2019). The neuroscience of meditation: classification, phenomenology, correlates, and mechanisms. *Prog. Brain Res.* 44 1–29. 10.1016/bs.pbr.2018.10.020 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
7. Bulf H., Johnson S. P., Valenza E. (2011). Visual statistical learning in the newborn infant. *Cognition* 121 127–132. 10.1016/j.cognition.2011.06.010 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
8. Buzsáki G. (2007). The structure of consciousness. *Nature* 446:267. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
9. Cabanac M., Cabanac A. J., Parent A. (2009). The emergence of consciousness in phylogeny. *Behav. Brain Res.* 198 267–272. 10.1016/j.bbr.2008.11.028 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
10. Cacha L. A., Poznanski R. R. (2014). Genomic instantiation of consciousness in neurons through a biophoton field theory. *J. Integr. Neurosci.* 13 253–292. 10.1142/S0219635214400081 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
11. Callender C. (2010). Is time an illusion? *Sci. Am.* 295 58–65. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
12. Carroll S. (2010). *From Eternity to Here: The Quest for the Ultimate Theory of Time*. New York, NY: Dutton/Penguin Group. [[Google Scholar](#)]
13. Carver C. S., Humphries C. (1981). Havana daydreaming: a study of self-consciousness and the negative reference group among Cuban Americans. *J. Personal. Soc. Psychol.* 40 545–552. 10.1037/0022-3514.40.3.545 [[DOI](#)] [[Google Scholar](#)]
14. Chalmers D. J. (1996). *The Conscious Mind*. Oxford: Oxford University Press. [[Google Scholar](#)]
15. Churchland P. S. (2002). *Brain-Wise. Studies in Neurophilosophy*. Cambridge, MA: MIT Press. [[Google Scholar](#)]
16. Churchland P. S. (2005). A neurophilosophical slant on consciousness research. *Prog. Brain Res.* 149 285–293. 10.1016/S0079-6123(05)49020-2 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
17. Cosco T. D., Howse K., Brayne C. (2017). Healthy ageing, resilience and wellbeing. *Epidemiol. Psychiatr. Sci.* 26 579–583. 10.1017/S2045796017000324 [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]

18. Costello M. C., Bloesch E. K. (2017). Are older adults less embodied? a review of age effects through the lens of embodied cognition. *Front. Psychol.* 8:267. 10.3389/fpsyg.2017.00267 [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
19. Crick F. (1994). *The Astonishing Hypothesis: The Scientific Search for the Soul*. New York, NY: Charles Scribner's Sons. [[Google Scholar](#)]
20. 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* 132 2531–2540. 10.1093/brain/awp111 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
21. Di Fabio A., Palazzeschi L. (2015). Hedonic and eudaimonic well-being: the role of resilience beyond fluid intelligence and personality traits. *Front. Psychol.* 6:1367. 10.3389/fpsyg.2015.01367 [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
22. Dresch-Langley B. (2011). Why the brain knows more than we do: non-conscious representations and their role in the construction of conscious experience. *Brain Sci.* 2 1–21. 10.3390/brainsci2010001 [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
23. Dresch-Langley B. (2018). The experience dependent dynamics of human consciousness. *Open J. Philos.* 8 116–143. 10.4236/ojpp.2018.82010 [[DOI](#)] [[Google Scholar](#)]
24. Dresch-Langley B. (2020). Children's Health in the digital age. *Int. J. Environ. Res. Public Health* 17:3240. [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
25. Dresch-Langley B., Durup J. (2009). A plastic temporal brain code for conscious state generation. *Neural Plasticity* 2009:482696. 10.1155/2009/482696 [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
26. Dresch-Langley B., Durup J. (2012). Does consciousness exist independently of present time and present time independently of consciousness? *Open J. Philos.* 2 45–49. [[Google Scholar](#)]
27. Edelman G. M. (1993). Neural darwinism: selection and reentrant signaling in higher brain function. *Neuron* 10, 115–125. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
28. Edelman G. M. (2003). Naturalizing consciousness: a theoretical framework. *Proc. Natl. Acad. Sci. U S A.* 100 5520–5524. 10.1073/pnas.0931349100 [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]

29. Encyclopedia Britannica (2021). *Energy*. Chicago, ILL: Britannica. [[Google Scholar](#)]
30. Fabbro F., Aglioti S. M., Bergamasco M., Clarici A., Panksepp J. (2015). Evolutionary aspects of self- and world consciousness in vertebrates. *Front. Hum. Neurosci.* 9:157. 10.3389/fnhum.2015.00157 [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
31. Feinberg T. E., Mallatt J. (2013). The evolutionary and genetic origins of consciousness in the Cambrian Period over 500 million years ago. *Front. Psychol.* 4:667. 10.3389/fpsyg.2013.00667 [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
32. Fingelkurts A. A., Fingelkurts A. A. (2002). Operational architectonics of the human brain biopotential field: towards solving the mind-brain problem. *Brain Mind.* 2 261–296. [[Google Scholar](#)]
33. Fouquet C., Tobin C., Rondi-Reig L. (2010). A new approach for modeling episodic memory from rodents to humans: the temporal order memory. *Behav. Brain Res.* 215 172–179. 10.1016/j.bbr.2010.05.054 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
34. Franklin T. B., Saab B. J., Mansuy I. M. (2012). Neural mechanisms of stress resilience and vulnerability. *Neuron* 75 747–761. 10.1016/j.neuron.2012.08.016 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
35. Garland E. L., Howard M. O. (2018). Mindfulness-based treatment of addiction: current state of the field and envisioning the next wave of research. *Addict. Sci. Clin. Pract.* 13:14. 10.1186/s13722-018-0115-3 [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
36. Gilead M., Liberman N., Maril A. (2014). From mind to matter: neural correlates of abstract and concrete mindsets. *Soc. Cogn. Affect. Neurosci.* 9 638–645. 10.1093/scan/nst031 [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
37. Ginzburg K., Tsur N., Barak-Nahum A., Defrin R. (2014). Body awareness: differentiating between sensitivity to and monitoring of bodily signals. *J. Behav. Med.* 37 564–575. 10.1007/s10865-013-9514-9 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
38. Grau C., Ginhoux R., Riera A., Nguyen T. L., Chauvat H., Berg M., et al. (2014). Conscious brain-to-brain communication in humans using non-invasive technologies. *PLoS One* 9:e105225. 10.1371/journal.pone.0105225 [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]

39. Grobstein P. (2005). Making the unconscious conscious, and vice versa: a bi-directional bridge between neuroscience/cognitive science and psychotherapy? *Cortex* 41 663–668. 10.1016/s0010-9452(08)70283-1 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
40. Grossberg S. (1999). The link between brain learning, attention, and consciousness. *Consciousness Cogn.* 8 1–44. 10.1006/ccog.1998.0372 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
41. Grossberg S. (2000). How hallucinations may arise from brain mechanisms of learning, attention, and volition. *J. Int. Neuropsychol. Soc.* 6 583–592. 10.1017/s135561770065508x [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
42. Hari R. (2017). From brain-environment connections to temporal dynamics and social interaction: principles of human brain function. *Neuron* 94 1033–1039. 10.1016/j.neuron.2017.04.007 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
43. Hegel G. W. F. (1807). *System der Wissenschaft. Erster Teil: Die Phänomenologie des Geistes*. Bamberg/Würzburg: Josef Anton Goebhardt Verlag. [[Google Scholar](#)]
44. Heidegger M. (1927). *Sein und Zeit*. Tübingen: Max Niemeyer Verlag. [[Google Scholar](#)]
45. Helekar S. A. (1999). On the possibility of universal neural coding of subjective experience. *Consciousness Cogn.* 8 423–446. 10.1006/ccog.1998.0377 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
46. Hilton L., Hempel S., Ewing B. A., Apaydin E., Xenakis L., Newberry S., et al. (2017). Mindfulness meditation for chronic pain: systematic review and meta-analysis. *Ann. Behav. Med.* 51 199–213. 10.1007/s12160-016-9844-2 [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
47. Hume D. (1740). *Treatise of Human Nature*. Oxford: Oxford University Press. [[Google Scholar](#)]
48. Jaynes J. (1990). *The Origin of Consciousness in the Breakdown of the Bicameral Mind*. Boston, MA: Houghton-Mifflin. [[Google Scholar](#)]
49. John E. R. (2002). The neurophysics of consciousness. *Brain Res. Rev.* 39 1–28. 10.1016/s0165-0173(02)00142-x [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
50. Jordan J. S., Ghin M. (2006). Proto-consciousness as a contextually emergent property of self-sustaining systems. *Mind Matter.* 4 45–68. [[Google Scholar](#)]

51. Keppler J. (2018). The role of the brain in conscious processes: a new way of looking at the neural correlates of consciousness. *Front. Psychol.* 9:1346. 10.3389/fpsyg.2018.01346 [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
52. Koch C., Massimini M., Boly M., Tononi G. (2016). Neural correlates of consciousness: progress and problems. *Nat. Rev. Neurosci.* 17, 307–321. 10.1038/nrn.2016.22. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
53. Köhler W. (1940). *Dynamics in Psychology*. New York, NY: Liveright. [[Google Scholar](#)]
54. Kotchoubey B. (2018). Human consciousness: where is it from and what is it for. *Front. Psychol.* 9:567. 10.3389/fpsyg.2018.00567 [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
55. Kuehn E., Perez-Lopez M. B., Diersch N., Döhler J., Wolbers T., Riemer M. (2018). Embodiment in the aging mind. *Neurosci. Biobehav. Rev.* 86 207–225. 10.1016/j.neubiorev.2017.11.016 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
56. Kurth F., Cherbuin N., Luders E. (2017). Promising links between meditation and reduced (Brain) aging: an attempt to bridge some gaps between the alleged fountain of youth and the youth of the field. *Front. Psychol.* 8:860. 10.3389/fpsyg.2017.00860 [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
57. LaBerge S. (1990). “Lucid dreaming: psychophysiological studies of consciousness during REM sleep,” in *Sleep and Cognition*, eds Bootzen R. R., Kihlstrom J. F., Schacter D. L. (Washington, DC: APA press;), 109–126. 10.1016/j.tics.2021.04.004 [[DOI](#)] [[Google Scholar](#)]
58. LaBerge S., Levitan L., Dement W. C. (1986). Lucid dreaming: physiological correlates of consciousness during REM sleep. *J. Mind Behav.* 7 251–258. 10.1093/sleep/32.9.1191 [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
59. Lashley K. S., Chow K. L., Semmes J. (1951). An examination of the electric field theory of cerebral integration. *Psychol. Rev.* 58 123–136. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
60. Libet B. (1993). “Experimental and theoretical studies of consciousness,” in *The Neural Time Factor in Conscious and Unconscious Events* (New York, NY: John Wiley & Sons;) 282–303. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
61. Libet B. (2004). *Mind Time*. Cambridge, MA: Harvard University Press. [[Google Scholar](#)]

62. Liu X., Jiang J., Zhang Y. (2021). Effects of logotherapy-based mindfulness intervention on internet addiction among adolescents during the COVID-19 pandemic. *Iran J. Public Health* 50 789–797. 10.18502/ijph.v50i4.6005 [\[DOI\]](#) [\[PMC free article\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
63. Logan R. K. (2007). *The Extended Mind: The Emergence of Language: The Human Mind and Culture*. Toronto, ON: University of Toronto Press. [\[Google Scholar\]](#)
64. Lutz A., Klimecki O. M., Collette F., Poisnel G., Arenaza-Urquijo E., Marchant N. L., et al. (2018). The age-well observational study on expert meditators in the medit-ageing european project. *Alzheimers Dement (N Y)* 4 756–764. 10.1016/j.trci.2018.11.002 [\[DOI\]](#) [\[PMC free article\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
65. Mahone M. C., Travis F., Gevirtz R., Hubbard D. (2018). fMRI during transcendental meditation practice. *Brain Cogn.* 123 30–33. 10.1016/j.bandc.2018.02.011 [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
66. Maltby J., Day L., Hall S. S., Chivers S. (2019). The measurement and role of ecological resilience systems theory across domain-specific outcomes: the domain-specific resilient systems scales. *Assessment* 26 1444–1461. 10.1177/1073191117738045 [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
67. Marchetti G. (2014). Attention and working memory: two basic mechanisms for constructing temporal experiences. *Front. Psychol.* 5:880. 10.3389/fpsyg.2014.00880 [\[DOI\]](#) [\[PMC free article\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
68. McFadden J. (2002). The conscious electromagnetic information field theory: the hard problem made easy? *J. Consciousness Stud.* 9 45–60. [\[Google Scholar\]](#)
69. Metzinger T., Gallese V. (2003). The emergence of a shared action ontology: building blocks for a theory. *Conscious. Cogn.* 12 549–571. 10.1016/s1053-8100(03)00072-2 [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
70. Muehsam D., Lutgendorf S., Mills P. J., Rickhi B., Chevalier G., Bat N., et al. (2017). The embodied mind: a review on functional genomic and neurological correlates of mind-body therapies. *Neurosci. Biobehav. Rev.* 73 165–181. 10.1016/j.neubiorev.2016.12.027 [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
71. Nair A. K., Sasidharan A., John J. P., Mehrotra S., Kutty B. M. (2017). Just a minute meditation: rapid voluntary conscious state shifts in long term meditators. *Conscious. Cogn.* 53 176–184. 10.1016/j.concog.2017.06.002 [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)

72. Nani A., Manuello J., Mancuso L., Liloia D., Costa T., Cauda F. (2019). The neural correlates of consciousness and attention: two sister processes of the brain. *Front. Neurosci.* 13:1169. 10.3389/fnins.2019.01169 [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
73. Natsoulas T. (1999). A rediscovery of presence. *J. Mind Behav.* 20 17–42. [[Google Scholar](#)]
74. Nyberg L., Kim A. S. N., Habib R., Levine B., Tulving E. (2010). Consciousness of subjective time in the brain. *Proc. Natl. Acad. Sci. U S A.* 107 22356–22359. 10.1073/pnas.1016823108 [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
75. Paoletti P., Ben-Soussan T. D. (2020). Reflections on inner and outer silence and consciousness without contents according to the sphere model of consciousness. *Front. Psychol.* 11:01807. 10.3389/fpsyg.2020.01807 [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
76. Paulson S., Berlin H. A., Ginot E., Makari G. (2017a). Delving within: the new science of the unconscious. *Ann. N. Y. Acad. Sci.* 1406 12–27. 10.1111/nyas.13390 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
77. Paulson S., Hustvedt S., Solms M., Shamdasani S. (2017b). The deeper self: an expanded view of consciousness. *Ann. N. Y. Acad. Sci.* 1406 46–63. 10.1111/nyas.13403 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
78. Pepperell R. (2018). Consciousness as a physical process caused by the organization of energy in the brain. *Front. Psychol.* 9:2091. 10.3389/fpsyg.2018.02091 [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
79. Piaget J. (1967). *La Construction du Réel Chez L'enfant*. Neuchâtel: Delachaux et Niestlé. [[Google Scholar](#)]
80. Pistoia F., Sacco S., Sarà M., Carolei A. (2013). The perception of pain and its management in disorders of consciousness. *Curr. Pain Headache Rep.* 17:374. 10.1007/s11916-013-0374-3 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
81. Pockett S. (2004). Does consciousness cause behaviour? *J. Conscious. Stud.* 11 23–40. [[Google Scholar](#)]
82. Pockett S. (2013). Field theories of consciousness. *Scholarpedia* 8:4951. 10.4249/scholarpedia.4951 [[DOI](#)] [[Google Scholar](#)]
83. Rees G., Kreiman G., Koch C. (2002). Neural correlates of consciousness in humans. *Nat. Rev. Neurosci.* 3 261–270. 10.1038/nrn783 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]

84. Rosenfeld A. J. (2019). The neuroscience of happiness and well-being: what brain findings from optimism and compassion reveal. *Child Adolesc. Psychiatr. Clin. N. Am.* 28 137–146. 10.1016/j.chc.2018.11.002 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
85. Ryff C. D., Heller A. S., Schaefer S. M., van Reekum C., Davidson R. J. (2016). Purposeful engagement, healthy aging, and the brain. *Curr. Behav. Neurosci. Rep.* 3 318–327. 10.1007/s40473-016-0096-z [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
86. Searle J. R. (1998). How to study consciousness scientifically. *Philos. Trans. R. Soc. B* 353 1935–1942. [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
87. Sheldrake R. (2013). Setting science free from materialism. *Explore (NY)* 9 211–218. 10.1016/j.explore.2013.04.004 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
88. Singer J. L. (1975). Navigating the stream of consciousness: research in daydreaming and related inner experience. *Am. Psychol.* 30 727–738. [[Google Scholar](#)]
89. Singer W. (2000). “Phenomenal awareness and consciousness from a neurobiological perspective,” in *Neural Correlates of Consciousness: Empirical and Conceptual Questions*, ed. Metzinger T. (Cambridge, MA: MIT Press;), 121–137. [[Google Scholar](#)]
90. Thorpe S., Delorme A., Van Rullen R. (2001). Spike-based strategies for rapid processing. *Neural Networks* 14 715–725. 10.1016/s0893-6080(01)00083-1 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
91. Tononi G., Edelman G. M. (1998). Consciousness and complexity. *Science* 282 1846–1851. 10.3390/e23030293 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
92. Travis F. (2014). Transcendental experiences during meditation practice. *Ann. N. Y. Acad. Sci.* 1307 1–8. 10.1111/nyas.12316 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
93. Treves I. N., Tello L. Y., Davidson R. J., Goldberg S. B. (2019). The relationship between mindfulness and objective measures of body awareness: a meta-analysis. *Sci. Rep.* 9:17386. 10.1038/s41598-019-53978-6 [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
94. Vieten C., Wahbeh H., Cahn B. R., MacLean K., Estrada M., Mills P., et al. (2018). Future directions in meditation research: recommendations for

expanding the field of contemplative science. *PLoS One* 13:e0205740. 10.1371/journal.pone.0205740 [\[DOI\]](#) [\[PMC free article\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)

95. Vivot R. M., Pallavicini C., Zamberlan F., Vigo D., Tagliazucchi E. (2020). Meditation increases the entropy of brain oscillatory activity. *Neuroscience* 431 40–51. 10.1016/j.neuroscience.2020.01.033 [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
96. von der Malsburg C. (1997). “The coherence definition of consciousness,” in *Cognition, Computation and Consciousness*, eds Ito M., Miyashita Y., Rolls E. T. (Oxford: Oxford University Press;), 193–204. 10.1037/10247-013 [\[DOI\]](#) [\[Google Scholar\]](#)
97. Wahbeh H., Sagher A., Back W., Pundhir P., Travis F. (2018). A systematic review of transcendent states across meditation and contemplative traditions. *Explore (NY)* 14 19–35. 10.1016/j.explore.2017.07.007 [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
98. Wall J. T., Xu J., Wang X. (2002). Human brain plasticity: an emerging view of the multiple substrates and mechanisms that cause cortical changes and related sensory dysfunctions after injuries of sensory inputs from the body. *Brain Res. Rev.* 39 181–215. 10.1016/s0165-0173(02)00192-3 [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
99. Walsh R. (2015). What is wisdom? cross-cultural and cross-disciplinary syntheses. *Rev. General Psychol.* 19:278. 10.1037/gpr0000045 [\[DOI\]](#) [\[Google Scholar\]](#)
100. Wong R. Y. (2018). A new strategic approach to successful aging and healthy aging. *Geriatrics (Basel)* 3:86. 10.3390/geriatrics3040086 [\[DOI\]](#) [\[PMC free article\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)

The Neuroscience of Consciousness

First published Tue Oct 9, 2018; substantive revision Wed Apr 3, 2024

Conscious experience in humans depends on brain activity, so neuroscience will contribute to explaining consciousness. What would it be for neuroscience to explain consciousness? How much progress has neuroscience made in doing so? What challenges does it face? How can it meet those challenges? What is the philosophical significance of its findings? This entry addresses these and related questions.

To bridge the gulf between brain and consciousness, we need neural data, computational and psychological models, and philosophical analysis to identify principles to connect brain activity to conscious experience in an illuminating way. This entry will focus on identifying such principles without shying away from the neural details. The theories and data to be considered will be organized around constructing answers to two questions (see [section 1.4](#) for more precise formulations):

- Generic Consciousness: How might neural properties explain when a state is conscious rather than not?
- Specific Consciousness: How might neural properties explain what the content of a conscious state is?

A challenge for an objective science of consciousness is to dissect an essentially subjective phenomenon. As investigators cannot experience another subject's conscious states, they rely on the subject's observable behavior to track consciousness. Priority is given to a subject's introspective reports as these express the subject's take on her experience. Introspection thus provides a fundamental way, perhaps *the* fundamental way, to track consciousness. That said, consciousness pervasively influences human behavior and affects physiological responses, so other forms of behavior and physiological data beyond introspective reports provide a window on consciousness. How to leverage disparate evidence is a central issue.

The term “neuroscience” covers those scientific fields whose explanations advert to the properties of neurons, populations of neurons, or larger parts of the nervous system.^[1] This includes, but is not limited to, psychologists' and cognitive neuroscientists' use of various neuroimaging methods to monitor the activity of tens of millions of neurons, computational theorists' modeling of biological and artificial neural networks, neuroscientists' use of electrodes inserted into brain tissue to record neural activity from individual or populations of neurons, and clinicians' study of patients with altered conscious experiences in light of damage to brain areas. Given the breadth of neuroscience so conceived, this review focuses mostly on cortical activity that sustains perceptual consciousness, with emphasis on vision. This is not because visual consciousness is more important than other forms of consciousness. Rather, the level of detail in empirical work on vision often speaks more comprehensively to the issues that we shall confront.

That said, there are many forms of consciousness that we will not discuss. Some are covered in other entries such as split-brain phenomena (see the entry on [the unity of consciousness](#), section 4.1.1), animal consciousness (see the entry on [animal consciousness](#)), and neural correlates of the will and agency (see the entry on [agency](#), section 5). In addition, this entry will not discuss the neuroscience of consciousness in audition, olfaction, or gustation; disturbed consciousness in mental disorders such as schizophrenia; conscious aspects of pleasure, pain and the emotions; the phenomenology of thought; the neural basis of dreams; and modulations of consciousness during sleep and anesthesia among other issues. These are important topics, and the principles and approaches highlighted in this discussion will apply to many of these domains.

1. Fundamentals

1.1 A Map of the Brain

The brain can be divided into the cerebral *cortex* and the *subcortex*. The cortex is divided into two hemispheres, left and right, each of which can be divided into four lobes: frontal, parietal, temporal and occipital.

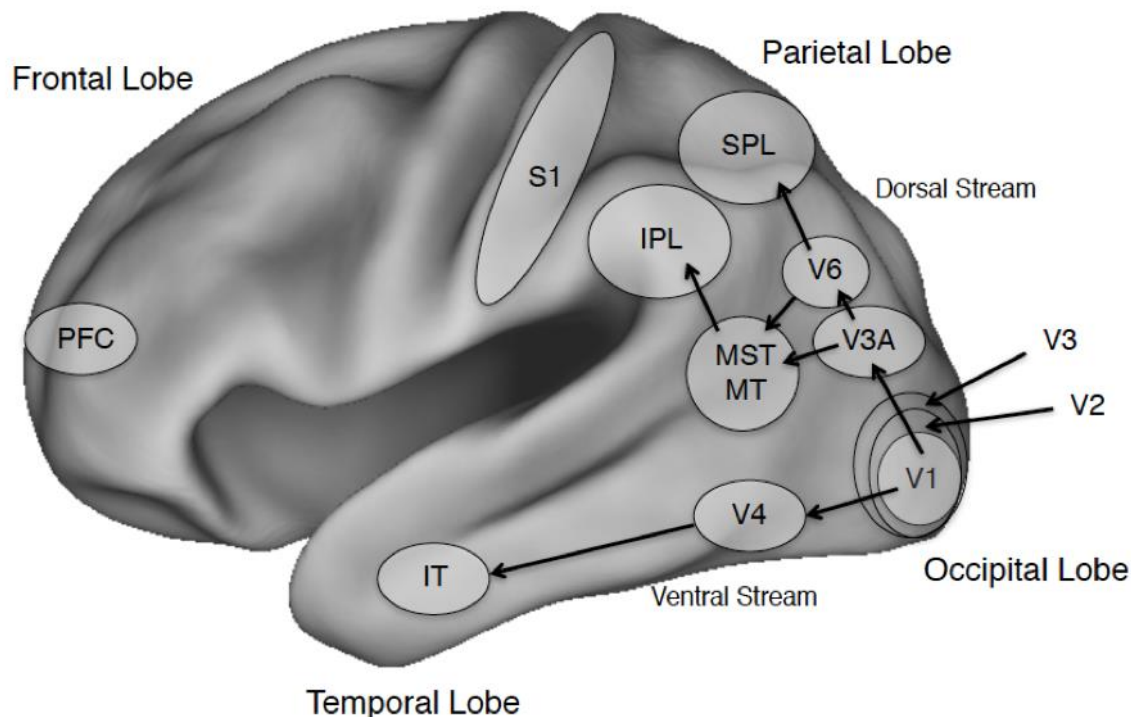


FIGURE 1. THE CEREBRAL CORTEX AND SALIENT AREAS

Figure Legend: The four lobes of the primate brain, shown for the left hemisphere. Some areas of interest are highlighted. Abbreviations: PFC: prefrontal cortex; IT: inferotemporal cortex; S1: primary somatosensory

cortex; IPL and SPL: Inferior and Superior Parietal Lobule; MST: medial superior temporal visual area; MT: middle temporal visual area (also called V5 in humans); V1: primary visual cortex; V2-V6 consist of additional visual areas.

The discussion that follows will highlight specific areas of cortex including the *prefrontal cortex* that will figure in discussions of confidence ([section 2.3](#)), the global neuronal workspace ([section 3.1](#)) and higher order theories ([section 3.3](#)); the *dorsal visual stream* that projects into parietal cortex and the *ventral visual stream* that projects into temporal cortex including visual areas specialized for processing places, faces, and word forms (see sections [2.5](#) on places, [4.1](#) on visual agnosia and [5.3](#) on seeing words); *primary somatosensory cortex* S1 (see section [5.3.2](#) on tactile sensation); and early visual areas in the occipital cortex including the *primary visual area*, V1 (see sections [4.2](#) on blindsight and [5.2](#) on binocular rivalry) and a motion sensitive area V5, also known as the *middle temporal area* (MT; [section 5.3.1](#) on seeing motion). Beneath the cortex is the subcortex, divided into the forebrain, midbrain, and hindbrain, which covers many regions although our discussion will largely touch on the *superior colliculus* and the *thalamus*. The latter is critical for regulating wakefulness and general arousal, and both areas play an important role in visual processing.

1.2 Neurons and Brain

A neuroscientific explanation of consciousness adduces properties of the brain, typically the brain's electrical properties. A salient phenomenon is neural signaling through *action potentials* or *spikes*. A spike is a large change in electrical potential across a neuron's cellular membrane which can be transmitted between neurons that form a neural circuit. For a sensory neuron, the spikes it generates are tied to its *receptive field*. For example, in a visual neuron, its receptive field is understood in spatial terms and corresponds to that area of external space where an appropriate stimulus triggers the neuron to spike. Given this correlation between stimulus and spikes, the latter carries information about the former. Information processing in sensory systems involves processing of information regarding stimuli within receptive fields.

Which electrical property provides the most fruitful explanatory basis for understanding consciousness remains an open question. For example, when looking at a single neuron, neuroscientists are not interested in spikes per se but the spike rate generated by a neuron per unit time. Yet spike rate is one among many potentially relevant neural properties. Consider the blood oxygen level dependent signal (BOLD) measure in functional magnetic resonance imaging (fMRI). The BOLD signal is a measure of changes in blood flow in the brain when neural tissue is active and is postulated to be a function of post-synaptic activity while spikes are tied to presynaptic activity. Furthermore, neuroscientists are often not interested in the response of a single neuron but rather that of a population of neurons, of whole brain regions, and/or their interactions. Higher order properties of brain regions include the *local field potential* generated by populations of neurons and correlated activity such as *synchrony* between activity in different areas of the brain as measured by, for example, electrocorticography (EEG).

The number of neural properties potentially relevant to explaining mental phenomena is dizzying. This review focuses on the facts that neural sensory systems carry information about the subject's environment and that neural information processing can be tied to a notion of *neural representation*. How precisely to understand neural representation is itself a vexed question (Cao 2012, 2014; Shea 2014; Baker, Lansdell, & Kording 2022), but we will deploy a simple assumption with respect to spikes which can be reconfigured for other properties: where a sensory neuron generates spikes when a stimulus is placed in its receptive field, the spikes carry information about the stimulus. The sensory neuron's activity thus *represent* the relevant aspect of the stimulus that drives its response (e.g., direction of motion or intensity of a sound).^[2] We shall return to neural representation in the final section when discussing how neural representations might explain conscious contents.

1.3 Access Consciousness and Phenomenal Consciousness

An important distinction separates *access consciousness* from *phenomenal consciousness* (Block 1995). "Phenomenal consciousness" refers to those properties of experience that correspond to *what it is like* for a subject to have those experiences (Nagel 1974 and the entry on [qualia](#)). These features are apparent to the subject from the inside, so tracking them arguably depends on one's having the relevant experience. For example, one understands what it is like to see red only if one has visual experiences of the relevant type (Jackson 1982).

As noted earlier, introspection is the first source of evidence about consciousness. Introspective reports bridge the subjective and objective. They serve as a behavioral measure that expresses the subject's own take on what it is like for them in having an experience. While there have been recent concerns about the reliability or empirical usefulness of introspection (Schwitzgebel 2011; Irvine 2012a), there are plausibly many contexts where introspection is reliable (Spener 2015; Wu 2023: chap. 7; see Irvine 2012b for an extended discussion of introspection in consciousness science; for philosophical theories, see Smithies & Stoljar 2012).

Introspective reports demonstrate that the subject can access the targeted conscious state. That is, the state is *access-conscious*: it is accessible for use in reasoning, report, and the control of action. Talk of access-consciousness must keep track of the distinction between *actual* access versus *accessibility*. When one reports on one's conscious state, one accesses the state. Thus, access consciousness provides much of the evidence for empirical theories of consciousness. Still, it seems plausible that a state can be conscious even if one does not access it in report so long as that state is *accessible*. One *can* report it. Access-consciousness is usually defined in terms of this dispositional notion of accessibility.

We must also consider the type of access/accessibility. Block's original characterization of access-consciousness emphasized accessibility in terms of the *rational* control of behavior, so we can summarize his account as follows:

A representation is access-conscious if it is poised for free use in reasoning and for direct “rational” control of action and speech.

Rational access contrasts with a broader conception of *intentional* access that takes a mental state to be access-conscious if it can inform goal-directed or intentional behavior including behavior that is not rational or done for a reason. This broader notion allows for additional measurable behaviors as relevant in assessing phenomenal consciousness especially in non-linguistic animals, non-verbal infants, and non-communicative patients. So, if access provides us with evidence for phenomenal consciousness, this can be (a) through introspective reports; (b) through rational behavior, (c) through intentional behavior including nonrational behavior. Indeed, in certain contexts, reflexive behavior and autonomic physiological measures provide measures of consciousness ([section 2.2](#)).

1.4 Generic and Specific Consciousness

There are two ways in which consciousness is understood in this entry. The first focuses on a mental state’s *being conscious* in general as opposed to not being conscious. Call this property *generic consciousness*, a property shared by specific conscious states such as seeing a red rose, feeling a touch, or being angry. Thus:

Generic Consciousness: What conditions/states N of nervous systems are necessary and/or sufficient for a mental state, M , to be conscious as opposed to not?

If there is such an N , then the presence of N entails that an associated mental state M is conscious and/or its absence entails that M is unconscious.

A second focus will be on the content of consciousness, say that associated with a perceptual experience’s being of some perceptible X . This yields a question about specific contents of consciousness such as experiencing the motion of an object (see [section 5.3.1](#)) or a vibration on one’s finger (see [section 5.3.2](#)):

Specific Consciousness: What neural states or properties are necessary and/or sufficient for a conscious perceptual state to have content X rather than Y ?

In introspectively accessing their conscious states, a subject reports what their experience is like by reporting what they experiences. Thus, the subject can report seeing an object moving, changing color, or being of a certain kind (e.g., a mug) and thus specify the content of the perceptual state. Discussion of specific consciousness will focus on perceptual states described as consciously perceiving X where X can be a particular such as a face or a property such as the color of an object.

Posing a clear question involves grasping its possible answers and in science, this is informed by identifying experiments that can provide evidence for such answers. The emphasis on necessary and sufficient conditions in our two questions indicates how to empirically test specific proposals. To test sufficiency, one would aim to produce or modulate a certain neural state and then demonstrate that consciousness of a certain form arises. To test necessity, one would eliminate a certain neural state and demonstrate that consciousness is abolished. Notice that such tests go beyond mere correlation between neural

states and conscious states (see [section 1.6](#) on neural correlates and sections [2.2](#), [4](#) and [5](#) for tests of necessity and sufficiency).

In many experimental contexts, the underlying idea is *causal* necessity and sufficiency. However, if $A = B$, then A 's presence is also necessary and sufficient for B 's presence since they are identical. Thus, a brain lesion that eliminates N and thereby eliminates conscious state S might do so either because N is causally necessary for S or because $N = S$. An intermediate relation is that N *constitutes* or *grounds* S which does not imply that $N = S$ (see the entry on [metaphysical grounding](#)). Whichever option holds for S , the first step is to find N , a neural correlate of consciousness ([section 1.6](#)).

In what follows, to explain generic consciousness, various global properties of neural systems will be considered ([section 3](#)) as well as specific anatomical regions that are tied to conscious versus unconscious vision as a case study ([section 4](#)). For specific consciousness, fine-grained manipulations of neural representations will be examined that plausibly shift and modulate the contents of perceptual experience ([section 5](#)).

1.5 The Hard Problem

David Chalmers presents the hard problem as follows:

It is undeniable that some organisms are subjects of experience. But the question of how it is that these systems are subjects of experience is perplexing. Why is it that when our cognitive systems engage in visual and auditory information-processing, we have visual or auditory experience: the quality of deep blue, the sensation of middle C? How can we explain why there is something it is like to entertain a mental image, or to experience an emotion? It is widely agreed that experience arises from a physical basis, but we have no good explanation of why and how it so arises. Why should physical processing give rise to a rich inner life at all? It seems objectively unreasonable that it should, and yet it does. If any problem qualifies as *the* problem of consciousness, it is this one. (Chalmers 1995: 212)

The Hard Problem can be specified in terms of generic and specific consciousness (Chalmers 1996). In both cases, Chalmers argues that there is an inherent limitation to empirical explanations of phenomenal consciousness in that empirical explanations will be fundamentally either structural or functional, yet phenomenal consciousness is not reducible to either. This means that there will be something that is left out in empirical explanations of consciousness, a missing ingredient (see also the *explanatory gap* [Levine 1983]).

There are different responses to the hard problem. One response is to sharpen the explanatory targets of neuroscience by focusing on what Chalmers calls *structural features* of phenomenal consciousness, such as the spatial structure of visual experience, or on the contents of phenomenal consciousness. When we assess explanations of specific contents of consciousness, these focus on the neural representations that fix conscious contents. These explanations leave open exactly what the secret ingredient is that shifts a state with that content from unconsciousness to consciousness. On ingredients explaining generic consciousness, a variety of options have been proposed (see [section 3](#)), but it is unclear whether these answer the Hard Problem, especially if any answer to that the Problem has a

necessary condition that the explanation must conceptually close off certain possibilities, say the possibility that the ingredient could be added yet consciousness not ignite as in a zombie, a creature without phenomenal consciousness (see the entry on [zombies](#)). Indeed, some philosophers deny the hard problem (see Dennett 2018 for a recent statement). Patricia Churchland urges: “Learn the science, do the science, and see what happens” (Churchland 1996: 408).

Perhaps the most common attitude for neuroscientists is to set the hard problem aside. Instead of explaining the existence of consciousness in the biological world, they set themselves to explaining generic consciousness by identifying neural properties that can turn consciousness on and off and explaining specific consciousness by identifying the neural representational basis of conscious contents.

1.6 Neural Correlates of Consciousness

Modern neuroscience of consciousness has attempted to explain consciousness by focusing on *neural correlates of consciousness* or NCCs (Crick & Koch 1990; LeDoux, Michel, & Lau 2020; Morales & Lau 2020). Identifying correlates is an important first step in understanding consciousness, but it is an early step. After all, correlates are not necessarily explanatory in the sense of answering specific questions posed by neuroscience. That one does not want a mere correlate was recognized by Chalmers who defined an NCC as follows:

An NCC is a minimal neural system N such that there is a mapping from states of N to states of consciousness, where a given state of N is sufficient under conditions C , for the corresponding state of consciousness. (Chalmers 2000: 31)

A similar way of putting this is that an NCC is “the minimal set of neuronal events and mechanisms jointly sufficient for a specific conscious percept” (Koch 2004: 16). One wants a minimal neural system since, crudely put, the brain is sufficient for consciousness but to point this out is hardly to explain consciousness even if it provides an answer to questions about sufficiency. There is, of course, much more to be said that is informative even if one does not drill down to a “minimal” neural system which is tricky to define or operationalize (see Chalmers 2000 for discussion; for criticisms of the NCC approach, see Noë & Thompson 2004; for criticisms of Chalmers’ definition, see Fink 2016).

The emphasis on sufficiency goes beyond mere correlation, as neuroscientists aim to answer more than the question: What is a neural correlate for conscious phenomenon C ? For example, Chalmers’ and Koch’s emphases on sufficiency indicate that they aim to answer the question: What neural phenomenon is sufficient for consciousness? Perhaps more specifically: What neural phenomenon is causally sufficient for consciousness? Accordingly, talk of “correlate” is unfortunate since sufficiency implies correlation but not vice versa. After all, assume that the NCC is type identical to a conscious state. Then many neural states will correlate with the conscious state: (1) the NCC’s typical effects, (2) its typical causes, and (3) states that are necessary for the NCCs obtaining (e.g., the presence of sufficient oxygen). Thus, some correlated effects will not be explanatory. For example, citing the effects of consciousness will not provide causally sufficient conditions for consciousness.

Establishing necessary conditions for consciousness is also difficult. Neuroplasticity, redundancy, and convergent evolution make necessity claims extremely hard to support experimentally. Under normal conditions, healthy humans may require certain brain areas or processes for supporting consciousness. However, this does not mean those regions or processes are necessary in any strong metaphysical sense. For example, after a lesion, the brain's functional connections may change allowing a different structural support for consciousness to emerge. There is also nothing in the cortical arrangement of specialized regions that fix them to a particular function. The so-called "visual cortex" is recruited in blind individuals to perform auditory, numerical, and linguistic processing (Bedny 2017). Similarly, as with many other complex structures, the brain is highly redundant. Different areas may perform the same function, preventing any single one from being strictly necessary. Finally, the way in which the mammalian brain operates is not the only way to support awareness. Birds—as well as cephalopods and insects (Barron & Klein 2016; Birch, Schnell, & Clayton 2020)—and even other primates have different anatomical and functional neural mechanisms and yet their nervous systems may support consciousness (Nieder, Wagener, & Rinnert 2020). Thus, there may not be a single necessary structure or process for supporting conscious awareness.

While many theorists are focused on explanatory correlates, it is not clear that the field has always grasped this, something recent theorists have been at pains to emphasize (Aru et al. 2012; Graaf, Hsieh, & Sack 2012). In other contexts, neuroscientists speak of the neural *basis* of a phenomenon where the basis does not simply correlate with the phenomenon but also explains and possibly grounds it. However, talk of correlates is entrenched in the neuroscience of consciousness, so one must remember that the goal is to find the subset of neural correlates that are explanatory, in answering concrete questions. Reference to neural correlates in this entry will always mean neural explanatory correlates of consciousness (on occasion, we will speak of these as the neural *basis* of consciousness). That is, our two questions about specific and generic consciousness focus the discussion on neuroscientific theories and data that contribute to explaining them. This project allows that there are limits to neural explanations of consciousness, precisely because of the explanatory gap (Levine 1983).

2. Methods for Tracking Consciousness

Since studying consciousness requires that scientists track its presence, it will be important to examine various methods used in neuroscience to isolate and probe conscious states.

2.1 Introspection and Report

Scientists primarily study phenomenal consciousness through subjective reports; objective measures such as performance in a task are often used too (for a critical assessment, see Irvine 2013). We can treat reports in neuroscience as conceptual in that they express how the subject recognizes things to be, whether regarding what they perceive (perceptual or observational reports, as in psychophysics) or regarding what mental states they are in (introspective reports). A report's conceptual content can be conveyed in words or other

overt behavior whose significance is fixed within an experiment (e.g., pressing a button to indicate that a stimulus is present or that one sees it). Subjective reports of conscious states draw on distinctively first-personal access to that state. The subject *introspects*.

Introspection raises questions that science has only recently begun to address systematically in large part because of longstanding suspicion regarding introspective methods or, in contrast, because of an unquestioning assumption that introspection is largely and reliably accurate. Early modern psychology relied on introspection to parse mental processes but ultimately abandoned it due to worries about introspection's reliability (Feest 2012; Spener 2018). Introspection was judged to be an unreliable method for addressing questions about mental processing (and it is still seen with suspicion by some; Schwitzgebel 2008). To address these worries, we must understand how introspection works, but unlike many other psychological capacities, detailed models of introspection of consciousness are hard to develop (Feest 2014; for theories of introspecting propositional attitudes, see Nichols & Stich 2003; Heal 1996; Carruthers 2011). This makes it difficult to address long-standing worries about introspective reliability regarding consciousness.

In science, questions raised about the reliability of a method are answered by calibrating and testing the method. This calibration has not been done with respect to the type of introspection commonly practiced by philosophers. Such introspection has revealed many phenomenal features that are targets of active investigation such as the phenomenology of mineness (Ehrsson 2009), sense of agency (Bayne 2011; Vignemont & Fournier 2004; Marcel 2003; Horgan, Tienson, & Graham 2003), transparency (Harman 1990; Tye 1992), self-consciousness (Kriegel 2003: 122), cognitive phenomenology (Bayne & Montague 2011); phenomenal unity (Bayne & Chalmers 2003) among others. A scientist might worry that philosophical introspection merely recycles rejected methods of a century ago, indeed without the stringent controls or training imposed by earlier psychologists. How can we ascertain and ensure the reliability of introspection in the empirical study of consciousness?

Model migration, applying well-understood concepts in less well-understood domains, provides conceptual resources to reveal patterns across empirical systems and to promote theoretical insights (Lin 2018; Knuuttila & Loettgers 2014; see the entry on [models in science](#)). Introspection's range of operation and its reliability conditions—when and why it succeeds and when and why it fails—can be calibrated by drawing parallels to how Signal Detection Theory (SDT) models perception (see Dołęga 2023 for an evaluation of introspective theories). In SDT (Tanner & Swets 1954; Hautus, Macmillan, & Creelman 2021), perceptually detecting or discriminating a stimulus is the joint outcome of the observer's perceptual sensitivity (e.g., how well a perceptual system can tell signal from noise) and a decision to classify the available perceptual evidence as signal or not (i.e., observers set a response criterion for what counts as signal). Consider trying to detect something moving in the brush at twilight versus at noon. In the latter, the signal will be greatly separated from noise (the object will be easier to detect because it generates a strong internal perceptual response). In contrast, at twilight the signal will not be easy to disentangle from noise (the object will be harder to detect because of the weaker internal perceptual response it generates in the observer). Importantly, even at noon when the signal is strong, rare as they might be, misses and false alarms are to be expected. Yet in either case, one

might operate with a conservative response criterion, say because one is afraid to be wrong. Thus, even if the signal is detectable, one might still opt not to report on it given a *conservative* bias (criterion), say if one is in the twilight scenario and would be ridiculed for “false alarms”, i.e., claiming the object to be present when it is not.

Introspection can be modeled as a signal detection mechanism that operates over conscious states. Pains can be strong or weak, mental images can be vivid or faint, and perceptions can be more or less striking. In other words, conscious experiences admit degrees of intensity (Lee 2023; Morales 2023). According to introspective Signal Detection Theory or iSDT (Morales forthcoming), the intensity of conscious experiences (their mental strength) modulates the introspective internal response generated by the intensity of the conscious experience under examination, which in turn modulates introspective sensitivity. Everything else being equal, introspectors are more likely to introspect accurately an intense experience (e.g., a strong pain, a vivid mental image, etc.) than a weak experience (e.g., a mild pain, a faint mental image, etc.). As in perception, a biased introspective criterion may affect an introspector’s response without necessarily implying changes in the introspectability of a state. Even when rare, introspection errors (misses and false alarms) should be possible and in fact expected. By modeling introspection as any other detection mechanism, iSDT aims to explain the range of reliability of introspection while preserving our intuitions that it provides highly accurate judgments in familiar cases such as detecting severe pain while also accounting for introspection’s potential fallibility.

Thus, in the context of scientific experiments of consciousness, iSDT provides a model that predicts higher introspective reliability when judging strong experiences. When the experiences are weak—as one may expect from stimuli presented quickly and at threshold—introspective judgments are expected to be more unreliable (Dijkstra & Fleming 2023).

Another way to link scientific introspection to consciousness is to connect it to models of attention. Philosophical conceptions of introspective attention construe it as capable of directly focusing on phenomenal properties and experiences. As this idea is fleshed out, however, it is clearly not a form of attention studied by cognitive science, for the posited direct introspective attention is neither perceptual attention nor what psychologists call internal attention (e.g., the retrieval of thought contents as in memory recollection; Chun, Golomb, & Turk-Browne 2011). Calibrating introspection as it is used in the science of consciousness would benefit from concrete models of introspection (Chirimuuta 2014; Spener 2015; Wu 2023; Kammerer & Frankish 2023; Morales forthcoming).

We can draw inspiration from a proposal inspired by Gareth Evans (1982): in introspecting perceptual states, say judging that one sees an object, one draws on the same perceptual capacities used to answer the question whether the object is present. In introspection, one then appends a further concept of “seeing” to one’s perceptual report.^[3] Thus, instead of simply reporting that a red stimulus is present, one reports that one sees the red stimulus. Paradoxically, introspection relies on externally directed perceptual attention, but as noted earlier, identifying what one perceives is a way of characterizing what one’s perception is like, so this “outward” perspective can provide information about the inner.

Further, the advantage of this proposal is that questions of reliability come down to questions of the reliability of psychological capacities that can be empirically assessed, say perceptual, attentional and conceptual reliability. For example, Peters and Lau (2015) showed that accuracy in judgments about the visibility of a stimulus, the introspective measure, coincided with accuracy in judgments about stimulus properties, the objective measure (see also Rausch & Zehetleitner 2016). Further, in many of the examples to be discussed, the perceptual attention-based account provides a plausible cognitive model of introspection (Wu, 2023). Subjects report on what they perceptually experience by attending to the object of their experience, and where perception and attention are reliable, a plausible hypothesis is that their introspective judgments will be reliable as well. Accordingly, the reliability of introspection in the empirical studies to be discussed can be assumed. Still, given that no scientist should assert the reliability of a method without calibration, introspection must be subject to the same standards. There is more work to be done.

2.2 Access Consciousness and No-Report Paradigms

Introspection illustrates a type of cognitive access, for a state that is introspected is access conscious. This raises a question that has epistemic implications: is access consciousness necessary for phenomenal consciousness? If it is not, then there can be phenomenal states that are not access conscious, so are in principle not reportable. That is, phenomenal consciousness can *overflow* access consciousness (Block 2007).

Access is tied to attention, and attention is tied to report. Some views hold that attention is necessary for access, which entails phenomenal consciousness (e.g., the Global Workspace theory [[section 3.1](#)]).^[4] In contrast, other theories (e.g., recurrent processing theory [section 3.2](#)), hold that there can be phenomenal states that are not accessible.

Many scientists of consciousness take there to be evidence for no phenomenal consciousness without access and little if no evidence of phenomenal consciousness outside of access. While those antagonistic to overflow have argued that it is not empirically testable (Cohen & Dennett 2011) testing that attention is necessary for consciousness may be equally untestable. After all, we must eliminate attention to a target while gathering evidence for the absence of consciousness. Yet if gathering evidence for consciousness requires attention, then in fulfilling the conditions for testing the necessity of attention, we undercut the access needed to substantiate the absence of (Wu 2017; for a monograph length discussion of attention and consciousness, see Montemayor & Haladjian 2015).^[5] How then can we gather the required evidence to assess competing theories?

One response is to draw on *no-report paradigms* which measure reflexive behaviors correlated with conscious states to provide a window on the phenomenal that is independent of access (Lumer & Rees 1999; Tse et al. 2005). In binocular rivalry paradigms (see [section 5.2](#)), researchers show subjects different images to each eye. Due to their very different features, these images cannot be fused into a single percept. This results in alternating experiences that transition back and forth between one image and the other. Since the presented images remain constant while experience changes, binocular rivalry has been used to find the neural correlates of consciousness (see Zou, He, & Zhang 2016; and Giles, Lau, &

Odegaard 2016 for concerns about using binocular rivalry for studying the neural correlates of consciousness). However, binocular rivalry tasks have typically involved asking subjects to report what their experience is like, requiring the recruitment of attention and explicit report. To overcome this issue, Hesse & Tsao (2020) recently introduced an important variation by adding a small fixation point on different locations of each image. For example, subjects' right eye was shown a photo of a person with a fixation point drawn on the bottom of the image, and their left eye was presented with a photo of some tacos with a fixation point on the top. People and monkeys were trained to look at the fixation point while ignoring the image. This way, the experimenters could use their eye movement behavior as a proxy for which image they were experiencing without having to collect explicit reports: if subjects look up, they are experiencing the tacos; if they look down, they are experiencing the person. They confirmed with explicit reports from humans that both monkeys and humans in the no-report condition behaved similarly. Importantly, they found from single-cell recordings in the monkeys that neurons in inferotemporal cortex—a downstream region associated with high-level visual processing—represented the experienced image. Because these monkeys were never trained to report their percept (they were just following the fixation point wherever it was), the experimenters could more confidently conclude that the activity linked to the alternating images was due to consciously experiencing them and not due to introspecting or reporting.

No-report paradigms use indirect responses to track the subject's perceptual experience in the absence of explicit (conceptualized) report. These can include eye movements, pupil changes, electroencephalographic (EEG) activity or galvanic skin conductance changes, among others. No-report paradigms seem to provide a way to track phenomenal consciousness even when access is eliminated. This would broaden the evidential basis for consciousness beyond introspection and indeed, beyond intentional behavior (our "broad" conception of access). However, in practice they do not always result in drastically different results (Michel & Morales 2020). Moreover, they do not fully circumvent introspection (Overgaard & Fazekas 2016). For example, the usefulness of any indirect physiological measure depends on validating its correlation with alternating experience *given subjective reports*. Once it is validated, monitoring these autonomic responses can provide a way to substitute for subjective reports within that paradigm. One cannot, however, simply extend the use of no-report paradigms outside the behavioral contexts within which the method is validated. With each new experimental context, we must revalidate the measure with introspective report. Moreover, no-report paradigms do not match post-perceptual processing between conscious and unconscious conditions (Block 2019). Even if overt report is matched, the cognitive consequences of perceiving a stimulus consciously and failing to do so are not the same. For example, systematic reflections may be triggered by one stimulus (the person) but not the other (the tacos). These post-cognitive differences would generate different neural activity that is not necessarily related to consciousness. To avoid this issue, post-perceptual cognition would also need to be matched to rule out potential confounds (Block 2019). However, this is easier said than done and no uncontroversial solutions around this issue have been found (Phillips & Morales 2020; Panagiotaropoulos, Dwarakanath, & Kapoor 2020).

Can we use no report paradigms to address whether access is necessary for phenomenal consciousness? A likely experiment would be one that validates no-report correlates for some conscious phenomenon *P* in a concrete experimental context *C*. With this validation in hand, one then eliminates accessibility and attention with respect to *P* in *C*. If the no-report correlate remains, would this clearly support overflow? Perhaps, it could still be argued that the result does not rule out the possibility that phenomenal consciousness disappears with access consciousness *despite* the no-report correlate remaining. For example, the reflexive response and phenomenal consciousness might have a common cause that remains even if phenomenal consciousness is selectively eliminated by removing access.

2.3 Confidence and Metacognitive Approaches

Given worries about calibrating introspection, researchers have asked subjects to provide a different metacognitive assessment of conscious states via reports about *confidence* (Fleming 2023a, 2020; Pouget, Drugowitsch, & Kepecs 2016). A standard approach is to have subjects perform a task, say perceptual discrimination of a stimulus, and then report how *confident* they are that their perceptual judgment was accurate. This metacognitive judgment, a confidence rating, about perception can be assessed for accuracy by comparing it with perceptual performance (for discussion of formal methods such as metacognitive signal detection theory, see Maniscalco & Lau 2012, 2014). Related paradigms include post-decision wagering where subjects place wagers on specific responses as a way of estimating their confidence (Persaud, McLeod, & Cowey 2007; but see Dienes & Seth 2010).

There are some advantages of using confidence judgments for studying consciousness (Michel 2023; Morales & Lau 2022; Peters 2022). While standard introspective judgments about conscious experiences may capture more directly the phenomenon of interest, confidence judgments are easier to explain to subjects and they are also more interpretable from the experimenter's point of view. Confidence judgments provide an objective measure of *metacognitive sensitivity*: how well subjects' confidence judgments track their performance in the task. Subjects can also receive feedback on those ratings and, at least in principle, improve their metacognitive sensitivity (though this has proved hard to achieve in laboratory tasks; Rouy et al. 2022; Haddara & Rahnev 2022). Metacognitive judgments also have the advantage over direct judgments about conscious experiences (e.g., Ramsøy & Overgaard 2004) in that they allow for comparisons between different domains that have very different phenomenology (e.g., perception vs memory) (e.g., Gardelle, Corre, & Mamassian 2016; Faivre et al. 2017; Morales, Lau, & Fleming 2018; Mazancieux et al. 2020).

One concern with metacognitive approaches is that they also rely on introspection (Rosenthal 2019; see also Sandberg et al. 2010; Dienes & Seth 2010). If metacognition relies on introspection, does it not accrue all the disadvantages of the latter? Perhaps, but an important gain in metacognitive approaches is that it allows for quantitative psychophysical analysis. While it does not replace introspection, it brings an analytical rigor to addressing certain aspects of conscious awareness.

How might we bridge metacognition with consciousness as probed by traditional introspection? The metacognitive judgment reflects introspective assessment of the quality of perceptual states and can provide information about the presence of consciousness. For example, Peters and Lau (2015) found that increases in metacognitive confidence tracked increases in perceptual sensitivity, which presumably underly the quality of perceptual experiences. They also did not find significant differences between asking for confidence or visibility judgments. Even studies that have found (small) differences between the two kinds of judgments indicate considerable association between the two ratings, suggesting similar behavioral patterns (Rausch & Zehetleitner 2016; Zehetleitner & Rausch 2013).

Beyond behavior, neuroscientific work shows that similar regions of the brain (e.g., prefrontal cortex) may be involved in supporting both conscious awareness and metacognitive judgments. Studies with non-human primates and rodents have begun to shed light on neural processing for metacognition (for reviews, see Grimaldi, Lau, & Basso 2015; Pouget, Drugowitsch, & Kepecs 2016). From animal studies, one theory is that metacognitive information regarding perception is already present in perceptual areas that guide observational judgments, and these studies implicate parietal cortex (Kiani & Shadlen 2009; Fetsch et al. 2014) and the superior colliculus (Kim & Basso 2008; but see Odegaard et al. 2018). Alternatively, information about confidence might be read out by other structures (see [section 3.3](#) on Higher-Order Theory; also the entry on [higher order theories of consciousness](#)). In both human and animal studies, the prefrontal cortex (specifically, subregions in dorsolateral and orbitofrontal prefrontal cortex) has been found to support subjective reports of awareness (Cul et al. 2009; Lau & Passingham 2006), subjective appearances (Liu et al. 2019); visibility ratings (Rounis et al. 2010); confidence ratings (Fleming, Huijgen, & Dolan 2012; Shekhar & Rahnev 2018) and conscious experiences without reports (Mendoza-Halliday & Martinez-Trujillo 2017; Kapoor et al. 2022; Michel & Morales 2020).

2.4 The Intentional Action Inference

Metacognitive and introspective judgments result from intentional action, so why not look at intentional action, broadly construed, for evidence of consciousness? Often, when subjects perform perception guided actions, we infer that they are relevantly conscious. It would be odd if a person cooks dinner and then denies having seen any of the ingredients. That they did something intentionally provides evidence that they were consciously aware of what they acted on. An emphasis on intentional action embraces a broader evidential basis for consciousness. Consider the *Intentional Action Inference* to phenomenal consciousness:

If some subject acts intentionally, where her action is guided by a perceptual state, then the perceptual state is phenomenally conscious.

An epistemic version takes the action to provide good evidence that the state is conscious. Notice that introspection is typically an intentional action so it is covered by the inference. In this way, the Inference effectively levels the evidential playing field: introspective reports are simply one form among many types of intentional actions that provide evidence for consciousness. Those reports are not privileged.

The intentional action inference and no-report paradigms highlight that the science of consciousness has largely restricted its behavioral data to one type of intentional action, introspection. What is the basis of privileging one intentional action over others? Consider the calibration issue. For many types of intentional action deployed in experiments, scientists can calibrate performance by objective measures such as accuracy. This has not been formally done for introspection of consciousness, so scientists have privileged an uncalibrated measure over a calibrated one. This seems empirically ill-advised. On the flip side, one worry about the intentional action inference is that it ignores guidance by unconscious perceptual states (see sections [4](#) and [5.3.1](#)).

2.5 Unresponsive Wakefulness Syndrome and the Intentional Action Inference

The Intentional Action Inference is operative when subjective reports are not available. For example, it is deployed in arguing that some patients diagnosed with unresponsive wakefulness syndrome are conscious (Shea & Bayne 2010; Drayson 2014).

A patient [with unresponsive wakefulness syndrome] appears at times to be wakeful, with cycles of eye closure and eye opening resembling those of sleep and waking. However, close observation reveals no sign of awareness or of a ‘functioning mind’: specifically, there is no evidence that the patient can perceive the environment or his/her own body, communicate with others, or form intentions. As a rule, the patient can breathe spontaneously and has a stable circulation. The state may be a transient stage in the recovery from coma or it may persist until death. (Working Party RCP 2003: 249)

These patients are not clinically comatose but fall short of being in a “minimally conscious state”. Unlike unresponsive wakeful patients, minimally conscious state patients seemingly perform intentional actions.

Recent work suggests that some patients diagnosed with unresponsive wakeful syndrome are conscious. Owen et al. (2006) used fMRI to demonstrate correlated activity in such patients in response to commands to deploy imagination. In an early study, a young female patient was scanned by fMRI while presented with three auditory commands: “imagine playing tennis”, “imagine visiting the rooms in your home”, “now just relax”. The commands were presented at the beginning of a thirty-second period, alternating between imagination and relax commands. The patient demonstrated similar activity when matched to control subjects performing the same task: sustained activation of the supplementary motor area (SMA) was observed during the motor imagery task while sustained activation of the parahippocampal gyrus including the parahippocampal place area (PPA) was observed during the spatial imagery task. Later work reproduced this result in other patients and in one patient, the tasks were used as a proxy for “yes”/ “no” responses to questions (Monti et al. 2010; for a review, see Fernández-Espejo & Owen 2013). Note that these tasks probe *specific* contents of consciousness by monitoring neural correlates of conscious imagery.

Several authors (Greenberg 2007; Nachev & Husain 2007) have countered that the observed activity was an automatic, non-intentional response to the command sentences, specifically

to the words “tennis” and “house”. In normal subjects, reading action words is known to activate sensorimotor areas (Pulvermüller 2005). Owen and colleagues (2007), responded that the sustained activity over thirty-seconds made an automatic response less likely than an intentional response. One way to rule out automaticity is to provide the patient with different sentences such as “do *not* imagine playing tennis” or “*Sharlene* was playing tennis”. Owen et al. (2007) demonstrated that presenting “Sharlene was playing tennis” to a normal subject did not induce the same activity as when the subject obeyed the command “imagine playing tennis”, but the same intervention was not tried on patients. However, subsequent experiments using other measures such as EEG (Goldfine et al. 2011; Curley et al. 2018; Cruse et al. 2012) and functional connectivity (Demertzi et al. 2015), indicate conscious awareness is indeed present in patients with (wrongly diagnosed) unresponsive wakefulness syndrome (for a review, see Edlow et al. 2021; for ethical reflections around diagnosing awareness in patients with disorders of consciousness, see Young et al. 2021).

Owen et al. draw on a neural correlate of imagination, a mental action. Arguing that the neural correlate provides evidence of the patient’s executing an intentional action, they invoke a version of the Intentional Action Inference to argue that performance provides evidence for specific consciousness tied to the information carried in the brain areas activated.^[6]

3. Neurobiological Theories of Consciousness

Recall that the Generic Consciousness question asks:

What conditions/states N of nervous systems are necessary and/or sufficient for a mental state, M , to be conscious as opposed to not?

Victor Lamme notes:

Deciding whether there is phenomenality in a mental representation implies putting a boundary—drawing a line—between different types of representations... We have to start from the intuition that consciousness (in the phenomenal sense) exists, and is a mental function in its own right. That intuition immediately implies that there is also *unconscious* information processing. (Lamme 2010: 208)

It is uncontroversial that there is unconscious information processing, say processing occurring in a computer. What Lamme means is that there are conscious and unconscious mental states (representations). For example, there might be visual states of seeing X that are conscious or not ([section 4](#)).

In what follows, the theories discussed provide higher level neural properties that are necessary and/or sufficient for generic consciousness of a given state. To provide a gloss on the hypotheses: For the Global Neuronal Workspace, entry into the neural workspace is necessary and sufficient for a state or content to be consciousness. For Recurrent Processing Theory, a type of recurrent processing in sensory areas is necessary and sufficient for perceptual consciousness, so entry into the Workspace is not necessary. For Higher-Order Theories, the presence of a higher-order state tied to prefrontal areas is necessary and

sufficient for phenomenal experience, so recurrent processing in sensory areas is not necessary nor is entry into the workspace. For Information Integration Theories, a type of integration of information is necessary and sufficient for a state to be conscious.

3.1 The Global Neuronal Workspace

One explanation of generic consciousness invokes the *global neuronal workspace*. Bernard Baars first proposed the global workspace theory as a cognitive/computational model (Baars 1988), but we will focus on the neural version of Stanislas Dehaene and colleagues: a state is conscious when and only when it (or its content) is present in the global neuronal workspace making the state (content) globally *accessible* to multiple systems including long-term memory, motor, evaluational, attentional and perceptual systems (Dehaene, Kerszberg, & Changeux 1998; Dehaene & Naccache 2001; Dehaene et al. 2006). Notice that the previous characterization does not commit to whether it is phenomenal or access consciousness that is being defined.

Access should be understood as a relational notion:

A system X accesses content from system Y iff X uses that content in its computations/processing.

The *accessibility* of information is then defined as its potential access by other systems. Dehaene (Dehaene et al. 2006) introduces a threefold distinction: (1) neural states that carry information that is not accessible (*subliminal* information); (2) states that carry information that is accessible but not accessed (not in the workspace; *preconscious* information); and (3) states whose information is accessed by the workspace (*conscious* information) and is globally accessible to other systems. So, a necessary and sufficient condition for a state's being conscious rather than not is the access of a state or content by the workspace, making that state or content accessible to other systems. Hence, only states in (3) are conscious.

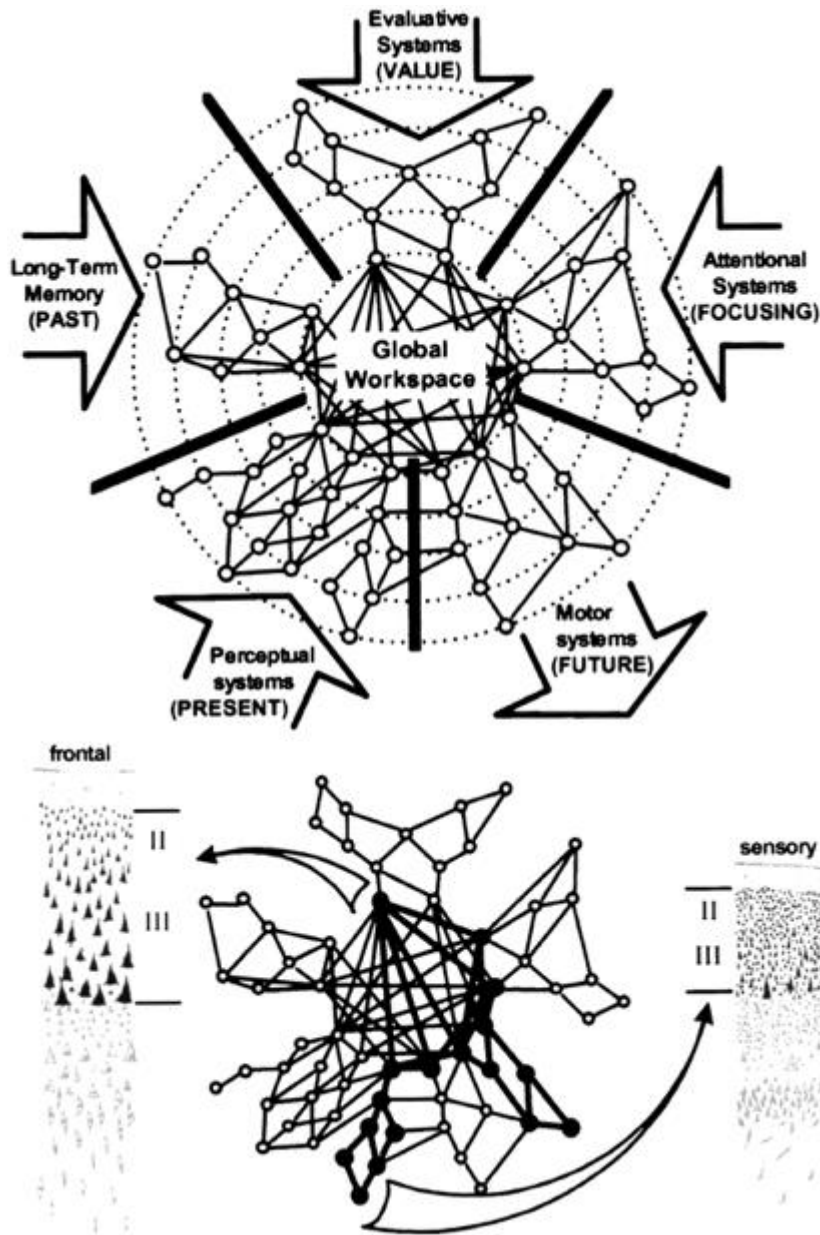


FIGURE 2. THE GLOBAL NEURONAL WORKSPACE

Figure Legend: The top figure provides a neural architecture for the workspace, indicating the systems that can be involved. The lower figure sets the architecture within the six layers of the cortex spanning frontal and sensory areas, with emphasis on neurons in layers 2 and 3. Figure reproduced from Dehaene, Kerszberg, and Changeux 1998. Copyright (1998) National Academy of Sciences.

The global neuronal workspace theory ties access to brain architecture. It postulates a cortical structure that involves workspace neurons with long-range connections linking systems: perceptual, mnemonic, attentional, evaluational and motoric.

What is the global workspace in neural terms? Long-range workspace neurons within different systems can constitute the workspace, but they should not necessarily be identified with *the* workspace. A subset of workspace neurons becomes the workspace when they

exemplify certain neural properties. What determines which workspace neurons constitute the workspace at a given time is the activity of those neurons given the subject's current state. The workspace then is not a rigid neural structure but a rapidly changing neural network, typically only a proper subset of all workspace neurons.

Consider then a neural population that carries content p and is constituted by workspace neurons. In virtue of being workspace neurons, the content p is accessible to other systems, but it does not yet follow that the neurons then constitute the global workspace. A further requirement is that workspace neurons are (1) put into an active state that must be sustained so that (2) the activation generates a *recurrent* activity between workspace systems. Only when these systems are recurrently activated are they, along with the units that access the information they carry, constituents of the workspace. This activity accounts for the idea of global *broadcast* in that workspace contents are accessible to further systems. Broadcasting explains the idea of consciousness as *for the subject*: globally broadcasted content is accessible for the subject's use in informing behavior.

The global neuronal workspace theory provides an account of *access* consciousness but what of phenomenal consciousness? The theory predicts widespread activation of a cortical workspace network as correlated with phenomenal conscious experience, and proponents often appeal to imaging results that reveal widespread activation when consciousness is reported (Dehaene & Changeux 2011). There is, however, a potential confound. We track phenomenal consciousness by access in introspective report, so widespread activity during reports of conscious experience correlates with both access and phenomenal consciousness. Correlation cannot tell us whether the observed activity is the basis of phenomenal consciousness or of access consciousness in report (Block 2007). This remains a live question for as discussed in [section 2.2](#), we do not have empirical evidence that overflow is false.

To eliminate the confound, experimenters ensure that performance does not differ between conditions where consciousness is present and where it is not. Where this was controlled, widespread activation was not clearly observed (Lau & Passingham 2006). Still, the absence of observed activity by an imaging technique does not imply the absence of actual activity for the activity might be beyond the limits of detection of that technique. Further, there is a general concern about the significance of null results given that neuroscience studies focused on prefrontal cortex are typically underpowered (for discussion, see Odegaard, Knight, & Lau 2017).

3.2 Recurrent Processing Theory

A different explanation ties perceptual consciousness to processing independent of the workspace, with focus on *recurrent* activity in sensory areas. This approach emphasizes properties of first-order neural representation as explaining consciousness. Victor Lamme (2006, 2010) argues that recurrent processing is necessary and sufficient for consciousness. Recurrent processing occurs where sensory systems are highly interconnected and involve feedforward and feedback connections. For example, forward connections from primary visual area V1, the first cortical visual area, carry information to higher-level processing

areas, and the initial registration of visual information involves a forward sweep of processing. At the same time, there are many feedback connections linking visual areas (Felleman & Van Essen 1991), and later in processing, these connections are activated yielding dynamic activity within the visual system.

Lamme identifies four stages of normal visual processing:

- Stage 1: Superficial feedforward processing: visual signals are processed locally within the visual system.
- Stage 2: Deep feedforward processing: visual signals have travelled further forward in the processing hierarchy where they can influence action.
- Stage 3: Superficial recurrent processing: information has traveled back into earlier visual areas, leading to local, recurrent processing.
- Stage 4: Widespread recurrent processing: information activates widespread areas (and as such is consistent with global workspace access).

Lamme holds that recurrent processing in Stage 3 is necessary and sufficient for consciousness. Thus, what it is for a visual state to be conscious is for a certain recurrent processing state to hold of the relevant visual circuitry. This identifies the crucial difference between the global neuronal workspace and recurrent processing theory: the former holds that recurrent processing at Stage 4 is necessary for consciousness while the latter holds that recurrent processing at Stage 3 is sufficient. Thus, recurrent processing theory affirms phenomenal consciousness without access by the global neuronal workspace. In that sense, it is an overflow theory (see [section 2.2](#)).

Why think that Stage 3 processing is sufficient for consciousness? Given that Stage 3 processing is not accessible to introspective report, we lack introspective evidence for sufficiency. Lamme appeals to experiments with brief presentation of stimuli such as letters where subjects are said to report seeing more than they can identify in report (Lamme 2010). For example, in George Sperling's partial report paradigm (Sperling 1960), subjects are briefly presented with an array of 12 letters (e.g., in 300 ms presentations) but are typically able to report only three to four letters even as they claim to see more letters (but see Phillips 2011). It is not clear that this is strong motivation for recurrent processing, since the very fact that subjects can report seeing more letters shows that they have some access to them, just not access to letter identity.

Lamme also presents what he calls *neuroscience arguments*. This strategy compares two neural networks, one taken to be sufficient for consciousness, say the processing at Stage 4 as per Global Workspace theories, and one where sufficiency is in dispute, say recurrent activity in Stage 3. Lamme argues that certain features found in Stage 4 are also found in

Stage 3 and given this similarity, it is reasonable to hold that Stage 3 processing suffices for consciousness. For example, both stages exhibit recurrent processing. Global neuronal workspace theorists can allow that recurrent processing in stage 3 is correlated, even necessary, but deny that this activity is explanatory in the relevant sense of identifying sufficient conditions for consciousness.

It is worth reemphasizing the empirical challenge in testing whether access is necessary for phenomenal consciousness (sections [2.1–2.2](#)). The two theories return different answers, one requiring access, the other denying it. As we saw, the methodological challenge in testing for the presence of phenomenal consciousness independently of access remains a hurdle for both theories.

3.3 Higher-Order Theory

A long-standing approach to conscious states holds that one is in a conscious state if and only if one relevantly represents oneself as being in such a state. For example, one is in a conscious visual state of seeing a moving object if and only if one suitably represents oneself being in that visual state. This higher-order state, in representing the first-order state that represents the world, results in the first order state's being conscious as opposed to not. The intuitive rationale for such theories is that if one were in a visual state but in no way aware of that state, then the visual state would not be conscious. Thus, to be in a conscious state, one must be aware of it, i.e., represent it (Rosenthal 2002; see the entry on [higher order theories of consciousness](#)). On certain higher-order theories (Higher-Order Thought Theory, Rosenthal 2005; and Higher-Order Representation of a Representation (HOROR), Brown 2015), one can be in a conscious visual state even if there is no visual system activity, so long as one represents oneself as being in that state (for a debate, see Block 2011; Rosenthal 2011). Other family of theories postulates that experiences are jointly determined by first- and higher-order states [e.g., Higher-Order State Space (HOSS) (Fleming 2020); Perceptual Reality Monitoring (PRM) (Lau 2019)]. An intermediate perspective proposes that higher-order states track our mental attitudes towards first-order states along different dimensions that include familiarity, vividness, value, and so on (Self-Organizing Meta-Representational Account (SOMA)—Cleeremans 2011; Cleeremans et al. 2020). These differences apart, higher-order theories merge with empirical work by tying high-order representations with brain activity, typically in the prefrontal cortex, which is taken to be the neural substrate of the required higher-order representations.

The focus on the prefrontal cortex allows for empirical tests of the higher-order theory against other accounts (Lau & Rosenthal 2011; LeDoux & Brown 2017; Brown, Lau, & LeDoux 2019; Lau 2022). For example, on the higher-order theory, lesions to prefrontal cortex should affect consciousness (see Kozuch 2013, 2022, 2023), testing the necessity of prefrontal cortex for consciousness. Against higher-order theories, some reports claim that patients with prefrontal cortex surgically removed maintain preserved perceptual consciousness (Boly et al. 2017) and that intracranial electrical stimulation (iES) to the prefrontal cortex does not alter consciousness (Racah, Block, & Fox 2021). This would lend support to recurrent processing theories that hold that prefrontal cortical activity is not necessary for consciousness (and would be evidence against both GWT and higher-order

theories). It is not clear, however, that the interventions succeeded in removing *all* of prefrontal cortex, leaving perhaps sufficient frontal areas needed to sustain consciousness (Odegaard, Knight, & Lau 2017), or that simple, localized stimulation to prefrontal cortex would be the right kind of stimulation for altering awareness (see Naccache et al. 2021). Bilateral suppression of prefrontal activity using transcranial magnetic stimulation also seems to selectively impair visibility as evidenced by metacognitive report (Rounis et al. 2010). Furthermore, certain syndromes and experimental manipulations suggest consciousness in the absence of appropriate sensory processing as predicted by some higher-order accounts (Lau & Brown 2019), a claim that coheres with the theory's sufficiency claims.

Subjective reports of conscious versus unconscious trials activate frontal regions as shown with EEG (Cul et al. 2009) and fMRI (Lau & Passingham 2006). Liu and colleagues (Liu et al. 2019) leveraged the “double-drift illusion” to show that real and apparent motion shared patterns of neural activity only in lateral and medial frontal cortex, not visual cortex. The double-drift illusion is a dramatic mismatch between physical and apparent motion created by a patch of gratings moving vertically while the gratings cycle horizontally; this creates the illusion of the patch's path to be more than 45° away from vertical. The conscious experience of seeing a stimulus veer off diagonally, whether physically or illusorily, was similarly encoded only in the prefrontal cortex, suggesting that the conscious experience of stimulus's diagonal motion was represented there, not in visual cortex. In a carefully designed no-report paradigm, Hatamimajoumerd and colleagues (2022) found that conscious stimuli were decodable from prefrontal cortex well above chance. Intracranial electrophysiological recording, where electrodes are placed directly on the surface of the brain, reveals prefrontal activity related to visual consciousness even when subjects were not required to respond to the stimulus (Noy et al. 2015). Fazekas and Nemeth (2018) discuss studies using different neuroimaging techniques showing significant increases in activity in the prefrontal cortex during dreams, a natural case of phenomenal awareness without report. Convergent evidence about the role of the prefrontal cortex sustaining awareness comes from single-cell recordings in macaques. Using binocular rivalry (see [section 5.2](#)), Dwarakanath et al. (2023) and Kapoor et al. (2022) show that dynamic subjective changes in monkeys' conscious experiences are robustly represented in prefrontal cortex.

3.4 Information Integration Theory

Information Integration Theory of Consciousness (IIT) draws on the notion of *integrated information*, symbolized by Φ , as a way to explain generic consciousness—specifically the quantity of consciousness present in a system (Tononi 2004, 2008; Oizumi, Albantakis, & Tononi 2014; Albantakis et al. 2023). IIT also aims to explain specific consciousness (i.e., the quality or content of conscious experiences) by appealing to the conceptual causal structure of the integrated information complex (i.e., the set of units of the physical substrate that is maximally integrated).

Integrated information theory (IIT) “starts from phenomenology and makes use of thought experiments to claim that consciousness is integrated information” (Tononi 2008: 216). IIT's first step is to find “*phenomenological axioms*”, that is, immediately given essential

properties of every conceivable experience. Once properly understood, these phenomenological axioms are taken by IIT to be irrefutably true (Albantakis et al. 2023: 3). These axioms, obtained by drawing on introspection and reason, are that consciousness exists, and that it is intrinsic, specific, unitary, definite, and structured. These axioms lead to postulates of *physical* existence. In other words, physical implementations that respect the properties first discovered through introspection. Finally, IIT develops mathematical formalisms that aim to preserve all these features and that can in principle help calculate the quality and quantity of integrated information in a system, or Φ .^[7]

Integrated information is defined in terms of the effective information carried by the parts of the system in light of its causal profile. For example, consider a circuit (neuronal, electrical, or otherwise). We can focus on a part of the whole circuit, say two connected nodes, and compute the effective information that can be carried by this microcircuit. The system carries *integrated information* if the effective informational content of the whole is greater than the sum of the informational content of the parts. If there is no partitioning where the summed informational content of the parts equals the whole, then the system as a whole carries integrated information and it has a positive value for Φ . Intuitively, the interaction of the parts adds more to the system than the parts do alone.

IIT holds that an above-zero value for Φ implies that a system is conscious, with more consciousness going with greater values for Φ . For example, Tononi argues that the human cerebellum does not contribute to consciousness due to its highly modular anatomical organization. Thus, it is hypothesized that the cerebellum has a value of zero Φ despite there being four to five times the number of neurons in the cerebellum versus in the human cortex. On IIT, what matters is the presence of appropriate connections and not the number of neurons (the soundness of this argument about cerebellum has been contested; Aaronson 2014b; Merker, Williford, & Rudrauf 2022). For this same reason, IIT also makes counterintuitive predictions. From panpsychist conclusions, such as admitting that even a 2D grid of inactivated logic gates “doing absolutely nothing, may in fact have a large value of PHI” (Tononi 2014 in [Other Internet Resources](#)), to neuroscientifically implausible conclusions: “a brain where no neurons were activated, but were kept ready to respond in a differentiated manner to different perturbations, would be conscious (perhaps that nothing was going on)” (Tononi 2004). (In [Other Internet Resources](#), see Aaronson 2014a and 2014b for striking counterexamples; see Tononi 2014 for a response).

It is important to note that IIT is not in and of itself a neuroscientific theory of consciousness. Rather, IIT is probably best understood as a metaphysical theory about the essential features of consciousness. Accordingly, these features could be present not just in organisms with neural systems but in any physical system (organic or not) that integrates information with a Φ larger than 0 (Tononi et al. 2016). Evidence of its metaphysical status are the theory’s idealist corollaries (see entry for Idealism). According to Tononi, and in stark contrast to current neuroscientific assumptions, only intrinsic entities (i.e., conscious entities) “truly exist and truly cause, whereas my neurons or my atoms neither truly exist nor truly cause” (Tononi, et al. 2022: 2 [[Other Internet Resources](#)]). Relatedly, IIT has received criticisms about its soundness and its neuroscientific status based on both empirical and theoretical

arguments. One notable concern is the lack of clarity between IIT’s metaphysical claims and their potential relevance for a neuroscientific understanding of consciousness.^[8]

3.5 Frontal or Posterior?

In recent years, one way to frame the debate between theories of generic consciousness is whether the “front” or the “back” of the brain is crucial. Using this rough distinction allows us to draw the following contrasts: Recurrent processing theories focus on sensory areas (in vision, the “back” of the brain) such that where processing achieves a certain recurrent state, the relevant contents are conscious even if no higher-order thought is formed or no content enters the global workspace. Similarly, proponents of IIT have recently emphasized a “posterior hot zone” covering parietal and occipital areas as a neural correlate for consciousness, as they speculate that this zone may have the highest value for Φ (Boly et al. 2017; but see Lau 2023). For certain higher-order thought theories, having a higher-order state, supported by prefrontal cortex, without corresponding sensory states can suffice for conscious states. In this case, the front of the brain would be sufficient for consciousness. Finally, the global neuronal workspace, drawing on workspace neurons that are present across brain areas to form the workspace, might be taken to straddle the difference, depending on the type of conscious state involved. They require entry into the global workspace such that neither sensory activity nor a higher order thought on its own is sufficient, i.e., neither just the front nor the back of the brain.

The point of talking coarsely of brain anatomy in this way is to highlight the neural focus of each theory and thus, of targets of manipulation as we aim for explanatory neural correlates in terms of what is necessary and/or sufficient for generic consciousness. What is clear is that once theories make concrete predictions of brain areas involved in generic consciousness, neuroscience can test them.

4. Neuroscience of Generic Consciousness: Unconscious Vision as Case Study

Since generic consciousness is a matter of a state’s being conscious or not, we can examine work on specific types of mental state that shift between being conscious or not and isolate neural substrates. Work on unconscious vision provides an informative example. In the past decades, scientists have argued for unconscious seeing and investigated its brain basis especially in *neuropsychology*, the study of subjects with brain damage. Interestingly, if there is unconscious seeing, then the intentional action inference must be restricted in scope since some intentional behaviors might be guided by unconscious perception ([section 2.4](#)). That is, the existence of unconscious perception blocks a direct inference from perceptually guided intentional behavior to perceptual consciousness. The case study of unconscious vision promises to illuminate more specific studies of generic consciousness along with having repercussions for how we attribute conscious states.

4.1 Unconscious Vision and the Two Visual Streams

Since the groundbreaking work of Leslie Ungerleider & Mortimer Mishkin (1982), scientists divide primate cortical vision into two streams: dorsal and ventral (for further dissection, see Kravitz et al. 2011). The dorsal stream projects into the parietal lobe while the ventral stream projects into the temporal lobe (see Figure 1). Controversy surrounds the functions of the streams. Ungerleider and Mishkin originally argued that the streams were functionally divided in terms of *what* and *where*: the ventral stream for categorical perception and the dorsal stream for spatial perception. David Milner and Melvyn Goodale (1995) have argued that the dorsal stream is for *action* and the ventral stream for “*perception*”, namely for guiding thought, memory and complex action planning (see Goodale & Milner 2004 for an engaging overview). There continues to be debate surrounding the Milner and Goodale account (Schenk & McIntosh 2010) but it has strongly influenced philosophers of mind.

Substantial motivation for Milner and Goodale’s division draws on lesion studies in humans. Lesions to the dorsal stream do not seem to affect conscious vision in that subjects are able to provide accurate reports of what they see (but see Wu 2014a). Rather, dorsal lesions can affect visual-guidance of action with *optic ataxia* being a common result. Optic ataxic subjects perform inaccurate motor actions. For example, they grope for objects, yet they can accurately report the object’s features (for reviews, see Andersen et al. 2014; Pisella et al. 2009; Rossetti, Pisella, & Vighetto 2003). Lesions in the ventral stream disrupt normal conscious vision, yielding visual agnosia, an inability to see visual form or to visually categorize objects (Farah 2004).

Dorsal stream processing is said to be unconscious. If the dorsal stream is critical in the visual guidance of many motor actions such as reaching and grasping, then those actions would be guided by unconscious visual states. The visual agnostic patient DF provides critical support for this claim.^[9] Due to carbon monoxide poisoning, DF suffered focal lesions largely in the ventral stream spanning the *lateral occipital complex* that is associated with processing of visual form (high resolution imaging also reveals small lesions in the parietal lobe; James et al. 2003). Like other visual agnosics with similar lesions, DF is at chance in reporting aspects of form, say the orientation of a line or the shape of objects. Nevertheless, she retains color and texture vision. Strikingly, DF can generate accurate visually guided action, say the manipulation of objects along specific parameters: putting an object through a slot or reaching for and grasping round stones in a way sensitive to their center of mass. Simultaneously, DF denies seeing the relevant features and, if asked to verbally report them, she is at chance. In this dissociation, DF’s verbal reports give evidence that she does not visually experience the features to which her motor actions remain sensitive.

What is uncontroversial is that there is a division in explanatory neural correlates of visually guided behavior with the dorsal stream weighted towards the visual guidance of motor movements and the ventral stream weighted towards the visual guidance of conceptual behavior such as report and reasoning (see section 5.3 on manipulation of seeing words via ventral stream stimulation). A substantial further inference is that consciousness is segregated away from the dorsal stream to the ventral stream. How strong is this inference?

Recall the intentional action inference. In performing the slot task, DF is doing something intentionally and in a visually guided way. For control subjects performing the task, we

conclude that this visually guided behavior is guided by conscious vision. Indeed, a folk-psychological assumption might be that consciousness informs mundane action (Clark 2001; for a different perspective see Wallhagen 2007). Since DF shows similar performance on the same task, why not conclude that she is also visually conscious? Presumably, one hesitates because DF's introspective reports clash with the intentional action inference. DF denies seeing features she is visually sensitive to in action. Should introspection then trump intentional action in attributing consciousness?

Two issues are worth considering. The first is that introspective reports involve a specific type of intentional action guided by the experience at issue. One type of intentional behavior is being prioritized over another in adjudicating whether a subject is conscious. What is the empirical justification for this prioritization? The second issue is that DF is possibly unique among visual agnosics. It is a substantial inference to move from DF to a general claim about the dorsal stream being unconscious in neurotypical individuals (see Mole 2009 for arguments that consciousness does not divide between the streams; see Wu 2013 for an argument for unconscious visually guided action in normal subjects). What this shows is that the methodological decisions that we make regarding how we track consciousness are substantial in theorizing about the neural bases of conscious and unconscious vision.

4.2 Blindsight

Two issues are worth considering. The first is that introspective reports involve a specific type of intentional action guided by the experience at issue. One type of intentional behavior is being prioritized over another in adjudicating whether a subject is conscious. What is the empirical justification for this prioritization? The second issue is that DF is possibly unique among visual agnosics. It is a substantial inference to move from DF to a general claim about the dorsal stream being unconscious in neurotypical individuals (see Mole 2009 for arguments that consciousness does not divide between the streams; see Wu 2013 for an argument for unconscious visually guided action in normal subjects). What this shows is that the methodological decisions that we make regarding how we track consciousness are substantial in theorizing about the neural bases of conscious and unconscious vision. ^[10]

The neuroanatomical basis of blindsight capacities remains unclear. Certainly, the loss of V1 deprives later cortical visual areas of a normal source of visual information. Still, there are other ways that information from the eye bypasses V1 to provide inputs to later visual areas. Alternative pathways include the superior colliculus (SC), the lateral geniculate nucleus (LGN) in the thalamus, and the pulvinar as likely sources.

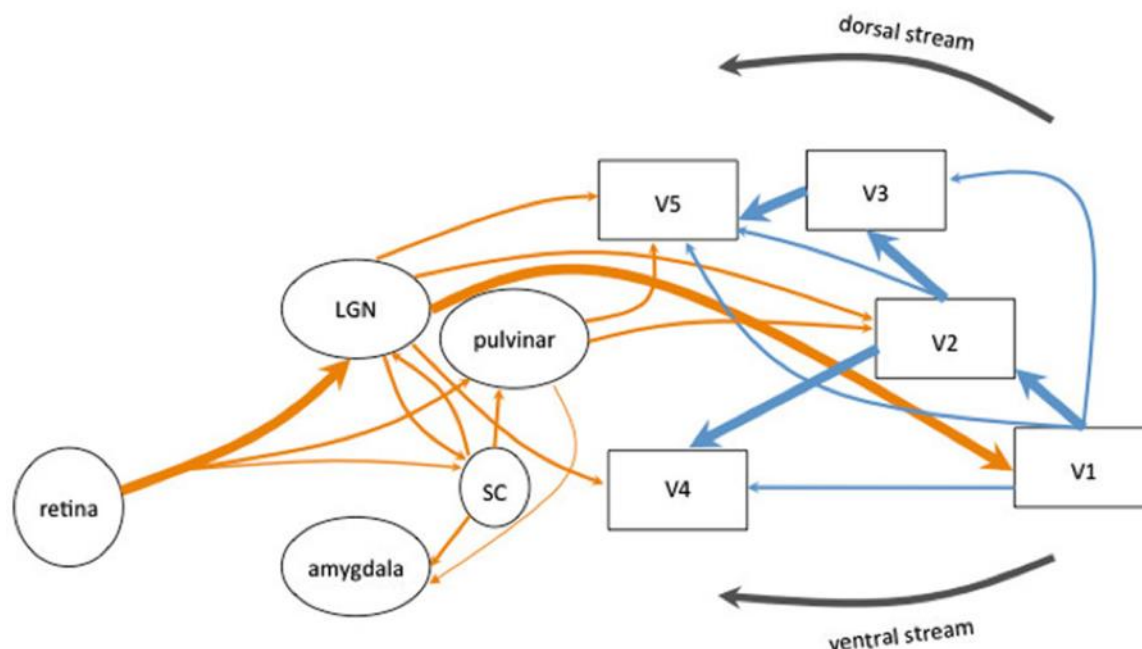


FIGURE 3. SUBCORTICAL PATHWAYS AND THEIR CONNECTION TO CORTICAL VISION (FROM URBANSKI, COUBARD, & BOURLON 2014)

Figure Legend: The front of the head is to the left, the back of the head is to the right. One should imagine that the blue-linked regions are above the orange-linked regions, cortex above subcortex. V4 is assigned to the base of the ventral stream; V5, called area MT in nonhuman primates, is assigned to the base of the dorsal stream.

The latter two have direct *extrastriate* projections (projections to visual areas in the occipital lobe outside of V1) while the superior colliculus synapses onto neurons in the LGN and pulvinar which then connect to extrastriate areas (Figure 3). Which of these provide for the basis for blindsight remains an open question though all pathways might play some role (Cowey 2010; Leopold 2012). If blindsight involves nonphenomenal, unconscious vision, then these pathways would be a substrate for it, and a functioning V1 might be necessary for normal conscious vision.

Campion et al. (1983) raised an important alternative explanation: blindsight subjects in fact have severely degraded conscious vision but merely report on them with low confidence (see Phillips 2021 for a recent version of this critique; see Michel & Lau 2021 for a response). In their reports, blindsight subjects feel like they are guessing about stimuli they can objectively discriminate. Campion et al. drew on *signal detection theory*, which emphasizes two determinants of detection behavior: *perceptual sensitivity* and *response criterion* (see [section 2.1](#)). Campion et al. hypothesized that blindsight patients are conscious in that they are aware of visual signal where discriminability is low. Further, blindsight patients are more conservative in their response so will be apt to report the absence of a signal by saying that they do not see the relevant stimulus even though the signal is there, and they can detect it, as verified by their above chance visually guided behavior.

This possibility was explicitly tested by Azzopardi and Cowey (1997) with the well-studied blindsight patient, GY. They compared blindsight performance with normal subjects at threshold vision using signal detection measures and found that with respect to motion

stimuli, the difference between discrimination and detection used to argue for blindsight can be explained by changes in response criterion, as Campion et al. hypothesized. That is, GYs claim that he does not see the stimulus is due to a conservative criterion and not to a detection incapacity. Interestingly, for static stimuli, his response criterion did not change but his sensitivity did, as if he was tapping into two different visual processing mechanisms in each task (for an alternative explanation based on shifting response criterion, see Ko & Lau 2012).

In introspecting, what concepts are available to subjects will determine their sensitivity in report. In many studies with blindsight, subjects are given a binary option: do you see the stimulus or do you not see it? The concern is that the *do not see* option would cover cases of degraded consciousness that subjects might be unwilling to classify as seeing due to a conservative response criterion. So, what if subjects are given more options for report? Ramsøy and Overgaard (2004; see also Overgaard et al. 2006) provided subjects with four categories for introspective report: *no experience*; *brief glimpse*; *almost clear experience*; *clear experience*. Using this *perceptual awareness scale*, they found that subjects' objective performance tracked their introspective reports where performance was at chance when subjects reported no visual experience. As visibility increased, so did performance. When the scale was used with a blindsight patient (Overgaard et al. 2008), no above chance performance was detected when the subject reported no visual experience (see also Mazzi, Bagattini, & Savazzi 2016 for further evidence). A live alternative hypothesis is that blindsight does not present a case of unconscious vision, but of degraded conscious vision with a conservative response bias that affects introspection. At the very least, the issue depends on how introspection is deployed, a topic that deserves further attention (see Phillips 2016, 2021 for further discussion of blindsight).

4.3 Unconscious Vision and the Intentional Action Inference

Blindsight and DF show that damage to specific regions of the brain disrupts normal visual processing, yet subjects can access visual information in preserved visual circuits to inform behavior despite failing to report on the relevant visual contents. The received view is that these subjects demonstrate unconscious vision. One implication is that the normal processing in the ventral stream, tied to normal V1 activity, plays a necessary role in normal conscious vision. Another is that dorsal stream processing or visual stream processing that bypasses V1 via subcortical processing yields only unconscious visual states. This points to a set of networks that begin to provide an answer to what makes visual states conscious or not. An important further step will be to integrate these results with the general theories noted earlier ([section 3](#)).

Still, the complexities of the empirical data bring us back to methodological issues about tracking consciousness and the following question: What behavioral data should form the basis of attributions of phenomenal consciousness? The intentional action inference is used in a variety of cases to attribute conscious states, yet the results of the previous sections

counsel us to be wary of applying that inference widely. After all, some intentional behavior might be unconsciously guided.

In the case of DF, we noted that unlike many other visual agnosics, she can direct motor actions towards stimuli that she cannot explicitly report and which she denies seeing. In her case, we prioritize introspective reports over intentional action as evidence for unconscious vision. Yet, one might take a broader view that vision for action is always conscious and that what DF vividly illustrates is that some visual contents (dorsal stream) are tied directly to performance of intentional motor behavior and are not directly available to conceptual capacities deployed in report. In contrast, other aspects of conscious vision, supported by the ventral stream, are directly available to guide reports. This functional divergence is explained by the anatomical division in cortical visual processing.

For some time now, these striking cases have been taken as clear cases of unconscious vision and if this hypothesis is correct, the work has begun to identify visual areas critical for creating seeing, sometimes conscious and sometimes not. The neuroanatomy demonstrates that visually-guided behavior has a complex neural basis involving cortical and subcortical structures that demonstrate a substantial level of specialization. Understanding consciousness and unconsciousness in vision will need to be sensitive to the complexities of the underlying neural substrate.

5. Specific Consciousness

We turn to experimental work on specific consciousness:

Specific Consciousness: What neural states or properties are necessary and/or sufficient for a conscious perceptual state to have content X rather than Y?

In this section, we examine attempts to address claims about necessity and sufficiency by manipulation of the contents of consciousness through direct modulation of neural representational content.

5.1 Neural Representationalism

In thinking about neural explanations of specific consciousness, namely the contents of consciousness, we will provisionally assume a type of first-order representationalism about phenomenal content, namely that such content supervenes on neural content (see the entry on [representational theories of consciousness](#)). One strong position would be that phenomenal content is *identical* to appropriate neural content. A weaker correlation claim affirms only supervenience: no change in phenomenal content without a change in neural content. This neural representationalism allows us to link phenomenal properties to the brain via linking neural contents to perceptual contents.

In invoking neural content, the assumption is that neural content mirrors perceptual content, so that if one is experiencing dots moving in a certain direction, there is a neural representation with the same content. This is a simplification and does not cohere with a common current approach to neural content that takes it to be *probabilistic* (for accessible

discussions, see Pouget, Dayan, & Zemel 2003; Colombo & Seriès 2012; Rescorla 2015). Yet perceptual content does not *seem* probabilistic. This emphasizes a *prima facie* disconnect between current theories of neural content and those of phenomenal content. One option is to find nonprobabilistic content at the neural level. The other is to find probabilistic content at the phenomenal level (for related ideas, see Morrison 2016 and a response by; Denison 2017; and Beck 2020; also Munton 2016; Block 2018; Shea 2018; Gross 2020; Vance 2021; Siegel 2022; Lee & Orlandi 2022). For simplicity, in what follows we assume a simple mapping between neural content and perceptual experiences.

5.2 The Contrast Strategy: Binocular Rivalry

A common approach, the *contrast strategy*, enjoins experimentalists to identify relevant correlates for some phenomenon *P* by contrasting cases where *P* is present from cases where *P* is not. Work on binocular rivalry illustrates this strategy (among many reviews, see Tong, Meng, & Blake 2006; Blake, Brascamp, & Heeger 2014; Blake 2022). When each eye receives a different image simultaneously, the subject does not see both, say one stimulus overlapping the other. Rather, visual experience alternates between them. Call this *phenomenal alternation*. An initial restatement of our question about specific consciousness in respect of binocular rivalry is:

Specific Rivalry: What neural property is necessary and/or sufficient for phenomenal alternation in binocular rivalry in condition *C*?

That is, empirical theories aim to explain how visual content alternates in binocular rivalry.^[11] Notice that this is a question about specific rather than generic consciousness, as the contrast is not between a state's being conscious versus not but about the contrast between two conscious states with different contents

Neural explanations of binocular rivalry concern competition at some level of visual processing: (a) “interocular” competition between *monocular* neurons early in the visual system, namely visual neurons that receive input from only one eye or (b) competition between *binocular* neurons later in the visual system, namely neurons that receive input from both eyes. The winner of competition fixes which stimulus the subject experiences at a given time. Some imaging studies in humans suggest that neural activity in V1 did correlate with alternation of the experienced images. For example, Polonsky et al. used fMRI to demonstrate that V1 activity to competing stimuli tracked perception (Polonsky et al. 2000; but see Maier et al. 2008). In contrast, some of the earliest electrophysiological studies (Leopold & Logothetis 1996; Logothetis, Leopold, & Sheinberg 1996; see also Hesse & Tsao 2020) on awake behaving monkeys supported later binocular processing as the neural basis of binocular rivalry. Processing in later (inferotemporal cortex, IT; see [Figure 1](#)) rather than earlier visual areas (V1 or V2) was observed to be best correlated to the monkey's reported perception based on the monkey's stimuli-specific response.

Recent accounts have taken binocular rivalry as resulting from processes at multiple levels (Wilson 2003; Freeman 2005; Tong, Meng, & Blake 2006). For example, when the two competing stimuli have parts that can be fused into a coherent stimulus, as when half of a picture is presented to each eye, the subject can perceive the fusion, integrating content from

each eye (Kovács et al. 1996; Ngo et al. 2000). This suggests that binocular rivalry can be sensitive to global properties of the stimulus (see Baker & Graf 2009). What unifies the mechanisms, perhaps, is the function of resolving a conflict generated by the stimuli.

Assume that some neural process R resolves interocular competition: when R resolves competition between stimuli X and Y in favor of X , then the subject is phenomenally conscious of X rather than Y and vice versa. Notice that R has the same “gating” function for any stimuli X and Y that are subject to binocular rivalry. So, while the presence of R can explain why the subject is having one conscious visual experience rather than another, R is not tied to a specific content. This suggests that in answering the question about rivalry, we will at best be identifying a necessary but not sufficient condition for a conscious visual state having a content X . R is a general gate for consciousness (cf. attention in global workspace theory).

An example of this gating of awareness can be found in recent studies using no-report paradigms, in which analysis of local field potentials in macaque prefrontal cortex has shown that transitions in frequency field activity track transitions in experience and encoding of the contents of conscious experiences (Dwarakanath et al. 2023). Moreover, feature-selective neurons in the prefrontal cortex are modulated by subjective changes in experiences (regardless if these were due to physical changes or rivalry). This modulation allows trial-by-trial successful decoding of the contents of experience from prefrontal cortex (Kapoor et al. 2022). Together, these recent studies offer robust evidence for a causal role of prefrontal cortex in supporting conscious experiences.

A narrower explanation of specific consciousness would identify the specific neural representations that explain a conscious state’s having the specific content X (rather than Y). By the representationalism assumption, this will involve identifying neural representations with the same content, X . Focusing on a gate in explaining alternation in rivalry stops short of identifying those representations. Still, binocular rivalry can provide a useful method for isolating neural populations that carry relevant content. In principle, for any stimulus type of interest, X (e.g., faces, words, etc.), so long as X is subject to binocular rivalry, we can use rivalry paradigms to isolate brain areas that carry the relevant information that correlate with the subject’s perceiving X . That would allow us to identify potential candidates for the neural basis of conscious content. (For a criticism of binocular rivalry paradigms as a means for discovering the neural basis of perceptual awareness—as opposed to mere perceptual processing—see Zou, He, & Zhang 2016; Giles, Lau, & Odegaard 2016).

5.3 Neural Stimulation

There are limited opportunities to manipulate human brain activity in a targeted way. Recent use of transcranial magnetic stimulation to activate or suppress neural activity has provided illumination, but such interventions are coarse-grained. Ultimately, to locate an explanatory correlate for specific conscious contents, we will need more fine-grained interventions in brain tissue. In humans, such opportunities are generally confined to manipulation before surgical interventions, say for brain tumors or epilepsy.

In the middle of the last century, neurosurgeon Wilder Penfield and colleagues performed a set of direct electrical microstimulations during preoperative procedures (Penfield & Perot 1963), and in certain cases induced hallucinations by stimulating primary sensory cortices such as V1 or S1 (see [Figure 1](#)). This provided evidence that endogenous activity could be causally sufficient for phenomenal experiences. Penfield's interventions, however, were not based on fine-grained targeting of specific neural representations. As Cohen and Newsome note:

Penfield's approach failed to generate substantial new insights into the neural basis of perception and cognition...because the gross electrical activation elicited by surface electrodes could not be related mechanistically to the information being processed within the excited neural tissue. (Cohen & Newsome 2004: 1)

A different approach begins with a more detailed understanding of underlying neural representations tied to different brain regions (Figure 4). For example, the fusiform face (FFA) area appears to be necessary for normal human face experience in that lesions in FFA lead to *prosopagnosia*, the inability to see faces even if one can see their parts [whether FFA is necessary for seeing faces specifically, or more generally for visual expertise is a point of contention (Kanwisher 2000; Tarr & Gauthier 2000)]. FFA is part of a larger network that is important in visual processing of faces (Behrmann & Plaut 2013). Recently, microstimulation of FFA in an awake human epilepsy patient induced visual distortions of actual faces as opposed to other objects (Parvizi et al. 2012). Alterations of visual experience were also reported during microstimulating the parahippocampal place (PPA) area in an awake pre-operative epileptic patient that induced visual hallucinations of scenes (Mégevand et al. 2014). PPA is the same area that showed activation in patients with unresponsive wakefulness syndrome when they putatively imagined walking around their home ([section 2.4](#)).

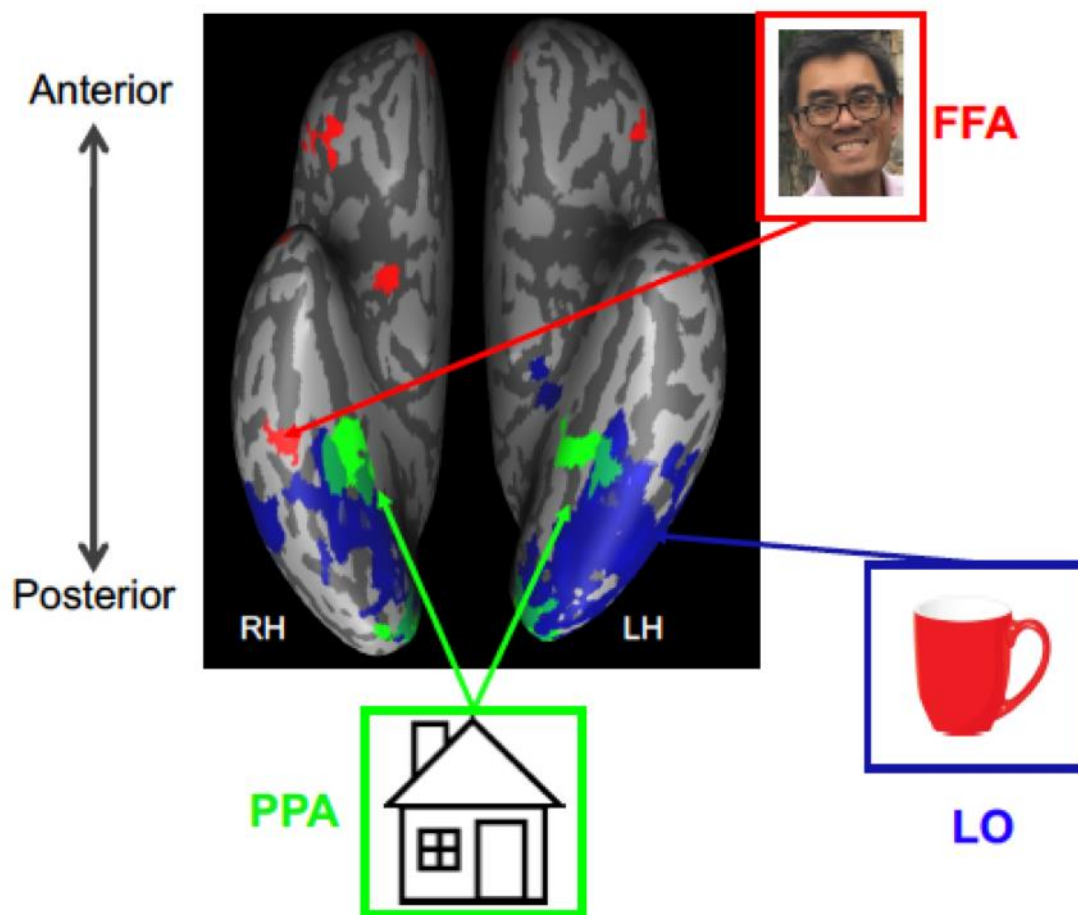


FIGURE 4. VENTRAL STREAM AREAS

A view from the bottom of cortex with location of areas FFA, PPA and LO identified. Occipital cortex is on the bottom. LO is lesioned in the visual agnostic patient, DF (see [section 4.1](#)). This figure is modified from figure 1 of Behrmann and Plaut 2013, kindly provided by Marlene Behrmann and used with her permission.

Another successful example of target microstimulation involves the visual word form area. This region in the left midfusiform gyrus (lmFG) is important for normal processing of visual word forms during reading. Microstimulation in human epileptic patients selectively disrupted word and letter reading—presumably altering the visual experience of letters—without changing general form perception (Hirshorn et al. 2016) (see Movies [S1](#) and [S2](#) in [Other Internet Resources](#)).

It is worth noting that many neuroscientists of vision take themselves to be investigating seeing in the ordinary sense, one that implies consciousness, but very few of them would characterize their work as about consciousness. That said, their work is of direct relevance to our understanding of specific consciousness even if it is not always characterized as such.^[12]

An important approach in visual neuroscience was articulated by A.J. Parker and William Newsome in “Sense and the Single Neuron” (1998) via “principles” to connect

electrophysiological data about information processing to perception (for a recent discussion, see Ruff & Cohen 2014). To probe the neural basis of perception, neuroscientists need to explanatorily link neural data to the subject's perception that guides behavior. The experimenter must ensure that recorded neural content correlates with perceptual content and not just response. Further, manipulation of the neurons carrying information should affect perception: inducing appropriate neural activity should shift perceptual response while abolishing or reducing that activity should eliminate or reduce perceptual response as measured in behavior. These proposals address concerns about necessity and sufficiency.

The intentional action inference is applicable (or at least its evidential version):

If some subject acts intentionally, where her action is guided by a perceptual state, then that state is phenomenally conscious.

We will consider the strength of this inference in two cases. The first case, visual motion perception, introduces the principles that guide the manipulation of neural content while the second case concerns tactile experience of vibration. These cases involve experiments with non-human primates, so we lack introspective reports. But direct manipulation of the human brain along with introspective reports have been performed too as discussed above with respect to face and letter perception.

These experiments involve microstimulation of small populations of neurons that are targeted precisely because of their informational content. Microstimulation involves injecting a small current from the tip of an electrode inserted into brain tissue that directly stimulates nearby neurons or, through synaptic connections to other neurons, indirectly activates more distant neurons (see Histed, Ni, & Maunsell 2013 for a review). It is assumed that neurons *tuned* in similar ways, that is neurons that respond to similar stimuli, tend to be interconnected, so microstimulation is taken to largely drive similarly tuned neurons.

5.3.1 Visual Motion Perception

We begin with visual motion perception in primates. Since the principles introduced here are central to much perceptual neuroscience and provide the basis for probing the link between neural representations and perceptual content, I examine it carefully. The salient question will be whether conscious experience is changed by the manipulations.

The work we shall discuss was done in awake behaving macaque monkeys. Visual area MT in the monkey brain (called V5 in humans) plays an important role in the visual experience of motion. MT is taken to lie in the dorsal visual stream ([Figure 1](#)). Lesions that disrupt MT are known to cause *akinetopsia*, the inability to see motion. One patient with an MT (V5) lesion reported the following phenomenology: “people were suddenly here or there but I have not seen them moving” (Zihl, von Cramon, & Mai 1983: 315). MT processing looks to be necessary for normal visual motion experience. Furthermore, MT neurons represent (carry information regarding) the direction of motion of visible stimuli: MT neurons are *tuned* for motion in specific directions with the highest firing rate for a specific direction of motion (for other functions and responses of MT, see Born & Bradley 2005). By placing motion stimuli in a neuron's receptive field, scientists can map its tuning:

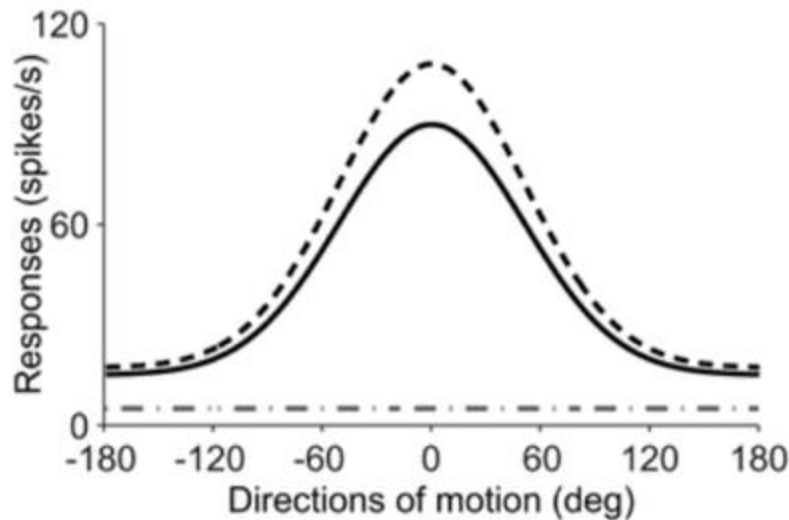


FIGURE 5. MT Neuron Tuning Curve

Figure Legend: Tuning of a neuron in MT showing a peak response in spiking rate at 0 degrees of motion. The dashed curve is generated when the animal is attending to the motion stimulus while it is in the receptive field (we shall not discuss the neural basis of attention, but see Wu 2014b: chap. 2 for a summary of the neuroscience of attention). The solid curve shows MT response when the animal is not attending to the motion stimulus in the receptive field. Figure from (Lee & Maunsell 2009).

What is plotted is the activity of an MT neuron, in spikes per second, to a specific type of motion stimulus placed within its receptive field. How to relate a tuning curve to a determinate content is complicated. Since the neural response is not simply to one stimulus value, it is not obvious that the neuron should be taken to represent 0 degrees of motion, namely the value at its peak response. Indeed, theorists have noted that the tuning curve looks like a probability density function, and many now take neurons to have *probabilistic* content ([section 5.1](#)).

Experimenters have trained macaque monkeys to perform discrimination tasks reporting direction of motion. Typically, the monkey maintains fixation while the moving stimulus is placed within the receptive field of the recorded neuron. The monkey reports the direction of the stimulus by moving its eyes to a target that stands for either leftward or rightward motion (other behavioral reports can be generated such as moving a joystick). Provisionally, we apply the intentional action inference, so we assume that such reports are guided by conscious visual experience of the stimuli. Thus, *changes in behavior will be evidence for changes in conscious content*.

Early work suggested that the activity of a single neuron provides a strong correlate of the animal's visually guided performance. This can be seen by plotting both the animal and the neuron's performance across different stimulus values. In these experiments, the value concerns the percent *coherence* of motion of a set of dots defined as the number of dots moving in the same direction (0% coherence being random motion; 100% being all dots moving in the same direction). In the first case, we construct a *psychometric curve* that plots the animal's percent correct reports relative to percent coherence of motion of the stimuli. As one might expect, percent correct reports drop as coherence drops, and the inflection point

reflects where the subject is equally likely to indicate left or right motion. We can do the same for the neural activity of the neuron across the same stimulus values, a *neurometric curve*.

The experimentalist's window onto conscious experience is through behavior, the assumption being that report about motion correlates with perceptual experience. Correlation is assessed by asking the following question: would an ideal observer, using the activity of the neuron in question, be able to *predict* the animal's visually guided performance? Essentially, do the psychometric and neurometric curves overlap? Strikingly, yes. MT neurons were observed to predict the animal's behavior (Britten et al. 1992).

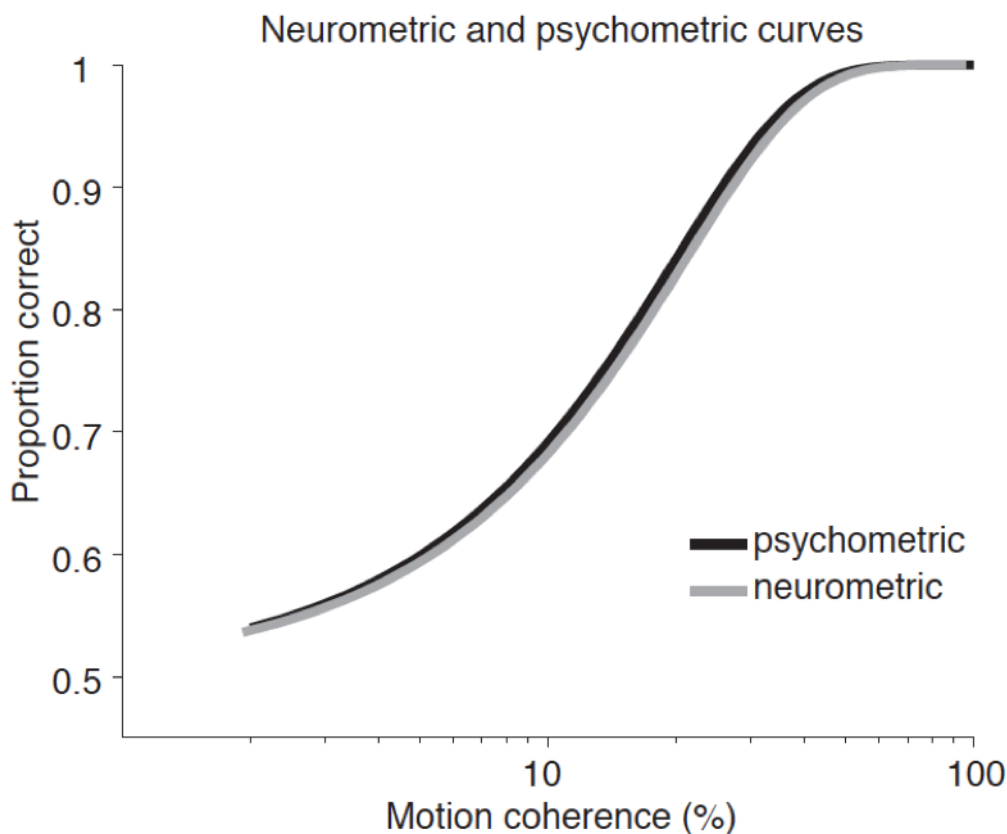


FIGURE 6. PSYCHOMETRIC AND NEUROMETRIC CURVES

Figure Legend: Psychometric and neurometric curves for a single MT neuron during performance of a motion direction detection task. Percent correct performance is plotted on the y-axis while percent motion coherence is plotted in a log scale on the x-axis. Figure modified from Ruff & Cohen 2014 and kindly provided by Doug Ruff.

This shows that the activity of a single MT neuron provides a neural correlate of the animal's visual discrimination of motion. Note that this is just a neural correlate of behavior. No one suggested that this neuron was causally sufficient for the behavior or for perception. Later results have suggested that individual neurons are not quite as sensitive as Britten suggested, but that small groups of MT neurons are sufficient to predict behavior (Cohen & Newsome 2009).

Earlier, we worried about mere correlates. To get causal or explanatory purchase, the content of the MT neurons correlated to the animal's behavior must be shown to contribute to perceptual guidance. This predicts that if we manipulate the content of the neurons, i.e., manipulate neural representations, then we should manipulate the content of the animal's visual experience of motion as reflected by predicted changes in behavior. This would be to test sufficiency with respect to specific consciousness.

Newsome and colleagues demonstrated that microstimulation of MT neurons shifted the animal's performance in predictable ways. Assume that neural population P , by encoding information about stimulus motion, can inform the subject's report of motion direction. This information is accessible for the control of behavior. Activation of P by microstimulation should shift behavior in a motion selective way correlated with the direction that P is tuned to (represents). This was first demonstrated by Salzman et al. (1990). They inserted electrodes into MT and identified neurons tuned to a particular orientation. During a motion discrimination task, microstimulation of neurons with that tuning led to a shift in the psychometric curve as if that neuron was given more weight in driving behavior.

In conditions of microstimulation relative to its absence, the monkey was more likely to report that there was motion in the stimulated neuron's preferred direction. In the original experiment, the psychophysical effect of microstimulation was equivalent to the addition of 7–20% coherence in the stimulus with respect to the neuron's preferred direction, depending on the experimental conditions. Further, as a test of necessity, a selective lesion of MT disrupted motion discrimination though the animals were able to recover some function suggesting that other visual information streams could be tapped so as to support performance (Newsome & Paré 1988).

Adopting the intentional action inference, one can conclude that the microstimulation shifted perceptual content (or again, that we have good evidence for this shift). That said, given our discussion of unconscious vision ([section 4](#)), another possibility is that MT microstimulation only changes unconscious visual representations. Newsome himself asked:

What is the conscious experience that accompanies the stimulation and the monkey's decision? Even if you knew everything about how the neurons encode and transmit information, you may not know what the monkey experiences when we stimulate his MT. (Singer 2006)

Clearly, having the monkey provide an introspective report would add evidential weight, but obtaining such reports from non-linguistic creatures is difficult. How can we get the animal to turn attention *inward* to their perceptual states in an experimental context?^[13]

5.3.2 Tactile Vibration

What of microstimulation in the absence of a stimulus? Might we induce *hallucinations* as Penfield did in his patients? Rather than modulation of ongoing perceptual processing, the issue here is to create an internal signal that mimics perception. Romo et al. (1998) demonstrated that monkeys can carry out sensory tasks via activation triggered by microstimulation. The monkeys' task was to discriminate the frequency of two sequential "flutters" on their fingertips, that is, mechanical vibrations on the skin at specific frequencies.

In an experimental trial, an initial *sample* flutter was presented for 500 ms and after a gap of 1–3 seconds, a second *test* flutter of either higher or lower frequency was presented. The animal reported whether the second test frequency was higher or lower than the sample.

The experimenters examined whether direct microstimulation in the absence of a stimulus could tap into the same neural representations that guided the animal's report. They isolated neurons in primary somatosensory cortex responsive to vibration frequency on the fingers (S1, the somatosensory homunculus discovered by Penfield & Boldrey [1937]; see [Figure 1](#)). The investigators then stimulated the same neurons in S1 in the absence of the test flutter, so used stimulation as a substitute for an actual vibration. Thus, the animal had to make a comparison between the frequency of a mechanical sample to either a subsequent (1) real mechanical test vibration (i.e., the good case with an actual stimulus) or (2) to a microstimulation test stimulus (i.e., the “hallucinatory” case where direct activation of the S1 neurons occurred in the absence of a stimulus). Romo et al. demonstrated that discrimination performance based either on mechanical stimulation or microstimulation was *equivalent*. In other words, the animals could match either mechanical or microstimulation to a remembered mechanical sample.^[14] (For a related approach with visual stimuli and optogenetic stimulation, see Azadi et al. 2023)

In subsequent work (Romo et al. 2000), the investigators inverted the experiment, using the microstimulation as the *sample*. In this case, the animals had to remember the information conveyed by the microstimulation (effectively, a hallucination) and then compare it to either a subsequent (a) mechanically generated stimulation on the finger (actual test stimulus) or (b) a microstimulation of S1 as test (i.e., no stimulus). In both cases, performance was similar to earlier results. The striking finding is that behavior could be driven entirely by microstimulation. At least for the tactile stimulations at issue, the animal might have been in the *Matrix*!

One might think that the intentional action inference is stronger in this paradigm, given the elegant flipping of stimuli in Romo et al. 2000. Still, the authors comment:

This study, therefore, has directly established a strong link between neural activity and perception. However, we do not know yet whether microstimulation of the QA circuit in S1 elicits a subjective flutter sensation in the fingertips. This can only be explored by microstimulating S1 in an attending human observer. (Romo et al. 2000: 277)^[15]

Like Newsome, the authors reach for introspection. Yet they might undersell their result, for it seems that the animals are having a tactile hallucination: (a) Penfield showed that stimulation of primary sensory cortices like S1 induces hallucinations in humans; (b) action is engaged not at low stimulation of S1 in monkeys but only at higher level stimulation; (c) at that point, when the stimulation grabs their attention, the monkeys do what they were trained to do, namely discriminate stimuli, either with (d) just mechanical stimulation (normal experience), or (e) with a mix of mechanical and microstimulation or with just microstimulation; (f) given the behavioral equivalence of these three cases, one might then argue that if performance in the mechanical stimulation cases involves conscious tactile experience, then that same experience is involved in the other cases.

Taken together, these cases provide examples of detailed manipulations in different sensory modalities, animals, and contents that test for causal sufficiency and necessity across different levels of the sensory processing hierarchy, from early levels (e.g., S1) to mid-levels (MT) and, as discussed at the outset of this section, to higher levels (ImFG or FFA). One issue that remains open is whether in tapping into neural processing by microstimulation, one has simply identified an earlier causal node in the neural processes that generate perceptual experience, there being more informative neural correlates later in the causal pathway.

6. The Future

Talk of the neuroscience of consciousness has, thus far, focused on the neural correlates of consciousness. Not all neural correlates are explanatory, so finding correlates is a first step in the neuroscience of consciousness. The next step involves manipulation of relevant correlates to test claims about sufficiency and necessity, as isolated in our two questions:

Generic Consciousness: What conditions/states N of nervous systems are necessary and (or) sufficient for a mental state, M , to be conscious as opposed to not?

Specific Consciousness: What neural states or properties are necessary and/or sufficient for a conscious perceptual state to have content X rather than Y ?

A productive neuroscience of consciousness requires that we understand the relevant neural properties at the right level of analysis. For generic consciousness, this will involve manipulation of relevant properties in a way that can avoid the access/phenomenal confound, and recent work focuses on pitting the many theories we have considered against each other. For specific consciousness, the critical issue will be to understand neural representational content and to find ways to link experimentally and explanatorily neural content to phenomenal content. We have tools to manipulate neural contents to affect phenomenal content, and in doing so, we can begin to uncover the neural basis of conscious contents. There is much interesting work yet to be done, philosophically and empirically, and we can look forward to a productive interdisciplinary research program.

What is consciousness, and could machines have it?

27 Oct 2017

Abstract

The controversial question of whether machines may ever be conscious must be based on a careful consideration of how consciousness arises in the only physical system that undoubtedly possesses it: the human brain. We suggest that the word “consciousness” conflates two different types of information-processing computations in the brain: the selection of information for global broadcasting, thus making it flexibly available for computation and report (C1, consciousness in the first sense), and the self-monitoring of those computations, leading to a subjective sense of certainty or error (C2, consciousness in the second sense). We argue that despite their recent successes, current machines are still mostly implementing computations that reflect unconscious processing (C0) in the human brain. We review the psychological and neural science of unconscious (C0) and conscious computations (C1 and C2) and outline how they may inspire novel machine architectures.

SIGN UP FOR THE AWARD-WINNING *SCIENCEADVISER* NEWSLETTER

The latest news, commentary, and research, free to your inbox daily

[SIGN UP](#)

Imagine that you are driving when you suddenly realize that the fuel-tank light is on. What makes you, a complex assembly of neurons, aware of the light? And what makes the car, a sophisticated piece of electronics and engineering, unaware of it? What would it take for the car to be endowed with a consciousness similar to our own? Are those questions scientifically tractable?

Alan Turing and John von Neumann, the founders of the modern science of computation, entertained the possibility that machines would ultimately mimic all of the brain’s abilities, including consciousness. Recent advances in artificial intelligence (AI) have revived this goal. Refinements in machine learning, inspired by neurobiology, have led to artificial neural networks that approach or, occasionally, surpass humans ([1](#), [2](#)). Although those networks do not mimic the biophysical properties of actual brains, their design benefitted from several neurobiological insights, including nonlinear input-output functions, layers with converging projections, and modifiable synaptic weights. Advances in computer hardware and training algorithms now allow such networks to operate on complex problems (such as machine translation) with success rates previously thought to be the privilege of real brains. Are they on the verge of consciousness? We argue that the answer is negative: The computations implemented by current deep-learning networks correspond mostly to nonconscious operations in the human brain. However, much like artificial neural networks took their inspiration from neurobiology, artificial consciousness may progress by investigating the architectures that allow the human brain to generate consciousness,

then transferring those insights into computer algorithms. Our aim is to foster such progress by reviewing aspects of the cognitive neuroscience of consciousness that may be pertinent for machines.

Multiple meanings of consciousness

The word “consciousness,” like many prescientific terms, is used in widely different senses. In a medical context, it is often used in an intransitive sense (as in, “the patient was no longer conscious”), in the context of assessing vigilance and wakefulness. Elucidating the brain mechanisms of vigilance is an essential scientific goal, with major consequences for our understanding of sleep, anesthesia, coma, or vegetative state. For lack of space, we do not deal with this aspect here, however, because its computational impact seems minimal: Obviously, a machine must be properly turned on for its computations to unfold normally.

We suggest that it is useful to distinguish two other essential dimensions of conscious computation. We label them using the terms global availability (C1) and self-monitoring (C2).

C1: Global availability

This corresponds to the transitive meaning of consciousness (as in “The driver is conscious of the light”). It refers to the relationship between a cognitive system and a specific object of thought, such as a mental representation of “the fuel-tank light.” This object appears to be selected for further processing, including verbal and nonverbal report. Information that is conscious in this sense becomes globally available to the organism; for example, we can recall it, act upon it, and speak about it. This sense is synonymous with “having the information in mind”; among the vast repertoire of thoughts that can become conscious at a given time, only that which is globally available constitutes the content of C1 consciousness.

C2: Self-monitoring

Another meaning of consciousness is reflexive. It refers to a self-referential relationship in which the cognitive system is able to monitor its own processing and obtain information about itself. Human beings know a lot about themselves, including such diverse information as the layout and position of their body, whether they know or perceive something, or whether they just made an error. This sense of consciousness corresponds to what is commonly called introspection, or what psychologists call “meta-cognition”—the ability to conceive and make use of internal representations of one’s own knowledge and abilities.

We propose that C1 and C2 constitute orthogonal dimensions of conscious computations. This is not to say that C1 and C2 do not involve overlapping physical substrates; in fact, as we review below, in the human brain both depend on the prefrontal cortex. But we argue that empirically and conceptually, the two may come apart because there can be C1 without C2, such as when reportable processing is not accompanied by accurate metacognition, or C2 without C1, such as when a self-monitoring operation unfolds without being consciously reportable. As such, it is advantageous to consider these computations separately before we consider their synergy. Furthermore, many computations involve neither C1 nor C2 and therefore are properly called “unconscious” (or C0 for short). It was Turing’s original insight that even sophisticated information processing can be realized by a mindless automaton. Cognitive neuroscience

confirms that complex computations such as face or speech recognition, chess-game evaluation, sentence parsing, and meaning extraction occur unconsciously in the human brain—under conditions that yield neither global reportability nor self-monitoring ([Table 1](#)). The brain appears to operate, in part, as a juxtaposition of specialized processors or “modules” that operate nonconsciously and, we argue, correspond tightly to the operation of current feedforward deep-learning networks.

Computation	Examples of experimental findings	References
<i>C0: Unconscious processing</i>		
Invariant visual recognition	Subliminal priming by unseen words and faces, invariant for font, size, or viewpoint.	(5)
	Functional MRI (fMRI) and single-neuron response to unseen words and faces	(33 , 37 , 78 , 79)
	Unconscious judgement of chess game configurations	(80)
Access to meaning	N400 response to unseen out-of-context words	(9 , 10)
Cognitive control	Unconscious inhibition or task set preparation by an unseen cue	(11 , 12)
Reinforcement learning	Subliminal instrumental conditioning by unseen shapes	(17)

Computation	Examples of experimental findings	References
<i>C1: Global availability of information</i>		
All-or-none selection and broadcasting of a relevant content	Conscious perception of a single picture during visual rivalry	(29)
	Conscious perception of a single detail in a picture or stream	(28 , 81)
	All-or-none memory retrieval	(82)
	Attentional blink: Conscious perception of item A prevents the simultaneous perception of item B	(27 , 30 , 83 , 84)
	All-or-none “ignition” of event-related potentials and fMRI signals, only on trials with conscious perception	(33–35 , 85–87)
	All-or-none firing of neurons coding for the perceived object in prefrontal cortex and other higher areas	(31 , 32 , 37 , 38 , 88)

Computation	Examples of experimental findings	References	
Stabilization of short-lived information for off-line processing	Brain states are more stable when information is consciously perceived; unconscious information quickly decays (~1 s)	(39 , 89)	
		Conscious access may occur long after the stimulus is gone	(90)
Flexible routing of information	Only conscious information can be routed through a series of successive operations (for example, successive calculations $3 \times 4 + 2$)	(91)	
Sequential performance of several tasks	Psychological refractory period: Conscious processing of item A delays conscious processing of item B	(34 , 92)	
	Serial calculations or strategies require conscious perception	(13 , 91)	

Computation	Examples of experimental findings	References	
		Serial organization of spontaneous brain activity during conscious thought in the “resting state”	(93)
<i>C2: Self-monitoring</i>			
Self-confidence	Humans accurately report subjective confidence, a probabilistic estimate in the accuracy of a decision or computation	(51, 55)	
Evaluation of one’s knowledge	Humans and animals can ask for help or “opt out” when unsure	(53, 65, 66)	
		Humans and animals know when they do not know or remember	(49, 53)
Error detection	Anterior cingulate response to self-detected errors	(61, 65, 94)	
Listing one’s skills	Children know the arithmetic procedures at their disposal, their speed, and error rate.	(70)	

Computation	Examples of experimental findings	References
Sharing one's confidence with others	Decision-making improves when two persons share knowledge	(69)

Expand for more

Table 1 Examples of computations pertaining to information-processing levels C0, C1 and C2 in the human brain.

[OPEN IN VIEWER](#)

We next review the experimental evidence for how human and animal brains handle C0-, C1-, and C2-level computations, before returning to machines and how they could benefit from this understanding of brain architecture.

Unconscious processing (C0): Where most of our intelligence lies

“We cannot be conscious of what we are not conscious of” (3). This truism has deep consequences. Because we are blind to our unconscious processes, we tend to underestimate their role in our mental life. However, cognitive neuroscientists developed various means of presenting images or sounds without inducing any conscious experience (Fig. 1) and then used behavioral and brain imaging to probe their processing depth.

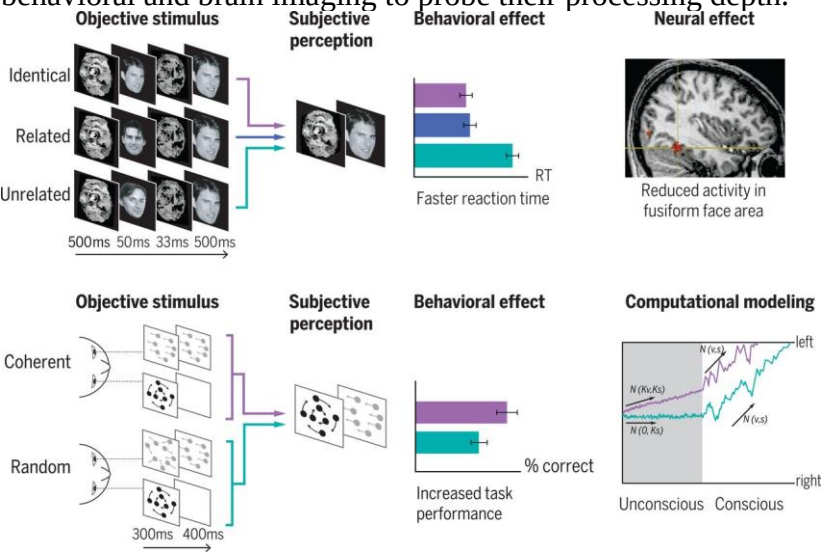


Fig. 1 Examples of paradigms probing unconscious processing (C0).

(Top) Subliminal view-invariant face recognition ([77](#)). On each trial, a prime face is briefly presented (50 ms), surrounded by masks that make it invisible, followed by a visible target face (500 ms). Although subjective perception is identical across conditions, processing is facilitated whenever the two faces represent the same person, in same or different view. At the behavioral level, this view-invariant unconscious priming is reflected in reduced reaction time in recognizing the target face. At the neural level, it is reflected in reduced cortical response to the target face (repetition suppression) in the fusiform face area of the human inferotemporal cortex. **(Bottom)** Subliminal accumulation of evidence during interocular suppression ([16](#)). Presentation of salient moving dots in one eye prevents the conscious perception of paler moving dots in the opposite eye. Despite their invisibility, the gray dots facilitate performance when they moved in the same direction as a subsequent dot display, an effect proportional to their amount of motion coherence. This facilitation only affects a first-order task (judging the direction of motion), not a second-order metacognitive judgement (rating the confidence in the first response). A computational model of evidence accumulation proposes that subliminal motion information gets added to conscious information, thus biasing and shortening the decision.

Expand for more

[OPEN IN VIEWER](#)

The phenomenon of priming illustrates the remarkable depth of unconscious processing. A highly visible target stimulus, such as the written word “four,” is processed more efficiently when preceded by a related prime stimulus, such as the Arabic digit “4,” even when subjects do not notice the presence of the prime and cannot reliably report its identity. Subliminal digits, words, faces, or objects can be invariantly recognized and influence motor, semantic, and decision levels of processing ([Table 1](#)). Neuroimaging methods reveal that the vast majority of brain areas can be activated nonconsciously.

Unconscious view-invariance and meaning extraction in the human brain

Many of the difficult perceptual computations, such as invariant face recognition or speaker-invariant speech recognition, that were recently addressed by AI correspond to nonconscious computations in the human brain ([4–6](#)). For instance, processing someone’s face is facilitated when it is preceded by the subliminal presentation of a totally different view of the same person, indicating unconscious invariant recognition ([Fig. 1](#)). Subliminal priming generalizes across visual-auditory modalities ([7, 8](#)), revealing that cross-modal computations that remain challenging for AI software (such as extraction of semantic vectors or speech-to-text) also involve unconscious mechanisms. Even the semantic meaning of sensory input can be processed without awareness by the human brain. Compared with related words (for example, animal-dog), semantic violations (for example, furniture-dog) generate a brain response as late as 400 ms after stimulus onset in temporal-lobe language networks, even if one of the two words cannot be consciously detected ([9, 10](#)).

Unconscious control and decision-making

Unconscious processes can reach even deeper levels of the cortical hierarchy. For instance, subliminal primes can influence prefrontal mechanisms of cognitive control involved in the selection of a task ([11](#)) or the inhibition of a motor response ([12](#)). Neural mechanisms of decision-making involve accumulating sensory evidence that affects the probability of the various choices until a threshold is attained. This accumulation of probabilistic knowledge continues to happen even with subliminal stimuli ([13–16](#)). Bayesian inference and evidence accumulation, which are cornerstone computations for AI ([2](#)), are basic unconscious mechanisms for humans.

Unconscious learning

Reinforcement learning algorithms, which capture how humans and animals shape their future actions on the basis of history of past rewards, have excelled in attaining supra-human AI performance in several applications, such as playing Go ([1](#)). Remarkably, in humans, such learning appears to proceed even when the cues, reward, or motivation signals are presented below the consciousness threshold ([17, 18](#)).

Complex unconscious computations and inferences routinely occur in parallel within various brain areas. Many of these C0 computations have now been captured by AI, particularly by using feedforward convolutional neural networks (CNNs). We next consider what additional computations are required for conscious processing.

C1: Global availability of relevant information

The need for integration and coordination

The organization of the brain into computationally specialized subsystems is efficient, but this architecture also raises a specific computational problem: The organism as a whole cannot stick to a diversity of probabilistic interpretations; it must act and therefore cut through the multiple

possibilities and decide in favor of a single course of action. Integrating all of the available evidence to converge toward a single decision is a computational requirement that, we contend, must be faced by any animal or autonomous AI system and corresponds to our first functional definition of consciousness: global availability (C1).

For example, elephants, when thirsty, manage to determine the location of the nearest water hole and move straight to it, from a distance of 5 to 50 km ([19](#)). Such decision-making requires a sophisticated architecture for (i) efficiently pooling over all available sources of information, including multisensory and memory cues; (ii) considering the available options and selecting the best one on the basis of this large information pool; (iii) sticking to this choice over time; and (iv) coordinating all internal and external processes toward the achievement of that goal.

Primitive organisms, such as bacteria, may achieve such decision solely through an unconscious competition of uncoordinated sensorimotor systems. This solution, however, fails as soon as it becomes necessary to bridge over temporal delays and to inhibit short-term tendencies in favor of longer-term winning strategies. Coherent, thoughtful planning required a specific C1 architecture.

Consciousness as access to an internal global workspace

We hypothesize that consciousness in the first sense (C1) evolved as an information-processing architecture that addresses this information-pooling problem ([20–23](#)). In this view, the architecture of C1 evolved to break the modularity and parallelism of unconscious computations. On top of a deep hierarchy of specialized modules, a “global neuronal workspace,” with limited capacity, evolved to select a piece of information, hold it over time, and share it across modules. We call “conscious” whichever representation, at a given time, wins the competition for access to this mental arena and gets selected for global sharing and decision-making. Consciousness is therefore manifested by the temporary dominance of a thought or train of thoughts over mental processes, so that it can guide a broad variety of behaviors. These behaviors include not only physical actions but also mental ones, such as committing information to episodic memory or routing it to other processors.

Relation between consciousness and attention

William James described attention as “the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought” ([24](#)). This definition is close to what we mean by C1: the selection of a single piece of information for entry into the global workspace. There is, however, a clear-cut distinction between this final step, which corresponds to conscious access, and the previous stages of attentional selection, which can operate unconsciously. Many experiments have established the existence of dedicated mechanisms of attention orienting and shown that, like any other processors, they can operate nonconsciously: (i) In the top-down direction, attention can be oriented toward an object, amplify its processing, and yet fail to bring it to consciousness ([25](#)); (ii) in the bottom-up direction, attention can be attracted by a flash, even if this stimulus ultimately remains unconscious ([26](#)). What we call attention is a hierarchical system of sieves that operate unconsciously. Such unconscious systems compute with probability distributions, but only a single sample, drawn from this probabilistic distribution, becomes conscious at a given time

([27](#), [28](#)). We may become aware of several alternative interpretations, but only by sampling their unconscious distributions over time ([29](#), [30](#)).

Evidence for all-or-none selection in a capacity-limited system

The primate brain comprises a conscious bottleneck and can only consciously access a single item at a time ([Table 1](#)). For instance, rivaling pictures or ambiguous words are perceived in an all-or-none manner; at any given time, we subjectively perceive only a single interpretation out of many possible ones [even though the others continue to be processed unconsciously ([31](#), [32](#))]. The serial operation of consciousness is attested by phenomena such as the attentional blink and the psychological refractory period, in which conscious access to a first item A prevents or delays the perception of a second competing item B ([9](#), [27](#), [30](#), [33–35](#)). Such interference with the perception of B is triggered by the mere conscious perception of A, even if no task is performed ([36](#)). Thus C1 consciousness is causally responsible for a serial information-processing bottleneck.

Evidence for integration and broadcasting

Brain imaging in humans and neuronal recordings in monkeys indicate that the conscious bottleneck is implemented by a network of neurons that is distributed through the cortex, but with a stronger emphasis on high-level associative areas. Listed in [Table 1](#) are some of the publications that have evidenced an all-or-none “ignition” of this network during conscious perception by using a variety of brain-imaging techniques. Single-cell recordings indicate that each specific conscious percept, such as a person’s face, is encoded by the all-or-none firing of a subset of neurons in high-level temporal and prefrontal cortices, whereas others remain silent ([Fig. 2](#)) ([31](#), [32](#), [37](#), [38](#)).

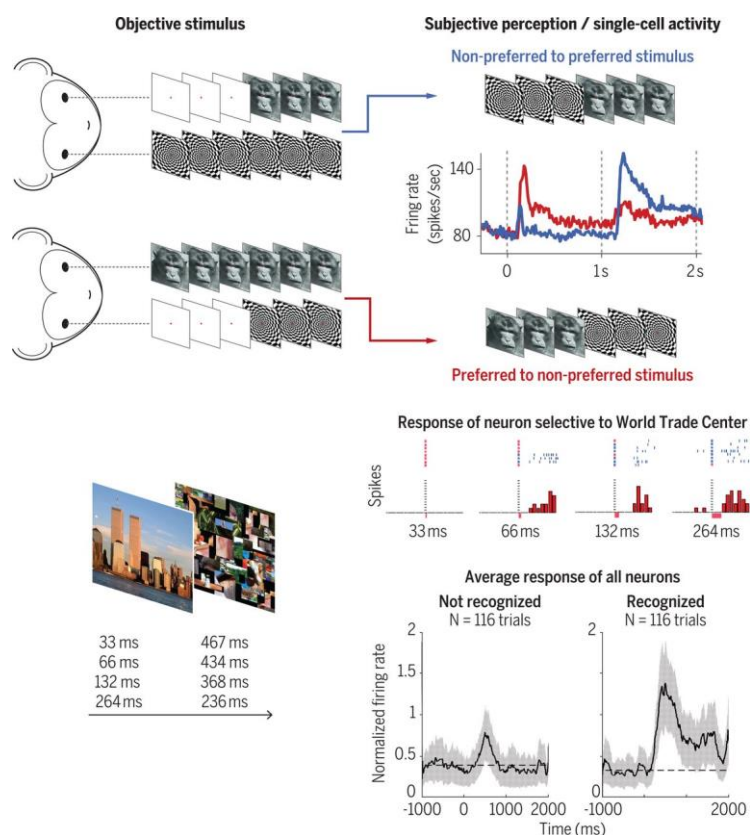


Fig. 2 Global availability: Consciousness in the first sense (C1).

Conscious subjective percepts are encoded by the sudden firing of stimulus-specific neural populations distributed in interconnected, high-level cortical areas, including the lateral prefrontal cortex, anterior temporal cortex, and hippocampus. **(Top)** During binocular flash suppression, the flashing of a picture to one eye suppresses the conscious perception of a second picture presented to the other eye. As a result, the same physical stimulus can lead to distinct subjective percepts. This example illustrates a prefrontal neuron sensitive to faces and unresponsive to checkers, whose firing shoots up in tight association with the sudden onset of subjective face perception ([31](#)). **(Bottom)** During masking, a flashed image, if brief enough and followed by a longer “mask,” can remain subjectively invisible. Shown is a neuron in the entorhinal cortex firing selectively to the concept of “World Trade Center.” Rasters in red indicate trials in which the subject reported recognizing the picture (blue

indicates no recognition). Under masking, when the picture is presented for only 33 ms there is little or no neural activity; but once presentation time is longer than the perceptual threshold (66 ms or larger), the neuron fires substantially only on recognized trials. Overall, even for identical objective input (same duration), spiking activity is higher and more stable for recognized trials ([38](#)).

Expand for more

[OPEN IN VIEWER](#)

Stability as a feature of consciousness

Direct contrasts between seen and unseen pictures or words confirm that such ignition occurs only for the conscious percept. As explained earlier, nonconscious stimuli may reach into deep cortical networks and influence higher levels of processing and even central executive functions, but these effects tend to be small, variable, and short-lived [although nonconscious information decays at a slower rate than initially expected ([39](#), [40](#))]. By contrast, the stable, reproducible representation of high-quality information by a distributed activity pattern in higher cortical areas is a feature of conscious processing ([Table 1](#)). Such transient “meta-stability” seems to be necessary for the nervous system to integrate information from a variety of modules and then broadcast it back to them, achieving flexible cross-module routing.

C1 consciousness in human and nonhuman animals

C1 consciousness is an elementary property that is present in human infants ([41](#)) as well as in animals. Nonhuman primates exhibit similar visual illusions ([31](#), [32](#)), attentional blink ([42](#)), and central capacity limits ([43](#)) as human subjects. The prefrontal cortex appears to act as a central information sharing device and serial bottleneck in both human and nonhuman primates ([43](#)). The considerable expansion of the prefrontal cortex in the human lineage may have resulted in a greater capacity for multimodal convergence and integration ([44–46](#)). Furthermore, humans possess additional circuits in the inferior prefrontal cortex for verbally formulating and reporting information to others. The capacity to report information through language is universally considered one of the clearest signs of conscious perception because once information has reached this level of representation in humans, it is necessarily available for sharing across mental modules and therefore conscious in the C1 sense. Thus, although language is not required for conscious perception and processing, the emergence of language circuits in humans may have resulted in a considerable increase in the speed, ease, and flexibility of C1-level information sharing.

C2: Self-monitoring

Whereas C1 consciousness reflects the capacity to access external information, consciousness in the second sense (C2) is characterized by the ability to reflexively represent oneself (47–50). A substantial amount of research in cognitive neuroscience and psychology has addressed self-monitoring under the term of “metacognition,” which is roughly defined as cognition about cognition or knowing about knowing. Below, we review the mechanisms by which the primate brain monitors itself, while stressing their implications for building self-reflective machines.

A probabilistic sense of confidence

When making a decision, humans feel more or less confident about their choice. Confidence can be defined as a sense of the probability that a decision or computation is correct (51). Almost anytime the brain perceives or decides, it also estimates its degree of confidence. Learning is also accompanied by a quantitative sense of confidence; humans evaluate how much trust they have in what they have learned and use it to weigh past knowledge versus present evidence (52). Confidence can be assessed nonverbally, either retrospectively, by measuring whether humans persist in their initial choice, or prospectively, by allowing them to opt out from a task without even attempting it. Both measures have been used in nonhuman animals to show that they too possess metacognitive abilities (53). By contrast, most current neural networks lack them: Although they can learn, they generally lack meta-knowledge of the reliability and limits of what has been learned. A noticeable exception is biologically constrained models that rely on Bayesian mechanisms to simulate the integration of multiple probabilistic cues in neural circuits (54). These models have been fruitful in describing how neural populations may automatically compute the probability that a given process is performed successfully. Although these implementations remain rare and have not addressed the same range of computational problems as has traditional AI, they offer a promising venue for incorporating uncertainty monitoring in deep learning networks.

Explicit confidence in prefrontal cortex

According to Bayesian accounts, each local cortical circuit may represent and combine probability distributions in order to estimate processing uncertainty (54). However, additional neural circuits may be required in order to explicitly extract and manipulate confidence signals. Magnetic resonance imaging (MRI) studies in humans and physiological recordings in primates and even in rats have specifically linked such confidence processing to the prefrontal cortex (55–57). Inactivation of the prefrontal cortex can induce a specific deficit in second-order (metacognitive) judgements while sparing performance on the first-order task (56, 58). Thus, circuits in the prefrontal cortex may have evolved to monitor the performance of other brain processes.

“...it is useful to distinguish two essential dimensions of conscious computation.”

Error detection: Reflecting on one’s own mistakes

Error detection provides a particularly clear example of self-monitoring; just after responding, we sometimes realize that we made an error and change our mind. Error detection is reflected by two components of electroencephalography (EEG) activity: the error-related negativity (ERN) and the positivity upon error (Pe), which emerge in cingulate and medial prefrontal cortex just

after a wrong response but before any feedback is received. How can the brain make a mistake and detect it? One possibility is that the accumulation of sensory evidence continues after a decision is made, and an error is inferred whenever this further evidence points in the opposite direction (59). A second possibility, more compatible with the remarkable speed of error detection, is that two parallel circuits, a low-level sensory-motor circuit and a higher-level intention circuit, operate on the same sensory data and signal an error whenever their conclusions diverge (60, 61).

Meta-memory

Humans do not just know things about the world; they actually know that they know or that they do not know. A familiar example is having a word “on the tip of the tongue.” The term “meta-memory” was coined to capture the fact that humans report feelings of knowing, confidence, and doubts on their memories. Meta-memory is thought to involve a second-order system that monitors internal signals (such as the strength and quality of a memory trace) to regulate behavior. Meta-memory is associated with prefrontal structures whose pharmacological inactivation leads to a metacognitive impairment while sparing memory performance itself (56). Metamemory is crucial to human learning and education by allowing learners to develop strategies such as increasing the amount of study or adapting the time allocated to memory encoding and rehearsal (49).

Reality monitoring

In addition to monitoring the quality of sensory and memory representations, the human brain must also distinguish self-generated versus externally driven representations. Indeed, we can perceive things, but we also conjure them from imagination or memory. Hallucinations in schizophrenia have been linked to a failure to distinguish whether sensory activity is generated by oneself or by the external world (62). Neuroimaging studies have linked this kind of reality monitoring to the anterior prefrontal cortex (63). In nonhuman primates, neurons in the prefrontal cortex distinguish between normal visual perception and active maintenance of the same visual content in memory (64).

Foundations of C2 consciousness in infants

Self-monitoring is such a basic ability that it is already present during infancy (Fig. 3). The ERN, indicating error monitoring, was observed when 1-year-old infants made a wrong choice in a perceptual decision task (65). Similarly, after 1-½-year-old infants pointed to one of two boxes in order to obtain a hidden toy, they waited longer for an upcoming reward (such as a toy) when their initial choice was correct than when it was wrong, suggesting that they monitored the likelihood that their decision was right (57, 65). Moreover, when given the opportunity to ask (nonverbally) their parents for help they chose this opt-out option specifically in trials in which they were likely to be wrong, revealing a prospective estimate of their own uncertainty (66). That infants can communicate their own uncertainty to other agents further suggests that they consciously experience metacognitive information. Thus, infants are already equipped with the ability to monitor their own mental states. Facing a world where everything remains to be

learned, C2 mechanisms allow them to actively orient toward domains that they know they do not know—a mechanism that we call “curiosity.”

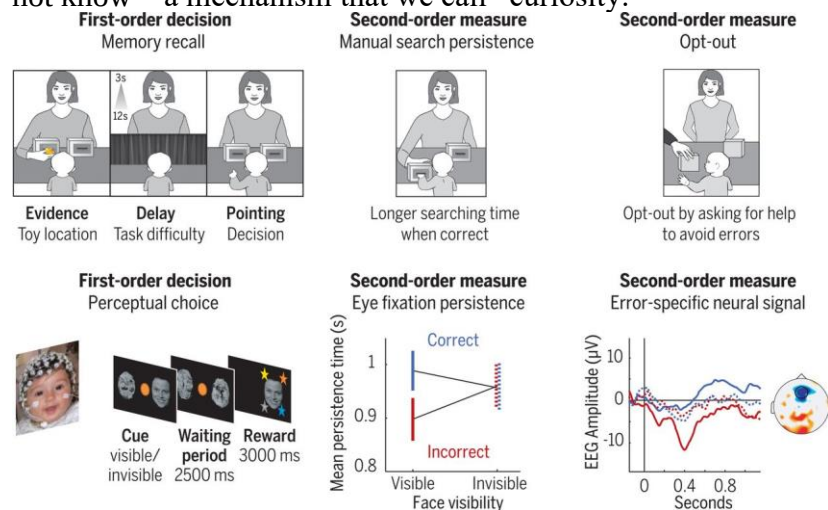


Fig. 3 Self-monitoring: Consciousness in the second sense (C2).

Self-monitoring (also called “meta-cognition”), the capacity to reflect on one’s own mental state, is available early during infancy. **(Top)** One-and-a-half-year-old infants, after deciding to point to the location of a hidden toy, exhibit two types of evidence for self-monitoring of their decision. (i) They persist longer in searching for the hidden object within the selected box when their initial choice was correct than when it was incorrect. (ii) When given the opportunity to ask for help, they use this option selectively to reduce the probability of making an error. **(Bottom)** One-year-old infants were presented with either a meaningless pattern or a face that was either visible or invisible (depending on its duration) and then decided to gaze left or right in anticipation of face reappearance. As for manual search, post-decision persistence in waiting at the same gaze location increased for correct compared with incorrect initial decisions. Moreover, EEG signals revealed the presence of the error-related negativity over fronto-central electrodes when infants make an incorrect choice. These markers of metacognition were elicited by visible but not by invisible stimuli, as also shown in adults ([61](#)).

Dissociations between C1 and C2

According to our analysis, C1 and C2 are largely orthogonal and complementary dimensions of what we call consciousness. On one side of this double dissociation, self-monitoring can exist for unreportable stimuli (C2 without C1). Automatic typing provides a good example: Subjects slow down after a typing mistake, even when they fail to consciously notice the error (67). Similarly, at the neural level, an ERN can occur for subjectively undetected errors (68). On the other side of this dissociation, consciously reportable contents sometimes fail to be accompanied with an adequate sense of confidence (C1 without C2). For instance, when we retrieve a memory, it pops into consciousness (C1) but sometimes without any accurate evaluation of its confidence (C2), leading to false memories. As noted by Marvin Minsky, “what we call consciousness [in the C1 sense] is a very imperfect summary in one part of the brain of what the rest is doing.” The imperfection arises in part from the fact that the global workspace reduces complex parallel sensory streams of probabilistic computation to a single conscious sample (27–29). Thus, probabilistic information is often lost on the way, and subjects feel over-confident in the accuracy of their perception.

Synergies between C1 and C2 consciousness

Because C1 and C2 are orthogonal, their joint possession may have synergistic benefits to organisms. In one direction, bringing probabilistic metacognitive information (C2) into the global workspace (C1) allows it to be held over time, integrated into explicit long-term reflection, and shared with others. Social information sharing improves decisions: By sharing their confidence signals, two persons achieve a better performance in collective decision-making than that of either person alone (69). In the converse direction, the possession of an explicit repertoire of one’s own abilities (C2) improves the efficiency with which C1 information is processed. During mental arithmetic, children can perform a C2-level evaluation of their available competences (for example, counting, adding, multiplying, or memory retrieval) and use this information to evaluate how to best face a given arithmetic problem (70). This functionality requires a single “common currency” for confidence across different modules, which humans appear to possess (71).

Endowing machines with C1 and C2

How could machines be endowed with C1 and C2 computations? Let us return to the car light example. In current machines, the “low gas” light is a prototypical example of an unconscious modular signal (C0). When the light flashes, all other processors in the machine remain uninformed and unchanged; fuel continues to be injected in the carburetor, and the car passes gas stations without stopping (although they might be present on the GPS map). Current cars or cell phones are mere collections of specialized modules that are largely “unaware” of each other. Endowing this machine with global information availability (C1) would allow these modules to share information and collaborate to address the impending problem (much like humans do when they become aware of the light, or elephants of thirst).

Although AI has met considerable success in solving specific problems, implementing multiple processes in a single system and flexibly coordinating them remain difficult problems. In the 1960s, computational architectures called “blackboard systems” were specifically designed to post information and make it available to other modules in a flexible and interpretable manner, similar in flavor to a global workspace (20). A recent architecture called Pathnet uses a genetic algorithm to learn which path through its many specialized neural networks is most suited to a given task (72). This architecture exhibits robust, flexible performance and generalization across tasks and may constitute a first step toward primate-like conscious flexibility.

“...the ability to reflexively represent oneself.”

To make optimal use of the information provided by the fuel-gauge light, it would also be useful for the car to possess a database of its own capacities and limits. Such self-monitoring (C2) would include an integrated image of itself—including its current location and fuel consumption, for example—as well as its internal databases (such as “knowing” that it possesses a GPS map that can locate gas stations). A self-monitoring machine would keep a list of its subprograms, compute estimates of their probabilities of succeeding at various tasks, and constantly update them (for example, noticing when a part fails).

Most present-day machine-learning systems are devoid of any self-monitoring; they compute (C0) without representing the extent and limits of their knowledge or the fact that others may have a different viewpoint than their own. There are a few exceptions: Bayesian networks (54) or programs (73) compute with probability distributions and therefore keep track of how likely they are to be correct. Even when the primary computation is performed by a classical CNN, and is therefore opaque to introspection, it is possible to train a second, hierarchically higher neural network to predict the first one’s performance (47). This approach, in which a system redescribes itself, has been claimed to lead to “the emergence of internal models that are metacognitive in nature and... make it possible for an agent to develop a (limited, implicit, practical) understanding of itself” (48). Pathnet (72) uses a related architecture to track which internal configurations are most successful at a given task and use this knowledge to guide subsequent processing. Robots have also been programed to monitor their learning progress and use it to orient resources toward the problems that maximize information gain, thus implementing a form of curiosity (74).

An important element of C2 that has received relatively little attention is reality monitoring. Bayesian approaches to AI (2, 73) have recognized the usefulness of learning generative models that can be jointly used for actual perception (present), prospective planning (future), and retrospective analysis (past). In humans, the same sensory areas are involved in both perception and imagination. As such, some mechanisms are needed to tell apart self-generated versus externally triggered activity. A powerful method for training generative models, called adversarial learning (75), involves having a secondary network “compete” against a generative network so as to critically evaluate the authenticity of self-generated representations. When such reality monitoring (C2) is coupled with C1 mechanisms, the resulting machine may more closely mimic human consciousness in terms of affording global access to perceptual representations while having an immediate sense that their content is a genuine reflection of the current state of the world.

Concluding remarks

Our stance is based on a simple hypothesis: What we call “consciousness” results from specific types of information-processing computations, physically realized by the hardware of the brain. It differs from other theories in being resolutely computational; we surmise that mere information-theoretic quantities (76) do not suffice to define consciousness unless one also considers the nature and depth of the information being processed.

We contend that a machine endowed with C1 and C2 would behave as though it were conscious; for instance, it would know that it is seeing something, would express confidence in it, would report it to others, could suffer hallucinations when its monitoring mechanisms break down, and may even experience the same perceptual illusions as humans. Still, such a purely functional definition of consciousness may leave some readers unsatisfied. Are we “over-intellectualizing” consciousness, by assuming that some high-level cognitive functions are necessarily tied to consciousness? Are we leaving aside the experiential component (“what it is like” to be conscious)? Does subjective experience escape a computational definition?

Although those philosophical questions lie beyond the scope of the present paper, we close by noting that empirically, in humans the loss of C1 and C2 computations covaries with a loss of subjective experience. For example, in humans, damage to the primary visual cortex may lead to a neurological condition called “blindsight,” in which the patients report being blind in the affected visual field. Remarkably, those patients can localize visual stimuli in their blind field but cannot report them (C1), nor can they effectively assess their likelihood of success (C2)—they believe that they are merely “guessing.” In this example, at least, subjective experience appears to cohere with possession of C1 and C2. Although centuries of philosophical dualism have led us to consider consciousness as unreducible to physical interactions, the empirical evidence is compatible with the possibility that consciousness arises from nothing more than specific computations.



Conscious robots and computers has long been a popular science fiction theme. Why do they lack consciousness in reality, and how might they develop it?

CREDIT: MENNOVANDIJK/ISTOCK

[OPEN IN VIEWER](#)

Acknowledgments

This work was supported by funding from Institut National de la Santé et de la Recherche Médicale, CEA, Canadian Institute for Advanced Research, Spoelberch Foundation, Collège de France, the European Research Council (ERC project NeuroConsc to S.D. and ERC project METAWARE to S.K), the French Agence National de la Recherche (grants ANR-10-LABX-0087 and ANR-10-IDEX-0001-02), the U.S. National Institute of Health (National Institute of Neurological Disorders and Stroke grant R01NS088628 to H.L.), and the U.S. Air Force Office of Scientific Research (grant FA9550-15-1-0110 to H.L.)

References and Notes

1

D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, D. Hassabis, Mastering the game of Go with deep neural networks and tree search. *Nature* **529**, 484–489 (2016). 10.1038/nature16961

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

2

B. M. Lake, T. D. Ullman, J. B. Tenenbaum, S. J. Gershman, Building machines that learn and think like people. *Behav. Brain Sci.* **2016**, 1–101 (2016). 10.1017/S0140525X16001837

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

3

J. Jaynes, *The Origin of Consciousness in the Breakdown of the Bicameral Mind* (Houghton Mifflin Company, 1976).

[GO TO REFERENCE](#)

[GOOGLE SCHOLAR](#)

4

S. Kouider, E. Dupoux, Subliminal speech priming. *Psychol. Sci.* **16**, 617–625 (2005). 10.1016/j.cognition.2008.06.012

[GO TO REFERENCE](#)

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

5

S. Kouider, S. Dehaene, Levels of processing during non-conscious perception: A critical review of visual masking. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **362**, 857–875 (2007). 10.1098/rstb.2007.2093

[GO TO REFERENCE](#)

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

6

E. Qiao, F. Vinckier, M. Szwed, L. Naccache, R. Valabrègue, S. Dehaene, L. Cohen, Unconsciously deciphering handwriting: Subliminal invariance for handwritten words in the visual word form area. *Neuroimage* **49**, 1786–1799 (2010). 10.1016/j.neuroimage.2009.09.034

[GO TO REFERENCE](#)

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

7

N. Faivre, L. Mudrik, N. Schwartz, C. Koch, Multisensory integration in complete unawareness: Evidence from audiovisual congruency priming. *Psychol. Sci.* **25**, 2006–2016 (2014). 10.1177/0956797614547916

[GO TO REFERENCE](#)

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

8

S. Kouider, S. Dehaene, Subliminal number priming within and across the visual and auditory modalities. *Exp. Psychol.* **56**, 418–433 (2009). 10.1027/1618-3169.56.6.418

[GO TO REFERENCE](#)

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

9

S. J. Luck, E. K. Vogel, K. L. Shapiro, Word meanings can be accessed but not reported during the attentional blink. *Nature* **383**, 616–618 (1996). 10.1038/383616a0

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

10

S. van Gaal, L. Naccache, J. D. I. Meuwese, A. M. van Loon, A. H. Leighton, L. Cohen, S. Dehaene, Can the meaning of multiple words be integrated unconsciously? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **369**, 20130212 (2014). 10.1098/rstb.2013.0212

[CROSSREF](#)
[PUBMED](#)
[WEB OF SCIENCE](#)
[GOOGLE SCHOLAR](#)

11

H. C. Lau, R. E. Passingham, Unconscious activation of the cognitive control system in the human prefrontal cortex. *J. Neurosci.* **27**, 5805–5811 (2007). 10.1523/JNEUROSCI.4335-06.2007

[CROSSREF](#)
[PUBMED](#)
[WEB OF SCIENCE](#)
[GOOGLE SCHOLAR](#)

12

S. van Gaal, V. A. Lamme, J. J. Fahrenfort, K. R. Ridderinkhof, Dissociable brain mechanisms underlying the conscious and unconscious control of behavior. *J. Cogn. Neurosci.* **23**, 91–105 (2011). 10.1162/jocn.2010.21431

[CROSSREF](#)
[PUBMED](#)
[WEB OF SCIENCE](#)
[GOOGLE SCHOLAR](#)

13

F. P. de Lange, S. van Gaal, V. A. Lamme, S. Dehaene, How awareness changes the relative weights of evidence during human decision-making. *PLOS Biol.* **9**, e1001203 (2011). 10.1371/journal.pbio.1001203

[CROSSREF](#)
[PUBMED](#)
[WEB OF SCIENCE](#)
[GOOGLE SCHOLAR](#)

14

D. Vorberg, U. Mattler, A. Heinecke, T. Schmidt, J. Schwarzbach, Different time courses for visual perception and action priming. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 6275–6280 (2003). 10.1073/pnas.0931489100

[CROSSREF](#)
[PUBMED](#)
[WEB OF SCIENCE](#)
[GOOGLE SCHOLAR](#)

15

S. Dehaene, L. Naccache, G. Le Clec'H, E. Koechlin, M. Mueller, G. Dehaene-Lambertz, P.-F. Ç. van de Moortele, D. Le Bihan, Imaging unconscious semantic priming. *Nature* **395**, 597–600 (1998). 10.1038/26967

[CROSSREF](#)
[PUBMED](#)
[WEB OF SCIENCE](#)
[GOOGLE SCHOLAR](#)

16

A. Vlassova, C. Donkin, J. Pearson, Unconscious information changes decision accuracy but not confidence. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 16214–16218 (2014).
10.1073/pnas.1403619111

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

17

M. Pessiglione, P. Petrovic, J. Daunizeau, S. Palminteri, R. J. Dolan, C. D. Frith, Subliminal instrumental conditioning demonstrated in the human brain. *Neuron* **59**, 561–567 (2008).
10.1016/j.neuron.2008.07.005

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

18

M. Pessiglione, L. Schmidt, B. Draganski, R. Kalisch, H. Lau, R. J. Dolan, C. D. Frith, How the brain translates money into force: A neuroimaging study of subliminal motivation. *Science* **316**, 904–906 (2007). 10.1126/science.1140459

[GO TO REFERENCE](#)

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

19

L. Polansky, W. Kilian, G. Wittemyer, Elucidating the significance of spatial memory on movement decisions by African savannah elephants using state-space models. *Proc. Biol. Sci.* **282**, 20143042 (2015). 10.1098/rspb.2014.3042

[GO TO REFERENCE](#)

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

20

B. Baars, *A Cognitive Theory of Consciousness* (Cambridge Univ. Press, 1988).

[GOOGLE SCHOLAR](#)

21

S. Dehaene, M. Kerszberg, J. P. Changeux, A neuronal model of a global workspace in effortful cognitive tasks. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14529–14534 (1998).
10.1073/pnas.95.24.14529

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

22

D. Dennett, Are we explaining consciousness yet? *Cognition* **79**, 221–237 (2001).
10.1016/S0010-0277(00)00130-X

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

23

S. Dehaene, L. Naccache, Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition* **79**, 1–37 (2001). 10.1016/S0010-0277(00)00123-2

[GO TO REFERENCE](#)

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

24

W. James, *The Principles of Psychology* (Holt, 1890).

[GO TO REFERENCE](#)

[GOOGLE SCHOLAR](#)

25

L. Naccache, E. Blandin, S. Dehaene, Unconscious masked priming depends on temporal attention. *Psychol. Sci.* **13**, 416–424 (2002). 10.1111/1467-9280.00474

[GO TO REFERENCE](#)

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

26

R. W. Kentridge, C. A. Heywood, L. Weiskrantz, Attention without awareness in blindsight. *Proc. Biol. Sci.* **266**, 1805–1811 (1999). 10.1098/rspb.1999.0850

[GO TO REFERENCE](#)

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

27

C. L. Asplund, D. Fugnie, S. Zughni, J. W. Martin, R. Marois, The attentional blink reveals the probabilistic nature of discrete conscious perception. *Psychol. Sci.* **25**, 824–831 (2014).
10.1177/0956797613513810

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

28

E. Vul, D. Hanus, N. Kanwisher, Attention as inference: Selection is probabilistic; responses are all-or-none samples. *J. Exp. Psychol. Gen.* **138**, 546–560 (2009). 10.1037/a0017352

[CROSSREF](#)
[PUBMED](#)
[WEB OF SCIENCE](#)
[GOOGLE SCHOLAR](#)

29

R. Moreno-Bote, D. C. Knill, A. Pouget, Bayesian sampling in visual perception. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 12491–12496 (2011). 10.1073/pnas.1101430108

[CROSSREF](#)
[PUBMED](#)
[WEB OF SCIENCE](#)
[GOOGLE SCHOLAR](#)

30

E. Vul, M. Nieuwenstein, N. Kanwisher, Temporal selection is suppressed, delayed, and diffused during the attentional blink. *Psychol. Sci.* **19**, 55–61 (2008). 10.1111/j.1467-9280.2008.02046.x

[CROSSREF](#)
[PUBMED](#)
[WEB OF SCIENCE](#)
[GOOGLE SCHOLAR](#)

31

T. I. Panagiotaropoulos, G. Deco, V. Kapoor, N. K. Logothetis, Neuronal discharges and gamma oscillations explicitly reflect visual consciousness in the lateral prefrontal cortex. *Neuron* **74**, 924–935 (2012). 10.1016/j.neuron.2012.04.013

[CROSSREF](#)
[PUBMED](#)
[WEB OF SCIENCE](#)
[GOOGLE SCHOLAR](#)

32

N. K. Logothetis, Single units and conscious vision. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **353**, 1801–1818 (1998). 10.1098/rstb.1998.0333

[CROSSREF](#)
[PUBMED](#)
[WEB OF SCIENCE](#)
[GOOGLE SCHOLAR](#)

33

C. Sergent, S. Baillet, S. Dehaene, Timing of the brain events underlying access to consciousness during the attentional blink. *Nat. Neurosci.* **8**, 1391–1400 (2005). 10.1038/nn1549

[CROSSREF](#)
[PUBMED](#)
[WEB OF SCIENCE](#)
[GOOGLE SCHOLAR](#)

34

S. Marti, M. Sigman, S. Dehaene, A shared cortical bottleneck underlying attentional blink and psychological refractory period. *Neuroimage* **59**, 2883–2898 (2012). 10.1016/j.neuroimage.2011.09.063

[GO TO REFERENCE](#)

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

35

S. Marti, J.-R. King, S. Dehaene, Time-resolved decoding of two processing chains during dual-task interference. *Neuron* **88**, 1297–1307 (2015). 10.1016/j.neuron.2015.10.040

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

36

M. Nieuwenstein, E. Van der Burg, J. Theeuwes, B. Wyble, M. Potter, Temporal constraints on conscious vision: On the ubiquitous nature of the attentional blink. *J. Vis.* **9**, 1–14 (2009). 10.1167/9.9.18

[GO TO REFERENCE](#)

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

37

G. Kreiman, I. Fried, C. Koch, Single-neuron correlates of subjective vision in the human medial temporal lobe. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 8378–8383 (2002). 10.1073/pnas.072194099

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

38

R. Q. Quiroga, R. Mukamel, E. A. Isham, R. Malach, I. Fried, Human single-neuron responses at the threshold of conscious recognition. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 3599–3604 (2008). 10.1073/pnas.0707043105

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

39

J.-R. King, N. Pescetelli, S. Dehaene, Brain mechanisms underlying the brief maintenance of seen and unseen sensory information. *Neuron* **92**, 1122–1134 (2016). 10.1016/j.neuron.2016.10.051

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

40

D. Trübutschek, S. Marti, A. Ojeda, J.-R. King, Y. Mi, M. Tsodyks, S. Dehaene, A theory of working memory without consciousness or sustained activity. *eLife* **6**, e23871 (2017). 10.7554/eLife.23871

[GO TO REFERENCE](#)

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

41

S. Kouider, C. Stahlhut, S. V. Gelskov, L. S. Barbosa, M. Dutat, V. de Gardelle, A. Christophe, S. Dehaene, G. Dehaene-Lambertz, A neural marker of perceptual consciousness in infants. *Science* **340**, 376–380 (2013). 10.1126/science.1232509

[GO TO REFERENCE](#)

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

42

R. T. Maloney, J. Jayakumar, E. V. Levichkina, I. N. Pigarev, T. R. Vidyasagar, Information processing bottlenecks in macaque posterior parietal cortex: An attentional blink? *Exp. Brain Res.* **228**, 365–376 (2013). 10.1007/s00221-013-3569-2

[GO TO REFERENCE](#)

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

43

K. Watanabe, S. Funahashi, Neural mechanisms of dual-task interference and cognitive capacity limitation in the prefrontal cortex. *Nat. Neurosci.* **17**, 601–611 (2014). 10.1038/nn.3667

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

44

G. N. Elston, Cortex, cognition and the cell: New insights into the pyramidal neuron and prefrontal function. *Cereb. Cortex* **13**, 1124–1138 (2003). 10.1093/cercor/bhg093

[GO TO REFERENCE](#)

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

45

F.-X. Neubert, R. B. Mars, A. G. Thomas, J. Sallet, M. F. S. Rushworth, Comparison of human ventral frontal cortex areas for cognitive control and language with areas in monkey frontal cortex. *Neuron* **81**, 700–713 (2014). 10.1016/j.neuron.2013.11.012

[CROSSREF](#)
[PUBMED](#)
[WEB OF SCIENCE](#)
[GOOGLE SCHOLAR](#)

46

L. Wang, L. Uhrig, B. Jarraya, S. Dehaene, Representation of numerical and sequential patterns in macaque and human brains. *Curr. Biol.* **25**, 1966–1974 (2015).
10.1016/j.cub.2015.06.035

[GO TO REFERENCE](#)
[CROSSREF](#)
[PUBMED](#)
[WEB OF SCIENCE](#)
[GOOGLE SCHOLAR](#)

47

A. Cleeremans, B. Timmermans, A. Pasquali, Consciousness and metarepresentation: A computational sketch. *Neural Netw.* **20**, 1032–1039 (2007). 10.1016/j.neunet.2007.09.011

[CROSSREF](#)
[PUBMED](#)
[WEB OF SCIENCE](#)
[GOOGLE SCHOLAR](#)

48

A. Cleeremans, Connecting conscious and unconscious processing. *Cogn. Sci.* **38**, 1286–1315 (2014). 10.1111/cogs.12149

[GO TO REFERENCE](#)
[CROSSREF](#)
[PUBMED](#)
[WEB OF SCIENCE](#)
[GOOGLE SCHOLAR](#)

49

J. Dunlosky, J. Metcalfe, *Metacognition* (Sage Publications, 2008).

[GOOGLE SCHOLAR](#)

50

A. Clark, A. Karmiloff-Smith, The cognizer's innards: A psychological and philosophical perspective on the development of thought. *Mind Lang.* **8**, 487–519 (1993).
10.1111/j.1468-0017.1993.tb00299.x

[GO TO REFERENCE](#)
[CROSSREF](#)
[GOOGLE SCHOLAR](#)

51

F. Meyniel, D. Schlunegger, S. Dehaene, The sense of confidence during probabilistic learning: A normative account. *PLOS Comput. Biol.* **11**, e1004305 (2015).
10.1371/journal.pcbi.1004305

[CROSSREF](#)
[PUBMED](#)
[WEB OF SCIENCE](#)
[GOOGLE SCHOLAR](#)

52

F. Meyniel, S. Dehaene, Brain networks for confidence weighting and hierarchical inference during probabilistic learning. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E3859–E3868 (2017). 10.1073/pnas.1615773114

[GO TO REFERENCE](#)

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

53

J. D. Smith, The study of animal metacognition. *Trends Cogn. Sci.* **13**, 389–396 (2009). 10.1016/j.tics.2009.06.009

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

54

W. J. Ma, J. M. Beck, P. E. Latham, A. Pouget, Bayesian inference with probabilistic population codes. *Nat. Neurosci.* **9**, 1432–1438 (2006). 10.1038/nn1790

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

55

S. M. Fleming, R. S. Weil, Z. Nagy, R. J. Dolan, G. Rees, Relating introspective accuracy to individual differences in brain structure. *Science* **329**, 1541–1543 (2010). 10.1126/science.1191883

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

56

K. Miyamoto, T. Osada, R. Setsuie, M. Takeda, K. Tamura, Y. Adachi, Y. Miyashita, Causal neural network of metamemory for retrospection in primates. *Science* **355**, 188–193 (2017). 10.1126/science.aal0162

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

57

A. Kepecs, N. Uchida, H. A. Zariwala, Z. F. Mainen, Neural correlates, computation and behavioural impact of decision confidence. *Nature* **455**, 227–231 (2008). 10.1038/nature07200

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

58

E. Rounis, B. Maniscalco, J. C. Rothwell, R. E. Passingham, H. Lau, Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness. *Cogn. Neurosci.* **1**, 165–175 (2010). 10.1080/17588921003632529

[GO TO REFERENCE](#)

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

59

A. Resulaj, R. Kiani, D. M. Wolpert, M. N. Shadlen, Changes of mind in decision-making. *Nature* **461**, 263–266 (2009). 10.1038/nature08275

[GO TO REFERENCE](#)

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

60

L. Charles, J.-R. King, S. Dehaene, Decoding the dynamics of action, intention, and error detection for conscious and subliminal stimuli. *J. Neurosci.* **34**, 1158–1170 (2014). 10.1523/JNEUROSCI.2465-13.2014

[GO TO REFERENCE](#)

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

61

L. Charles, F. Van Opstal, S. Marti, S. Dehaene, Distinct brain mechanisms for conscious versus subliminal error detection. *Neuroimage* **73**, 80–94 (2013). 10.1016/j.neuroimage.2013.01.054

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

62

C. D. Frith, *The Cognitive Neuropsychology of Schizophrenia* (Psychology Press, 1992).

[GO TO REFERENCE](#)

[GOOGLE SCHOLAR](#)

63

J. S. Simons, J. R. Garrison, M. K. Johnson, Brain mechanisms of reality monitoring. *Trends Cogn. Sci.* **21**, 462–473 (2017). 10.1016/j.tics.2017.03.012

[GO TO REFERENCE](#)

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

64

D. Mendoza-Halliday, J. C. Martinez-Trujillo, *Nat. Commun.* 10.1038/ncomms15471 (2017).
10.1038/ncomms15471

[GO TO REFERENCE](#)

[CROSSREF](#)

[PUBMED](#)

[GOOGLE SCHOLAR](#)

65

L. Goupil, S. Kouider, Behavioral and neural indices of metacognitive sensitivity in preverbal infants. *Curr. Biol.* **26**, 3038–3045 (2016). 10.1016/j.cub.2016.09.004

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

66

L. Goupil, M. Romand-Monnier, S. Kouider, Infants ask for help when they know they don't know. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 3492–3496 (2016). 10.1073/pnas.1515129113

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

67

G. D. Logan, M. J. Crump, Cognitive illusions of authorship reveal hierarchical error detection in skilled typists. *Science* **330**, 683–686 (2010). 10.1126/science.1190483

[GO TO REFERENCE](#)

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

68

S. Nieuwenhuis, K. R. Ridderinkhof, J. Blom, G. P. Band, A. Kok, Error-related brain potentials are differentially related to awareness of response errors: Evidence from an antisaccade task. *Psychophysiology* **38**, 752–760 (2001). 10.1111/1469-8986.3850752

[GO TO REFERENCE](#)

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

69

B. Bahrami, K. Olsen, P. E. Latham, A. Roepstorff, G. Rees, C. D. Frith, Optimally interacting minds. *Science* **329**, 1081–1085 (2010). 10.1126/science.1185718

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

70

R. S. Siegler, Strategy choice procedures and the development of multiplication skill. *J. Exp. Psychol. Gen.* **117**, 258–275 (1988). 10.1037/0096-3445.117.3.258

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

71

V. de Gardelle, P. Mamassian, Does confidence use a common currency across two visual tasks? *Psychol. Sci.* **25**, 1286–1288 (2014). 10.1177/0956797614528956

[GO TO REFERENCE](#)

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

72

C. Fernando, D. Banarse, C. Blundell, Y. Zwols, D. Ha, A. A. Rusu, A. Pritzel, D. Wierstra, arXiv:170108734 [cs.NE] (2017).

[GOOGLE SCHOLAR](#)

73

J. B. Tenenbaum, C. Kemp, T. L. Griffiths, N. D. Goodman, How to grow a mind: Statistics, structure, and abstraction. *Science* **331**, 1279–1285 (2011). 10.1126/science.1192788

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

74

J. Gottlieb, P.-Y. Oudeyer, M. Lopes, A. Baranes, Information-seeking, curiosity, and attention: Computational and neural mechanisms. *Trends Cogn. Sci.* **17**, 585–593 (2013). 10.1016/j.tics.2013.09.001

[GO TO REFERENCE](#)

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

75

I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative Adversarial Networks. arXiv:14062661 [stat.ML] (2014).

[GO TO REFERENCE](#)

[GOOGLE SCHOLAR](#)

76

G. Tononi, M. Boly, M. Massimini, C. Koch, Integrated information theory: From consciousness to its physical substrate. *Nat. Rev. Neurosci.* **17**, 450–461 (2016). 10.1038/nrn.2016.44

[GO TO REFERENCE](#)

[CROSSREF](#)

[PUBMED](#)
[WEB OF SCIENCE](#)
[GOOGLE SCHOLAR](#)

77

S. Kouider, E. Eger, R. Dolan, R. N. Henson, Activity in face-responsive brain regions is modulated by invisible, attended faces: Evidence from masked priming. *Cereb. Cortex* **19**, 13–23 (2009). 10.1093/cercor/bhn048

[GO TO REFERENCE](#)
[CROSSREF](#)
[PUBMED](#)
[WEB OF SCIENCE](#)
[GOOGLE SCHOLAR](#)

78

S. Dehaene, L. Naccache, L. Cohen, D. L. Bihan, J.-F. Mangin, J.-B. Poline, D. Rivière, Cerebral mechanisms of word masking and unconscious repetition priming. *Nat. Neurosci.* **4**, 752–758 (2001). 10.1038/89551

[GO TO REFERENCE](#)
[CROSSREF](#)
[PUBMED](#)
[WEB OF SCIENCE](#)
[GOOGLE SCHOLAR](#)

79

P. Vuilleumier, N. Sagiv, E. Hazeltine, R. A. Poldrack, D. Swick, R. D. Rafal, J. D. E. Gabrieli, Neural fate of seen and unseen faces in visuospatial neglect: A combined event-related functional MRI and event-related potential study. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 3495–3500 (2001). 10.1073/pnas.051436898

[GO TO REFERENCE](#)
[CROSSREF](#)
[PUBMED](#)
[WEB OF SCIENCE](#)
[GOOGLE SCHOLAR](#)

80

A. Kiesel, W. Kunde, C. Pohl, M. P. Berner, J. Hoffmann, Playing chess unconsciously. *J. Exp. Psychol. Learn. Mem. Cogn.* **35**, 292–298 (2009). 10.1037/a0014499

[GO TO REFERENCE](#)
[CROSSREF](#)
[PUBMED](#)
[WEB OF SCIENCE](#)
[GOOGLE SCHOLAR](#)

81

M. Aly, A. P. Yonelinas, Bridging consciousness and cognition in memory and perception: Evidence for both state and strength processes. *PLOS ONE* **7**, e30231 (2012). 10.1371/journal.pone.0030231

[GO TO REFERENCE](#)
[CROSSREF](#)
[PUBMED](#)

[WEB OF SCIENCE](#)
[GOOGLE SCHOLAR](#)

82

I. M. Harlow, A. P. Yonelinas, Distinguishing between the success and precision of recollection. *Memory* **24**, 114–127 (2016). 10.1080/09658211.2014.988162

[GO TO REFERENCE](#)
[CROSSREF](#)
[PUBMED](#)
[WEB OF SCIENCE](#)
[GOOGLE SCHOLAR](#)

83

H. L. Pincham, H. Bowman, D. Szucs, The experiential blink: Mapping the cost of working memory encoding onto conscious perception in the attentional blink. *Cortex* **81**, 35–49 (2016). 10.1016/j.cortex.2016.04.007

[GO TO REFERENCE](#)
[CROSSREF](#)
[PUBMED](#)
[WEB OF SCIENCE](#)
[GOOGLE SCHOLAR](#)

84

C. Sergent, S. Dehaene, Is consciousness a gradual phenomenon? Evidence for an all-or-none bifurcation during the attentional blink. *Psychol. Sci.* **15**, 720–728 (2004). 10.1111/j.0956-7976.2004.00748.x

[GO TO REFERENCE](#)
[CROSSREF](#)
[PUBMED](#)
[WEB OF SCIENCE](#)
[GOOGLE SCHOLAR](#)

85

A. Del Cul, S. Baillet, S. Dehaene, Brain dynamics underlying the nonlinear threshold for access to consciousness. *PLOS Biol.* **5**, e260 (2007). 10.1371/journal.pbio.0050260

[GO TO REFERENCE](#)
[CROSSREF](#)
[PUBMED](#)
[WEB OF SCIENCE](#)
[GOOGLE SCHOLAR](#)

86

R. Marois, D. J. Yi, M. M. Chun, The neural fate of consciously perceived and missed events in the attentional blink. *Neuron* **41**, 465–472 (2004). 10.1016/S0896-6273(04)00012-1

[CROSSREF](#)
[PUBMED](#)
[WEB OF SCIENCE](#)
[GOOGLE SCHOLAR](#)

87

C. Moutard, S. Dehaene, R. Malach, Spontaneous fluctuations and non-linear ignitions: Two dynamic faces of cortical recurrent loops. *Neuron* **88**, 194–206 (2015).
10.1016/j.neuron.2015.09.018

[GO TO REFERENCE](#)

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

88

H. G. Rey, I. Fried, R. Quiñan Quiroga, Timing of single-neuron and local field potential responses in the human medial temporal lobe. *Curr. Biol.* **24**, 299–304 (2014).
10.1016/j.cub.2013.12.004

[GO TO REFERENCE](#)

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

89

A. Schurger, I. Sarigiannidis, L. Naccache, J. D. Sitt, S. Dehaene, Cortical activity is more stable when sensory stimuli are consciously perceived. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E2083–E2092 (2015). 10.1073/pnas.1418730112

[GO TO REFERENCE](#)

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

90

C. Sergent, V. Wyart, M. Babo-Rebelo, L. Cohen, L. Naccache, C. Tallon-Baudry, Cueing attention after the stimulus is gone can retrospectively trigger conscious perception. *Curr. Biol.* **23**, 150–155 (2013). 10.1016/j.cub.2012.11.047

[GO TO REFERENCE](#)

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

91

J. Sackur, S. Dehaene, The cognitive architecture for chaining of two mental operations. *Cognition* **111**, 187–211 (2009). 10.1016/j.cognition.2009.01.010

[CROSSREF](#)

[PUBMED](#)

[WEB OF SCIENCE](#)

[GOOGLE SCHOLAR](#)

92

R. Marois, J. Ivanoff, Capacity limits of information processing in the brain. *Trends Cogn. Sci.* **9**, 296–305 (2005). 10.1016/j.tics.2005.04.010

[GO TO REFERENCE](#)

[CROSSREF](#)
[PUBMED](#)
[WEB OF SCIENCE](#)
[GOOGLE SCHOLAR](#)

93

P. Barttfeld, L. Uhrig, J. D. Sitt, M. Sigman, B. Jarraya, S. Dehaene, Signature of consciousness in the dynamics of resting-state brain activity. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 887–892 (2015). 10.1073/pnas.1418031112

[GO TO REFERENCE](#)
[CROSSREF](#)
[PUBMED](#)
[WEB OF SCIENCE](#)
[GOOGLE SCHOLAR](#)

94

W. J. Gehring, B. Goss, M. G. H. Coles, D. E. Meyer, E. Donchin, A neural system for error detection and compensation. *Psychol. Sci.* **4**, 385–390 (1993). 10.1111/j.1467-9280.1993.tb00586.x

[GO TO REFERENCE](#)
[CROSSREF](#)
[WEB OF SCIENCE](#)
[GOOGLE SCHOLAR](#)

Theoretical Models of Consciousness: A Scoping Review

[Davide Sattin](#)^{1,2,*}, [Francesca Giulia Magnani](#)¹, [Laura Bartesaghi](#)¹, [Milena Caputo](#)¹, [Andrea Veronica Fittipaldo](#)³, [Martina Cacciatore](#)¹, [Mario Picozzi](#)⁴, [Matilde Leonardi](#)¹

Editor: Gerry Leisman

- Author information
- Article notes
- Copyright and License information

PMCID: PMC8146510 PMID: [33923218](#)

Abstract

The amount of knowledge on human consciousness has created a multitude of viewpoints and it is difficult to compare and synthesize all the recent scientific perspectives. Indeed, there are many definitions of consciousness and multiple approaches to study the neural correlates of consciousness (NCC). Therefore, the main aim of this article is to collect data on the various theories of consciousness published between 2007–2017 and to synthesize them to provide a general overview of this topic. To describe each theory, we developed a thematic grid called the dimensional model, which qualitatively and quantitatively analyzes how each article, related to one specific theory, debates/analyzes a specific issue. Among the 1130 articles assessed, 85 full texts were included in the prefinal step. Finally, this scoping review analyzed 68 articles that described 29 theories of consciousness. We found heterogeneous perspectives in the theories analyzed. Those with the highest grade of variability are as follows: subjectivity, NCC, and the consciousness/cognitive function. Among sub-cortical structures, thalamus, basal ganglia, and the hippocampus were the most indicated, whereas the cingulate, prefrontal, and temporal areas were the most reported for cortical ones also including the thalamo-cortical system. Moreover, we found several definitions of consciousness and 21 new sub-classifications.

Keywords: consciousness, neural correlates, cognition, consciousness theory, consciousness definition, subjectivity

1. Introduction

The scientific literature debating themes on consciousness is huge. Contributing scientists come from different disciplines such as medicine, philosophy, physics, psychology, neurobiology, mathematics, and computer science. Authors from these disciplines have led to an increase in knowledge about consciousness, giving new and interesting inputs for neuroscience. However, all this information has created a multitude of viewpoints and it is difficult to compare (and synthesize) some of the various issues that have emerged in the last few decades. For example, one of the main problems related to consciousness is its definition. Indeed, as remarked by Sommerhoff [1], “A precise definition of the word [‘consciousness’], of course, can only be the endpoint of a theory of consciousness, just as the concepts of work and energy found a

precise definition only as part of a theory of mechanics” [1]. Therefore, the probability of having different definitions of the same object correlates directly to the number of theories that try to explain the nature of consciousness itself. In this sense, the concept of ‘consciousness’ can be found in historical documents since ancient times, and it can be proven in various traditions and cultures with some perspectives that, although the expression language has become different, seem to be concepts that had evolved throughout history [2]. Fritjof Capra [3] describes how some eastern ancient concepts on a universal consciousness are consistent with modern physics, considering that some physics perspectives try to invalidate the concept of atomic particles as consisting of smaller independent building blocks made of different substances. Capra believes that one substance exists that comes from a ‘cosmic consciousness’ (this was the definition of consciousness used) out of which all matter is made of. This is a position commonly used also by those who argue that consciousness supervenes on the whole embodied animal in dynamic interaction with the environment [4]. Another example proposed in the last decade derived by some eastern reductionist perspective highlighted that consciousness could be defined as “the ability to maintain an alert state, attention, and awareness of self and environment” [5] or as the “the subjective character of our mental states” [6], according to the clinical and representational perspective used to describe it. This provides particular attention to the role of introspection and metarepresentation, two themes cited by Aristotle that contribute to current debates over the nature of consciousness [7].

Therefore, the analysis of the general framework and the perspectives proposed by each model of consciousness is important for improving the science of consciousness and, consequently, its definitions.

Another issue is represented by the definition of the neural correlates of consciousness (NCCs). If we consider consciousness as a human phenomenon deriving from matter, one of the main neurobiological questions is the definition of the minimum neuronal mechanisms jointly sufficient for any one specific conscious percept [8,9]. The definition of what we mean by consciousness and the brain neurons/areas/processes linked to the appearance of this phenomenon is a crucial issue for all theories of consciousness. The evolution of imaging and neurophysiological techniques, as well as the possibility to change brain activities online using tailored magnetic and electrical stimulations during tasks requiring wakefulness and awareness of external stimuli, has led to increases in the possibility of verifying/falsifying different hypotheses. Moreover, it has led to study the of consciousness modulation through the stimulation of different brain areas, overpassing the limits derived from the anatomo-clinical model based on the lesional approach. However, every theoretical approach has some assumptions according to which consciousness could be generated by a multitude of processes. This is grounded in the activity of a wide range of structures that vary from the single neuron to the whole brain. Consequently, the number of NCCs is presently very large.

There are other main issues debated in the scientific literature related to consciousness. The problem of how we can “quantify” the conscious level, for example, is one of the

most important questions that many scientists are still studying. This question is crucial for clinicians during the clinical diagnosis of patients with disorders of consciousness after a severe brain injury (to assess the consciousness level beyond behavioral abilities of patients), as well as for computer science/artificial intelligence researchers involved in machine and computer development.

All the examples reported above makeup only a minimal part of the literature on consciousness, which has increased over the last few years. Therefore, the general aim of this article is to offer an exhaustive framework on the theoretical models of consciousness that currently exist, without committing to a specific perspective or viewpoint. To achieve this goal, the specific aims of this research are twofold: Firstly, we sought to systematically review the scientific literature on theoretical models of consciousness published from 2007 to 2017 in major scientific databases and, secondly, to describe each theory collected using a common thematic grid called the dimensional model (DM) in order to qualitatively and quantitatively analyze how each article, related to a specific theory, debates/analyzes specific issues.

2. Materials and Methods

We included all the published studies that analyzed theoretical aspects of consciousness.

2.1. Search Strategies and Information Sources

The authors performed a systematic search in the Medline (PubMed), Embase (EMBASE.com), Scopus, Web of Science (WoS), and PsycINFO databases. A tailored search strategy was developed for each database according to their thesaurus characteristics. The search method for the identification of studies is reported in [Supplemental Materials \(Table S1\)](#). Results obtained from each database were exported in a web-based bibliographic management software (RefWorks; <https://www.refworks.com/it/>, accessed on May–December 2017) and duplicate deletion was done by two independent raters, which were then matched. Moreover, all records were imported in a tailored Excel spreadsheet including title, abstract, and record information for each article.

2.2. General Selection Criteria

Studies were excluded if they met any of the following criteria: (1) studies published before 2007; (2) article type: book, generic, magazine article, or monograph; (3) studies not published in the English or Italian languages.

This review was performed following the preferred reporting items for systematic reviews and meta-analyses criteria [\[10\]](#).

2.3. Screening and Eligibility Criteria for Each Step

Search results were independently screened by different reviewers using a three-step procedure due to the high number of articles selected. Duplicate articles were removed before step 1.

2.3.1. Step 1: Abstract Selection

In the first step, one author (Rater #1) screened the articles by title, abstract, and keywords using the following assessment scale: 0 = the article does not debate consciousness itself (the term consciousness appears in the title or the abstract but with general meaning (e.g., people are conscious of their life, etc.)); 1 = the rater doubts this article; 2 = the article debates some aspects related to consciousness. In total, 20% (4100 studies) of the total number of articles were selected with random data extraction (using SPSS software; SPSS Inc., Chicago, IL, USA) for an independent rating made by Rater #2. The agreement between the two raters (Cohen K analysis) was predetermined; it had to be over 0.90 to proceed with the second step of the screening process. If the agreement value was below 0.90, a third rater was involved in the screening of abstracts, and then a second analysis was performed. The third author (Rater #3), who had senior experience in systematic reviews and consciousness issues, checked the quality of data collection, helping to solve the discrepancy between the evaluations of the first- and second-rater.

2.3.2. Step 2: Full-Text Selection

In the second step, two raters (#2 and #3) analyzed all full-text articles extracted during step 1 to assess them for final eligibility. The assessment scale used in this step was as follows: 0 = the article does not report theoretical aspects related to consciousness; 1 = the rater doubts this article; 2 = the article discusses theoretical aspects related to consciousness. The term “theoretical aspects” was used in referring to papers that accomplished three things: (a) describe a theory of consciousness; (b) describe one dimension of the DM (see below for its description) concerning consciousness as a topic; (c) describe a specific issue concerning different models of consciousness. In the case of a disagreement between raters regarding a selected study, consensus was achieved by means of discussion involving Rater #4, who read the article/s and participated in the discussion.

2.3.3. Step 3: Data Extraction and Management

In the final step, full-text analysis of the selected records classified as “included” in the previous step was performed. Raters #2, #4, and #5 independently analyzed the selected articles. Each of them completed the DM (see below for the full description) and then summarized each theory of consciousness. Additional relevant papers were retrieved through a cross-reference check. The bibliography of each article was read to collect information on other studies not included in the final list.

2.4. Outcome Measures

Before beginning the review process, we developed a DM. The DM is a thematic grid used to quantify how each article, belonging to one specific theory, debates/analyzes specific issues. Dimensions are selected according to (i) important questions that emerged from previous consciousness studies [[11](#),[12](#),[13](#),[14](#),[15](#)] and (ii) questions on which authors wanted to point out.

2.5. Main Dimensions Analyzed

2.5.1. Neural Correlates of Consciousness (NCCs)

NCCs are defined as the minimum neuronal mechanisms jointly sufficient for any one specific conscious percept [[8](#),[9](#),[16](#)]. In each article, the rater analyzed how much NCC was debated (quantifying this dimension using the Likert scale described below). Further, the rater reported the DM in the main brain structures. Moreover, the raters indicated whether the NCCs were described at the neuronal or brain area levels and whether the NCCs were correlated to a process involving different neural circuits. In this review, we did not consider the distinction between content-related consciousness NCC (i.e., the local state) and awakening state NCC (i.e., the global state or the activity which determines a particular phenomenal distinction within an experience) as eligibility criteria. Rather, raters reported more specific information when a clear description was found in the analyzed articles.

2.5.2. Association between Consciousness and Other Cognitive Functions

In this dimension, raters were asked to analyze how much each article explained the relationships between consciousness and other cognitive functions such as attention, memory, mental imagery, and so on. In each study, raters were asked to analyze how much the theoretical process linked conscious information to cognitive functions (specifically about information processing). For instance, raters were required to analyze how stimuli conscious elaboration is related to memory or how consciousness generation mechanisms work in relation to attentional functions [[17](#)].

2.5.3. Translation from Theory to Clinical Practice

In this area, we spotlighted information useful for the clinical management of patients with neurological/psychiatric disorders related to consciousness manifestations. Raters evaluated if each article provided information about clinical implications for either the diagnosis or the treatments of patients (e.g., with disorders of consciousness or with loss of consciousness due to epilepsy, etc.). Moreover, raters evaluated if the article reported any implications that would determine some prognostic markers based on the theory described. Specifically, the translation from theoretical aspects to clinical practice was examined, looking for words such as “marker”, “recovery of consciousness”, “diagnosis”, “clinical impact”, and “clinical implications”, as is reported in the [supplementary materials](#).

2.5.4. Quantitative Measures of Consciousness

Raters observed if an article reported information about the quantification of consciousness levels. We analyzed if instruments/tools to measure consciousness were described and if tailored algorithms were reported. Special attention was paid to the “index” definition directly related to the consciousness definition reported in the theory.

2.5.5. Consciousness, Sensory Processes, and the Autonomic Nervous System

This area was chosen to explore if the selected theories of consciousness debate on the relationships between peripheral information derived from the autonomic nervous system (ANS) (also including sensory processes) and consciousness. In detail, raters were asked to quantify how much the elaboration of information coming from ANS was described in each article and how this information was linked to a consciousness percept (e.g., pathways that provide visceral sensation to conscious perception).

2.5.6. Subjectivity

The final dimension analyzed by raters refers to how each article/theory of consciousness explains where subjectivity comes from and how it is related to conscious experience. A specific definition of subjectivity was not provided to raters to avoid the exclusion of new concepts and alternative viewpoints about subjectivity. In this sense, raters were instructed to search within for text-specific words such as “subjective experience”, “subjectivity”, “first-person experience”, and other terms linked to the discussion about how subjective experience could emerge from neural activity, and then to analyze how much this topic is analyzed in the text of each article.

2.6. Outcomes

2.6.1. Quantitative Outcomes

For each study collected, we analyzed the full text, giving a score on a Likert scale ranging from 0 to 5. The score for each dimension was assigned according to the following criteria:

0 = the issue was not debated in the article.

1 = the article gave minimal attention to the issue. The theme was cited in the text (introduction or discussion) as a marginal topic that used fewer words.

2 = the issue was not reported in the abstract. The theme was debated in the text using some sentences as collateral/secondary issues in refers to other main topics.

3 = the issue was not reported in the abstract but there was a short paragraph that discussed the issue reporting specific references.

4 = the issue was cited in the abstract. In specific sections of the paper (introduction, results, or discussion) there was at least a long paragraph on the issue reporting specific references.

5 = the issue was cited in the title and/or in the abstract and/or as a keyword. The issue was the main topic of the article, which was debated in different parts of the text.

2.6.2. Qualitative Outcomes

A series of information, although not evaluable in a quantitative way, was also collected from full texts.

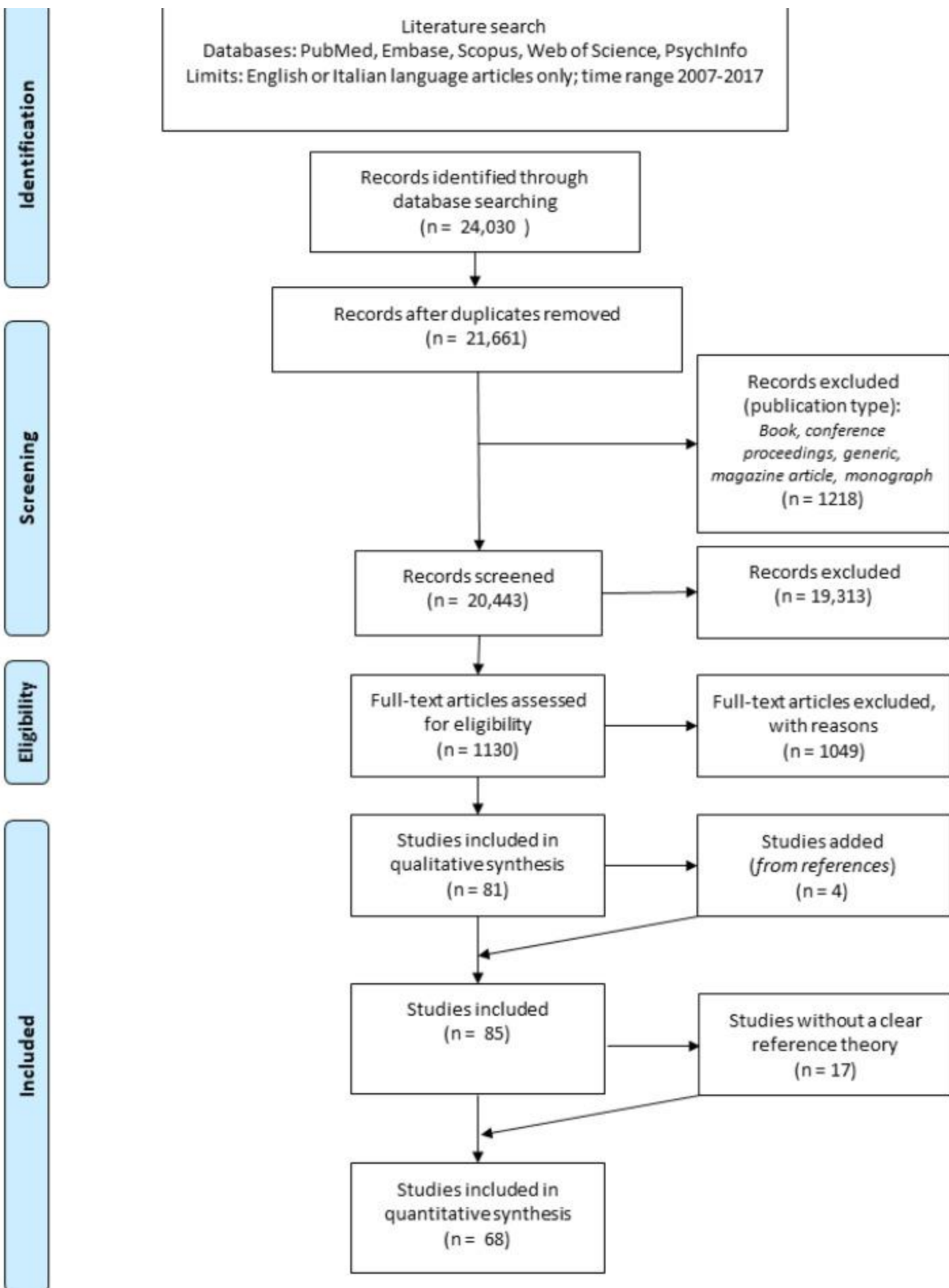
1. Main definitions of consciousness. Raters were asked to report if there was a main definition of consciousness (e.g., consciousness is/represents/serves/refers to/consists of/results in/defined as/has to do with) linked to the main theory described in the articles they read. The definition of consciousness expressed by other authors and only cited in the text was not reported unless strictly related to the theory presented in the article. Definitions of what is not consciousness were also included.
2. Definitions of consciousness' components/parts/sub-elements. Raters noted any definitions concerning parts and elements of consciousness and their definitions (if available).
Related terms/features. Raters reported specific terms and/or features concerning the nature of consciousness as described by the single article/theory that is complementary to the main definition.

3. Results

3.1. Review Results

The search identified 21,661 records after duplicates deletion performed in the step. In total, 11130 full-text articles were analyzed in step 2. Further, 68 articles were analyzed using the DM (see [Figure 1](#)).

Figure 1.

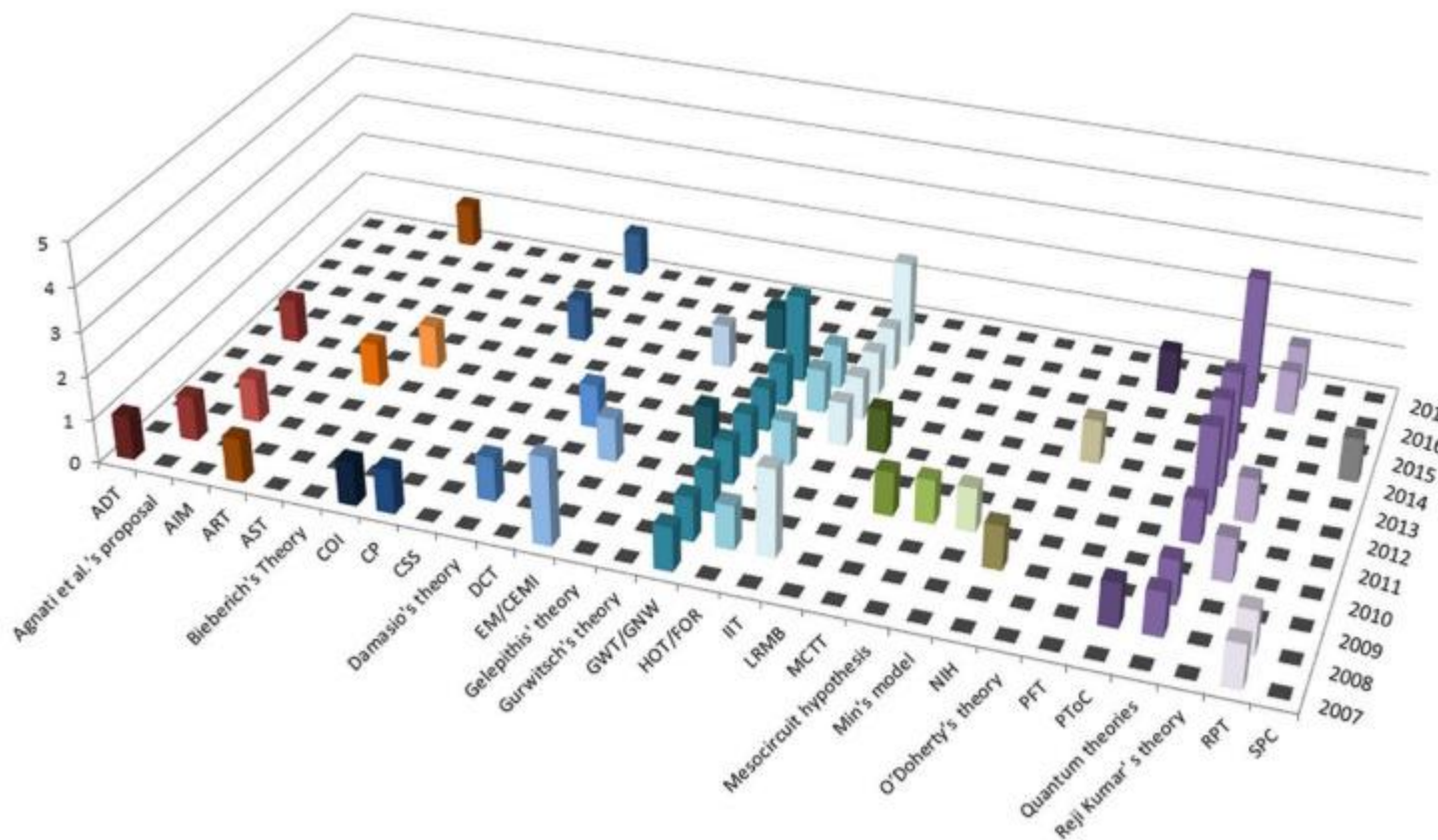


Prisma flow diagram.

The criteria used to exclude records before the final analysis consisted of using the term “consciousness” with a generic meaning (e.g., “conscious about their condition”) or with a moral viewpoint (e.g., “conscious about the consequences of his actions”).

During the analysis of the final 68 articles, we identified 29 different theories of consciousness ([Figure 2](#)). Each of them was synthesized following alphabetical order. [Figure 2](#) graphically represents the number of record distribution over the years.

Figure 2.

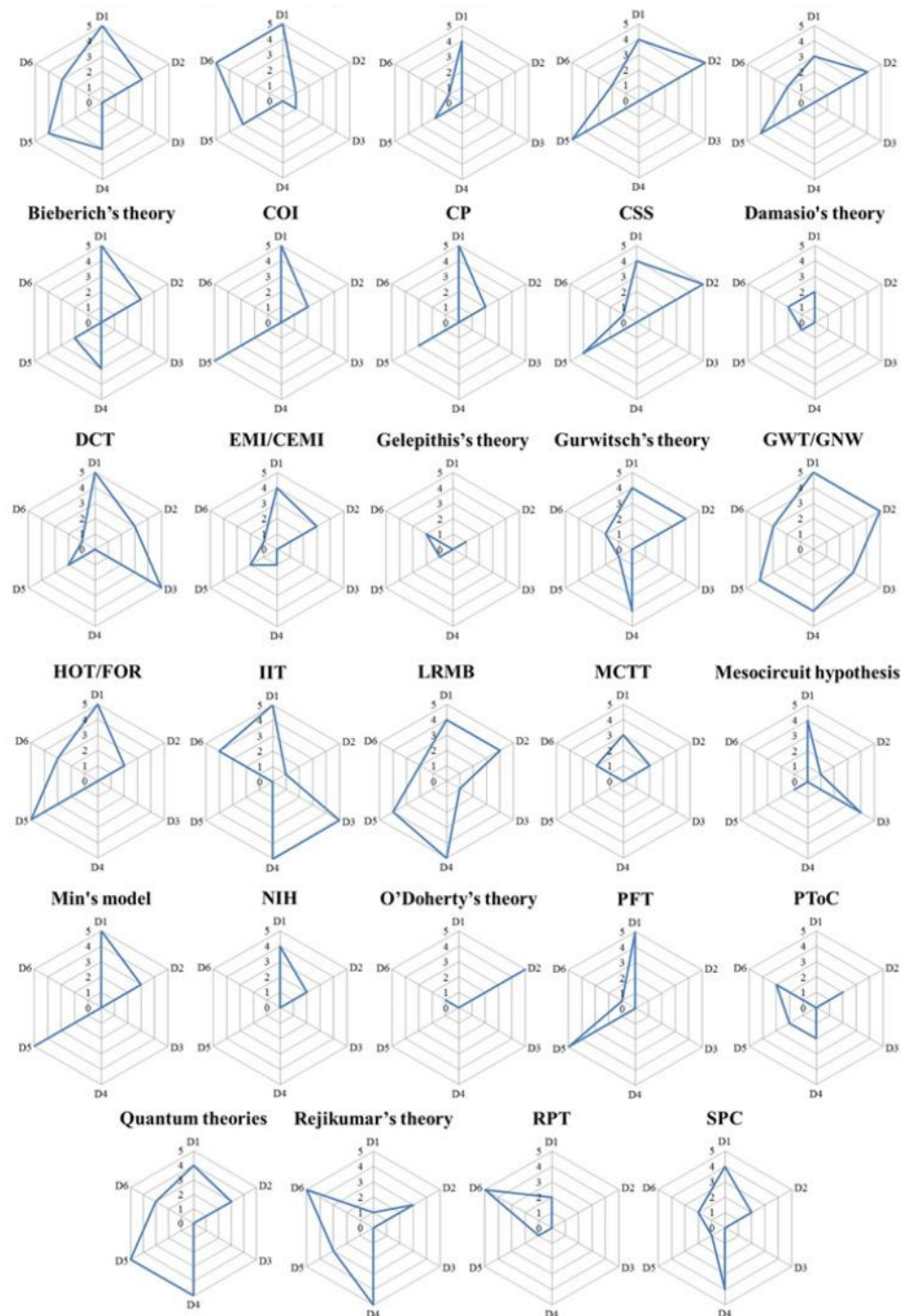


Graphic representation of the number of articles found for each theory in the period analyzed. ADT: apical dendrite theory; AIM: activation/information/mode-synthesis hypothesis; ART: adaptive resonance theory; AST: attention schema theory; COI: cross-order integration theory of consciousness; CP: centrencephalic proposal; CCS: consciousness state space model; DCT: dynamic core theory; EM/CEMI: electromagnetic field hypothesis/consciousness electromagnetic Information field theory; GWT: global workspace theory; GNW: global neural workspace; HOT/FOR: higher order theories of consciousness/first order representational theory; IIT: integrated

information theory; LRMB: layered reference model of the brain; MCTT: memory consciousness and temporality theory; NIH: network inhibition hypothesis; PFT: passive frame theory; PToC: psychological theory of consciousness. Q theories—Orch OR: orchestrated objective reduction theory; single particle consciousness hypothesis; quantum no-go theorems; the three layer model; Koehler's mathematical approach; timeless and spaceless; the four-dimensional Einstein; RPT: radical plasticity thesis; SPC: semantic pointer competition theory of consciousness; Agnati et al.; Bieberich's; Damasio; Gelepithis's; Gurwitsch's; Min's; O'Doherty's; Rejikumar's. The Mesocircuit hypothesis is not abbreviated.

[Figure 3](#) reports the maximum score obtained by the articles analyzed for each theory. In in dimension number 1 (NCC), the maximum score was observed in 11 theories; in D2 (Association between consciousness and other cognitive functions) in three theories; in D3 (translation from theory to clinical practice) in two theories; in D4 (quantitative measures of consciousness) in four theories; in D5 (consciousness, sensory processes, and the autonomic nervous system) in six theories; and in D6 (subjectivity) in two theories. Regarding the total number of dimensions debated by each theory, we found that the range was comprised of a minimum of two dimensions debated in the selected articles (NIH and O'Doherty's theory) to a maximum of all dimensions, as reported in [Table S2](#). The theories with the highest number of analyzed articles were the group of the quantum theories of consciousness ($n = 12$), the IIT ($n = 8$), and the GWT/GNW ($n = 9$), whereas 18 theories were described by a single article.

Figure 3.



[Open in a new tab](#)

Graphic representation of the maximum scores obtained in the six dimensions analyzed by each theory. The figures depicted the scores achieved by each theory (displayed in alphabetical order) in each dimension of the dimensional model. We considered the maximum score for each theory among the articles belonging to it. The y-axis represents the scores ranging from 0 to 5. **D1** refers to the neural correlates of consciousness dimension. **D2** refers to the association between consciousness and other cognitive functions dimension. **D3** refers to translation from theory to clinical practice dimension. **D4** refers to the quantitative measures of consciousness

dimension. **D5** refers to consciousness, sensory processes, and the autonomic nervous system dimension. **D6** refers to the subjectivity dimension.

For each analyzed article, we reported the definition of consciousness ([Table 1](#)). Regarding the various definitions found, we decided to report the exact sentences found in each article to avoid misinterpretation. Frequencies of the main terms used in the main definitions of consciousness by each theory were also reported in [Table 2](#).

Table 1.

Consciousness definitions found in the collected articles.

Theory	Authors (year)	Main Definition	Other Terms/Subcategorization Related to Consciousness (Definitions Reported If Available)
ADT	LaBerge and Kasevich (2007) [18]	Consciousness is an activity that is extended in time and typically continues from the time we awake to the time we fall asleep.	- Background consciousness: some processes in the brain create “cognitive event that may be called ‘having an impression’ of something”. - Elevated consciousness for selected aspects of background consciousness is assumed to arise when sustained activity of a primary sensory area is sent to higher sensory areas, and where a selected part of the sensory scene is amplified by attentional activity controlled from the frontal lobes. - Foreground consciousness: the elevated attentional activity of a part of the sensory scene in higher sensory areas. - Content of consciousness: the simultaneous impressions from both foreground and background consciousness together;
Agnati et al.’s proposal	Cook (2008) [19]	nf	
	Agnati et al. (2012) [20]	Consciousness may be thought as the global result of integrative processes taking place at different levels of miniaturization in plastic mosaics	
AIM	Hobson (2009) [21]	Waking consciousness can be defined as the awareness of the external world, our bodies, and ourselves (including the	- Primary consciousness can be defined as simple awareness that includes perception and emotion. - Secondary consciousness depends on

Theory	Authors (year)	Main Definition	Other Terms/Subcategorization Related to Consciousness (Definitions Reported If Available)
		awareness of our awareness) that humans experience when awake.	language and includes such features as self-reflective awareness, abstract thinking, volition, and metacognition.
ART	Grossberg (2007) [22]	nf	
	Grossberg (2017) [23]	Consciousness is not just a whirl of information-processing.	- Core consciousness (see Damasio's theory)
AST	Graziano and Kastner (2011) [24]	Consciousness is not an emergent property or a metaphysical emanation but is itself information computed by an expert system. [...] Consciousness=awareness (i.e., a perceptual model of attention).	
Bieberich's theory	Bieberich (2012) [25]	Consciousness determines what is perceived as reality.	
COI	Kriegel (2007) [26]	When a subject has a higher-order representation that is unified with the first-order representation it represents, their representational unity constitutes a conscious state. Consciousness determines what is perceived as reality.	
CP	Merker (2007) [27]	Consciousness may be regarded most simply as the "medium" of all possible experience. Consciousness as the state or condition presupposed by any experience whatsoever.	- Reflective consciousness: is one of many contents of consciousness available to creatures with sophisticated cognitive capacities; it is "a luxury of consciousness on the part of certain big-brained species, and not its defining property"; self-consciousness, unselfconsciously.
CSS	Berkovich-Ohana et al. (2014) [28]	Consciousness is an experienced property of mental states and processes, which is lost during a dreamless deep sleep, deep anesthesia, or coma	- Minimal self (MS): a self that is short of temporal extension and endowed with a sense of agency, ownership, and non-conceptual first-person content. - Narrative self (NS): it involves personal identity and continuity across time, as well as conceptual thought. - Core consciousness (CC): it supports the MS; its scope is in the here and

Theory	Authors (year)	Main Definition	Other Terms/Subcategorization Related to Consciousness (Definitions Reported If Available)
			now. - Extended consciousness (EC): it supports the NS and involves memory of past, imagination of future, and verbal thought. It deals with holding in mind, overtime, a multiplicity of neural patterns that describe the autobiographical self.
	Berkovich-Ohana et al. (2017) [29]	nf	
Damasio's theory	Bosse et al. (2008) [30]	Proto-self: The proto-self is a coherent collection of neural patterns which map, moment by moment, the state of the physical structure of the organism in its many dimensions. Core consciousness (or feeling a feeling) is what emerges when the organism detects that its representation of its own body state (the proto-self) has been changed by the occurrence of the stimulus. Thus, it becomes (consciously) aware of the feeling, i.e., extended consciousness;	
DCT	Ward (2011) [31]	Phenomenal consciousness is generated by synchronized neural activity in the dendritic trees of dorsal thalamic neurons: a thalamic dynamic core.	Meta-consciousness and Self-consciousness are viewed as contents of consciousness that are built from recursive application of primary consciousness to conscious contents.
EMI/CEMI	McFadden (2007) [32]	Consciousness is that component of the brain's electromagnetic information field that is downloaded to motor neurons and is thereby capable of communicating its state to the outside world. Consciousness is what electromagnetic field information feels like from the inside.	
	Pockett (2007) [33]	nf	

Theory	Authors (year)	Main Definition	Other Terms/Subcategorization Related to Consciousness (Definitions Reported If Available)
	Lewis and MacGregor (2010) [34]	Consciousness itself is a dispositional form of energy that may relate to physical forms of energy as the phase state of a gas relates in a material way to its liquid phase.	
Gelepithis' theory	Gelepithis (2014) [35]	nf	
Gurwitsch's theory	Yoshimi (2011) [36]	nf	
	Yoshimi and Vinson (2015) [37]	Consciousness is the totality of co-present data [...] Any field of consciousness can be parsed into these three co-present domains of data: theme, thematic field, and margin.	Marginal consciousness: unattended data not relevant to the theme. Theme: data at the focus of attention organized according to Gestalt law. Thematic field: unattended data relevant to the theme.
GWT/GNW	Baars and Franklin (2007) [38]	Consciousness serves as a lookout function to spot potential dangers or opportunities so that there is a particularly close relationship between conscious content and significant sensory input.	
	Prakash et al. (2008) [39]	nf	
	Baars and Franklin (2009) [40]	nf	
	Raffone and Pantani (2010) [41]	nf	Phenomenal consciousness consists in phenomenally conscious states that are not cognitively accessible. Access consciousness represents a content information which is 'broadcast' in the GW.
	Dehaene and Changeux (2011) [42]	The availability of information is what we subjectively experience in a conscious state. "Conscious" is an ambiguous word. In its intransitive use (e.g., "the patient was still conscious") it	

Theory	Authors (year)	Main Definition	Other Terms/Subcategorization Related to Consciousness (Definitions Reported If Available)
		refers to the state of consciousness, also called wakefulness or vigilance, which is thought to vary continuously from coma and slow-wave sleep to full vigilance. In its transitive use (e.g., “I was not conscious of the red light”), it refers to conscious access to and/or conscious processing of a specific piece of information.	
	Sergent and Naccache (2012) [43]	It is possible to use subjective reports to probe the content of consciousness and therefore define any representation which is not reported by the subject as non-conscious, even when questioned about it, the existence of which can be demonstrated through behavioral and/or functional brain-imaging measures.	Phenomenal consciousness is a much larger domain of conscious contents than the one accessible through reports perceptual consciousness; micro-consciousness
	Baars et al. (2013) [44]	Conscious experiences reflect a flexible “binding and broadcasting” function in the brain, which mobilize a large, distributed collection of specialized cortical networks and processes that are not conscious by themselves.	Sensory consciousness; perceptual consciousness.
	Dehaene et al. (2014) [45]	Conscious access is the process by which a piece of information becomes conscious content. Conscious processing refers to the various operations that can be applied to a conscious content (as when multiplying two numbers mentally). Conscious report is the process by which a conscious content can be described verbally or by various gestures. Such reportability remains the main criterion for whether a piece of information is or is not conscious, i.e., I can report something if and only if I am aware of it.	Self-consciousness is a particular instance of conscious accessibility where the conscious ‘spotlight’ is oriented toward internal states [...] The state of consciousness, associated with fluctuations in wakefulness or vigilance, refers to the brain’s very ability to entertain stream of conscious contents.
	Bartolomei et al. (2014) [46]	The global availability of information through the workspace is what we	

Theory	Authors (year)	Main Definition	Other Terms/Subcategorization Related to Consciousness (Definitions Reported If Available)
		subjectively experience as a conscious state.	
HOT	Lau (2007) [47]	Perceptual consciousness depends on the representation of the probability distributions that describe the behavior of the internal signal.	
	Lau and Rosenthal (2011) [48]	Consciousness consists in perceptual processes that occur with subjective experience, of which we are aware and about which we can report under normal circumstances. This contrasts with perceptual processes that occur without subjectivity, of which we are unaware and about which we cannot report. The term ‘conscious awareness’ can also apply to thoughts and volitional states of which one is subjectively aware	
	Friesen (2014) [49]	nf	
	Mehta and Mashour (2013) [50]	Consciousness is the story that our perceptual system tells us about the world. Any conscious state is a representation of what it is like to be in a conscious state, which is wholly determined by the content of that representation.	General consciousness: it pertains the levels of consciousness; - Specific consciousness pertains the content of consciousness
IIT	Balduzzi and Tononi (2008) [51]	Consciousness has to do with a system’s capacity to generate integrated information. This suggestion stems from considering two basic properties of consciousness: (i) each conscious experience generates a large amount of information by ruling out alternative experiences; (ii) the information is integrated, meaning it cannot be decomposed into independent parts.	
	Tononi (2008) [52]	Consciousness is integrated information and its quality is given by the	The quantity of consciousness generated by a complex of elements is determined by the amount of integrated

Theory	Authors (year)	Main Definition	Other Terms/Subcategorization Related to Consciousness (Definitions Reported If Available)
		informational relationships generated by a complex of elements.	information it generates above and beyond its parts. The quality of consciousness is determined by the set of all the informational relationships its mechanisms generate.
	Tononi (2012) [53]	Consciousness is what vanishes every night when we fall into dreamless sleep and reappears when we wake up or when we dream. Consciousness is synonymous with experience.	
	Casali et al. (2013) [54]	Conscious experience is both differentiated (i.e., it has many specific features that distinguish it from a large repertoire of other experiences) and integrated (i.e., it cannot be divided into discrete, independent components).	
	Oizumi et al. (2014) [55]	An experience (i.e., consciousness) is thus an intrinsic property of a complex of elements in a state: how they constrain—in a compositional manner—its space of possibilities, in the past, and in the future.	
	Tononi and Koch (2015) [15]	Consciousness is a fundamental property possessed by physical systems having specific causal properties. Consciousness is a fundamental, observer-independent property that can be accounted for by the intrinsic cause–effect power of certain mechanisms in a state—how they give form to the space of possibilities in their past and their future. Consciousness is a fundamental property of certain physical systems, one that requires having real cause–effect power, specifically the power of shaping the space of possible past and future states in	

Theory	Authors (year)	Main Definition	Other Terms/Subcategorization Related to Consciousness (Definitions Reported If Available)
		a way that is maximally irreducible intrinsically.	
	Tononi et al. (2016) [56]	Consciousness is subjective experience.	The quality or content of consciousness is identical to the form of the conceptual structure specified by the PSC. The quantity or level of consciousness corresponds to its irreducibility (integrated information Φ).
	Tsuchiya et al. (2016) [57]	Consciousness usually refers to either level or contents of consciousness.	The level of consciousness ranges from very high in the aroused awake state to low as in coma, vegetative states, deep dreamless sleep, and deep general anesthesia. At a given level of consciousness, every experiential moment contains various contents of consciousness. The contents of consciousness are synonymous with the other concepts such as qualia (its singular is a quale) of phenomenal consciousness.
LRMB	Wang (2012) [58]	Consciousness is a collective mental state of self-awareness that represents the bodily and mental status and their relations to the external environment, which is inductively generated or synthesized from the levels of metabolic homeostasis, unconsciousness, subconsciousness, and consciousness from the bottom-up. Consciousness is the basic characteristic of life and the mind, which is the state of being awareness of oneself, of perception to both internal and external worlds, and of responsive to one's surroundings. Consciousness is a collective state at the perception layer of the 7-layer LRMB model, such as the sensation, action, memory, perception, metacognition, inference, and cognitive layers from the	- Subconsciousness

Theory	Authors (year)	Main Definition	Other Terms/Subcategorization Related to Consciousness (Definitions Reported If Available)
		bottom-up. Consciousness is the sense of self and sign of life in natural intelligence. Consciousness is a collective and general state of advanced living organisms encompassing all cognitive attributes, which is generated based on the physiological structures, memory, and the brain's power for acquiring and manipulating sensory and mental information.	
MCTT	Dalla Barba and Boissè (2010) [59]	Consciousness is not a generic and a specific dimension that then becomes the consciousness of its object. It is immediately conscious of something. We would also add that consciousness is always consciousness of something in a certain way. This means that consciousness takes a certain point of view of its object and of the same object consciousness can take various points of view.	Knowing consciousness (KC) describes what is usually referred to as semantic memory. KC is temporal since it is a synthesis of what I have been. At the same time, KC is also atemporal in the sense that the time of which it is made is unrecognizable. Temporal consciousness is an organized, original, and irreducible form of consciousness for addressing the world. Unlike KC, TC transcends the mere presence of the object to set it in time.
Mesocircuit Hypothesis	Schiff (2010) [60]	Consciousness is the combination of different components: arousal, which relates to behavioral and physiological observations and establishes a threshold level that allows other aspects of higher consciousness to occur; awareness and motivation, which presupposes the will to interact with the environment and therefore can be intense as a “tension towards something”.	
Min's Model	Min (2010) [61]	Consciousness is a mental state embodied through TRN-modulated synchronization of thalamocortical networks. Consciousness consists of each mental unit, which is an individual thalamocortical looping mechanism, no	Awareness: conscious perception of an attended mental representation by strengthening relevant neural networks through thalamocortical reiterating.

Theory	Authors (year)	Main Definition	Other Terms/Subcategorization Related to Consciousness (Definitions Reported If Available)
		matter what cognitive stages it involves. Consciousness is referred to as thalamocortical response modes controlled by the TRN and is embodied in the form of dynamically synchronized thalamocortical networks ready for upcoming attentional processes	
NIH	Yu and Blumenfeld (2009) [5]	Consciousness is the ability to maintain an alert state, attention, and awareness of self and environment	
O'Doherty's theory	O'Doherty (2013) [62]	Consciousness represents the storage of past events for use in future situations and it is altered by external experience of the organism. Consciousness results from the gradual evolutionary development of the human information processing function. Consciousness is a phenomenon resulting from interactions between organisms rather than being located within an organism. Consciousness may consist of higher level of signal detection that has evolved through categorization and storage of past events.	
PFT	Morsella et al. (2015) [63]	Consciousness is a phenomenon associated with perceptual-like processing and interfacing with the somatic nervous system.	
PToC	Shannon (2008) [64]	Consciousness is the subjective experience and there are three types of consciousness: the sensed being, mental awareness, and meta-mentation.	Sensed being or sentience concerns the primitive and elementary aspect of consciousness that distinguishes the living from inanimate organisms. It has no specific context or structure; it is pervasive, and it is present during all our life. Mental awareness, typical of higher-order mammals, relates subjective experience that is distinct and differentiated, and the contents and

Theory	Authors (year)	Main Definition	Other Terms/Subcategorization Related to Consciousness (Definitions Reported If Available)
			forms of those experiences, such as mental images, ideations, flows of consciousness, and internal verbal monologues. Meta-mentation is the mind's ability to take its own productions as object for further reflection. This type of ability has different manifestation like meta-observation, reflection, monitoring, and control.
Quantum theories	Klemm and Klink (2008) [65]	Consciousness is the capacity of a system to opt among presented alternatives.	Unreflected consciousness refers to conscious acts and meanings in their immediate experiential and direct givenness.
	Das (2009) [66]	Consciousness is a property of the Nambu-Goldstone bosons created by Yukawa coupling between the Nambu-Goldstone boson scalar field and the electron Dirac field.	
	Di Biase (2009) [67]	Consciousness is non-local information with a status equal to matter and energy.	
	Koehler (2011) [68]	nf	
	Argonov (2012) [69]	Consciousness in the phenomenal sense is a synonym of "subjective reality", i.e., perception, volition, mental images, and emotions	
	Georgiev (2013) [6]	Consciousness is a collective term that refers to the subjective character of our mental states, and of our ability to experience or to feel. A conscious state is a state of experience.	
	Li (2013) [70]	nf	
	Hameroff and Penrose (2014) [71]	Consciousness consists of a sequence of discrete events, each being a moment of 'objective reduction' (OR) of a quantum state (according to the DP scheme),	

Theory	Authors (year)	Main Definition	Other Terms/Subcategorization Related to Consciousness (Definitions Reported If Available)
		where it is taken that these quantum states exist as parts of quantum computations carried on primarily in neuronal microtubules. Such OR events would have to be ‘orchestrated’ in an appropriate way (Orch OR), for genuine consciousness to arise.	
	Hoffman and Prakash (2014) [72]	A conscious agents as a six tubules $C = ((X,X), (G,G), P,D, A,N))$, where: (X, X) and (G, G) are measurable spaces; $P: W \times X \rightarrow [0, 1]$, $D: X \times G \rightarrow [0, 1]$, $A: G \times W \rightarrow [0, 1]$ are mathematical formalism of Markovian kernels, where N is an integer.	
	Kak et al. (2016) [73]	States of consciousness are mediated by languages and metalanguages with varying capacity or ability to recruit memories.	
	Sieb (2016) [74]	Conscious experience is defined as the direct observation of conscious events. A conscious event [...] is the fundamental entity of conscious experience (observed physical reality) represented by three coordinates of space and one coordinate of time in the space–time continuum: conscious event that consist of a set of qualia; conscious experience that is intimately tied to perception; postulated for conscious experience; conscious experience that is essentially an orientation in space and time.	
	Brabant (2016) [75]	nf	
Reji Kumar’s theory	Reji Kumar (2010) [76]	Human consciousness is built with models and these models are involved in the development of consciousness at a very basic level. As such, consciousness is subject to all advantages and	

Theory	Authors (year)	Main Definition	Other Terms/Subcategorization Related to Consciousness (Definitions Reported If Available)
		limitations of models. Consciousness is the result of all information processing activities that occur in the mind.	
	Ahmad and Khan (2012) [77]	Consciousness is the core component of the mind. Consciousness can be defined as the spirit of consciousness and the transitivity of an information system. Transitivity can be defined as a certain correspondence between the system of information and outside matters. Consciousness is the final output of all information processing activities that occur in the mind.	
	Reji Kumar (2016) [78]	Consciousness is the result of all processes that occur in the brain.	
	Reji Kumar (2016) [79]	Consciousness is the totality of all effects of the functions of the brain and the supporting nervous system that produces the feeling of the objective world and the subjective mind or self	
RPT	Cleeremans et al. (2007) [80]	Consciousness is the brain's theory about itself, gained through experience interacting with the world, and, crucially, with itself [...] Conscious experience occurs if and only if an information processing system has learned about its own representations of the world [...] Experience is, almost by definition ("what it feels like"), something that takes place not in any physical entity but rather only in special physical entities, namely cognitive agents.	
	Cleeremans (2008) [81]	nf	
SPC	Thagard and Stewart (2014) [82]	Consciousness is a neural process resulting from three mechanisms: representation by firing patterns in	

Theory	Authors (year)	Main Definition	Other Terms/Subcategorization Related to Consciousness (Definitions Reported If Available)
		neural populations, binding of representations into more complex representations called semantic pointers, and competition among semantic pointers to capture the most important aspects of an organism's current state.	

[Open in a new tab](#)

Note. The table enlists the qualitative data retrieved from each article included in the review grouped by theory (first column). The main definitions of consciousness are quoted from the original articles (please see [Table S3](#) for other definitions of consciousness and related terms). nf = not found in the article analyses.

Table 2.

Number of citations. The table shows how many times (second column) a specific term (first column) was used when referring to consciousness by theories (third column).

Terms	N. of Citations	Theory
Activity	4	ADT, DCT
Awareness	13	AIM, AST, HOT, LRMB, Mesocircuit Hypothesis, NIH, PToC
Brain	6	EMI/CEMI, GWT/GNW, LRMB, Reji Kumar's theory, RPT
Cognitive	4	LRMB, Min's Model, RPT
Collective	6	LRMB, Quantum theories
Component	4	EMI/CEMI, IIT, Mesocircuit Hypothesis, Reji Kumar's theory
Content	7	GWT/GNW, HOT, IIT
Events	7	O'Doherty's theory, Quantum theories
Experience	26	AIM, CP, CSS, GWT/GNW, HOT, IIT, O'Doherty's theory, PToC, Quantum theories, RPT
External	4	AIM, LRMB, O'Doherty's theory
Feeling	7	Damasio's theory, Reji Kumar's theory
Field	6	EMI/CEMI, Gurwitsch's theory, Quantum theories
Form	4	EMI/CEMI, IIT, Min's Model
Function	5	GWT/GNW, O'Doherty's theory, Reji Kumar's theory
Information	20	ART, AST, EMI/CEMI, GWT/GNW, IIT, LRMB, O'Doherty's theory, Reji Kumar's theory, RPT

Terms	N. of Citations	Theory
Integrated	5	IIT
Level	6	Agnati et al.'s proposal, IIT, LRMB, Mesocircuit Hypothesis, O'Doherty's theory, Reji Kumar's theory
Mental	10	CSS, GWT/GNW, LRMB, Min's Model, Quantum theories
Mind	5	LRMB, Reji Kumar's theory
Neural	4	Damasio's theory, DCT, SPC,
Object	5	MCTT, Quantum theories, Reji Kumar's theory
Organism	7	Damasio's theory, LRMB, O'Doherty's theory, SPC
Past	5	IIT, O'Doherty's theory
Perception	5	LRMB, Quantum theories
Perceptual	5	AST, GWT/GNW, HOT, PFT
Physical	7	AST, Damasio's theory, EMI/CEMI, IIT, Quantum theories, RPT
Possible	5	CP, GWT/GNW, IIT
Power	4	IIT, LRMB
Processes	17	Agnati et al.'s proposal, CSS, GWT/GNW, HOT, Min's Model, Reji Kumar's theory
Property	9	AST, CSS, IIT, Quantum theories
Reality	4	Bieberich's theory, COI, Quantum theories
Representation	12	COI, Damasio's theory, GWT/GNW, HOT, RPT, SPC
Something	5	GWT/GNW, MCTT, Mesocircuit Hypothesis, RPT
Space	6	IIT, Quantum theories
State	28	CP, Damasio's theory, EMI/CEMI, GWT/GNW, HOT, LRMB, Min's Model, Quantum theories
Subjective	14	GWT/GNW, HOT, IIT, PToC, Quantum theories, Reji Kumar's theory
System	11	AST, HOT, IIT, PFT, PToC, Reji Kumar's theory, RPT,
Thalamocortical	4	Min's Model
World	7	AIM, EMI/CEMI, HOT, LRMB, Reji Kumar's theory, RPT

[Open in a new tab](#)

Moreover, other terms related to definitions or sub-categorizations of consciousness were also reported in [Table 1](#). Considering only adjectives/nouns reported before the

word consciousness, we found 21 different terms related to consciousness or consciousness sub-categorization (in alphabetical order): access consciousness; background consciousness; core consciousness; elevated consciousness; extended consciousness; foreground consciousness; general consciousness; knowing consciousness; marginal consciousness; meta-consciousness; micro-consciousness; perceptual consciousness; phenomenal consciousness; primary consciousness; reflective consciousness; secondary consciousness; self-consciousness; specific consciousness; subconsciousness; temporal consciousness; and unreflected consciousness.

[Table 3](#) shows the main NCC reported in the various theories. The most cited neural structures were related to sub-cortical structures, although several theories reported cortical structures as well, defining in detail NCCs that could refer to cortico-thalamic systems.

Table 3.

Main neural correlates of consciousness reported in the various theories. In the second column, the percentage of theories referring to a specific neural structure is reported. The third column contains the name of theories that were reported for each neural structure. The fourth column indicates the percentage of theories that refer to either micro-, subcortical-, cortical-structures, or entire systems as neural correlates of consciousness.

Brain Structures	Theories Percentage Single Structure	Theories	Theories Percentage
Micro-structures			
Single Neuron	3.44%	Agnati's theory	24%
Pyramidal neurons	6.89%	ADT, Bieberich's Theory	
GABAergic inhibitory neurons	3.44%	Min's model	
Periaqueductal gray neurons	3.44%	CP	
Astroglial-pyramidal cells	3.44%	CEMI	
Glia cells	3.44%	Agnati's theory	
Dendritic cells	10.34%	ADT, Bieberich's Theory, Quantum Theories	
Neural Microtubules	3.44%	Quantum Theories	
Ionic channels	10.34%	Agnati's theory, Bieberich's Theory, Quantum Theories	
Cortical minicolumns	3.44%	ADT	
Cortical single structures			

Brain Structures	Theories Percentage Single Structure	Theories	Theories Percentage
Prefrontal cortex (PFC)	24.13%	Agnati's theory, CP, GWT, HOT, Min's model, Quantum Theories, RPT	48%
Orbitofrontal areas	3.44%	ART	
Dorsolateral prefrontal cortex (dlPFC)	6.89%	COI, CSS	
Medial pre-frontal cortex (mPFC)	10.34%	AST, COI, CSS	
Cingulate cortex	13.79%	CSS, Damasio's theory, GWT, NIH	
Anterior cingulate cortex	13.79%	Agnati's theory, COI, CSS, GWT	
Experiadorsal anterior cingulate cortex (dACC)	3.44%	CSS	
Medial cingulate gyrus	3.44%	CP	
Posterior cingulate cortex (PCC)	3.44%	CSS	
Precuneus	6.89%	CSS, NIH	
Medial frontal cortex	3.44%	NIH	
Supplementary motor area (SMA)	3.44%	CSS	
Premotor cortex	6.89%	CSS, GWT	
Parietal cortex	10.34%	CSS, GWT	
Superior parietal lobule	3.44%	CSS	
Inferior parietal lobule	3.44%	CSS	
Posterior parietal cortex (PPC)	10.34%	CP, HOT, quantum theories	
Temporoparietal junction	6.89%	AST, CSS	
Temporal lobe	17.24%	ART, AST, CSS, GWT, MCTT	
Occipital cortex	3.44%	GWT	
Subcortical single structures			
Basal ganglia	10.34%	ART, CP, quantum theories	52%
Nucleus accumbens	3.44%	Agnati's theory	

Brain Structures	Theories Percentage Single Structure	Theories	Theories Percentage
Clastrum	3.44%	Agnati's theory	
Insula	6.89%	Agnati's theory, CSS	
Amygdala	3.44%	ART	
Hippocampus	17.24%	ART, CSS, GWT, quantum theories, RPT	
Cerebellum	10.34%	GWT, LRMB, Reji Kumar's theory	
Thalamus	24.13%	Agnati's theory, CP, Damasio's theory, DCT, GWT, HOT, LRMB	
Anterior Pretectal nucleus	3.44%	DCT	
Medial thalamus nuclei	3.44%	NIH	
Thalamic reticular nucleus	6.89%	DCT, Min's model	
Pulvinar	3.44%	DCT	
Hypothalamus	6.89%	ART, CP	
Midbrain	3.44%	CP	
Superior colliculus	6.89%	CP, Damasio's theory	
Ventral tegmental area	3.44%	Agnati's theory	
Brain stem	3.44%	NIH	
Raphe nuclei	3.44%	Agnati's theory	
Brain systems/regions			
Cortico-thalamic system	17.24%	ART, AIM, IIT, Mesocircuit hypothesis, quantum Theories	28%
Fronto-parietal areas	6.89%	CSS, GWT	
Sensory cortical areas	3.44%	GWT	
Perceptual regions	3.44%	PFT	
Limbic system	3.44%	AIM	
Striate regions	3.44%	GWT	

[Open in a new tab](#)

3.2. Analytical Description of Each Theory

In the next sections, we report a brief description of each theory analyzed (in alphabetical order).

3.2.1. The Apical Dendrite Theory (ADT)

The apical dendrite theory (ADT) of consciousness, proposed by LaBerge and Kasevich [18], focuses on input-processing, which seeks to understand how consciousness is generated by neurons and how information processing sustains itself over a prolonged time period without apparent outputs.

The theory affirmed that the pyramidal neurons within the thalamus-cortical circuit are important anatomical structures for consciousness due to their long apical dendrite, which produce a special kind of electrical activity.

Authors focused on structures called “minicolumns”, which are mini structures composed of major pyramidal neurons, stellate neurons, inhibitory neurons, and axons. The minicolumn is regarded as the functional unit of the cortex [83], based on the observation that neurons in each minicolumn share many receptive field properties. This theory analyzed two main circuits of the minicolumn: the shell and axis circuit.

The first one is a one-directional input–output circuit that connects one minicolumn with another, and the processing component consists of connections between layer 2 and 3 of neurons within the minicolumn, creating a sort of “horizontal” circuit. The axis circuit, instead, is considered as an “input-stayput” circuit due to its apparent role as a holder of neural activity over time. It constitutes the “vertical” circuit of the layer 5 cortical minicolumn, which extended to the thalamus in a reciprocally connected manner in a looping way.

The theory assumed that sustained activity in a primary sensory area provides the ongoing activities for a background consciousness. Elevated consciousness for selected aspects of background consciousness is assumed to arise when sustained activity of a primary sensory area is sent to higher sensory areas, where a selected part of the sensory scene is amplified by attentional activity controlled by the frontal lobes.

Within this attentional selection process, the control for the selective amplification/suppression activity is hypothesized to be provided by frontal cortex axons, which connect with the thalamic nuclei that serve those microcolumns. On the other side, the thalamic nuclei receive both this frontal information and data from primary sensory microcolumns, which remain relatively constant.

The shell circuits in early sensory areas that receive V1 inputs identify object appearances and locations. Later, sensory areas process these identifications into

categories. In the frontal areas, categories are organized into propositions that may be based mainly on activity in the shell circuitry, which is spread across many minicolumns. On the other side, the global axis pathway begins with the initial cortical registration of the sensory inputs in the primary sensory minicolumns, and it extends into the frontal cortex (where presumably special supporting functions are performed and act as holding circuits that sustain activity for extended periods of time). When the intensity of the apical dendrite component of the thalamus-cortical holding circuit is sufficiently high, it produces the cognitive event, which may be labeled as “having an impression” of something. Two kinds of pathways are shown to sustain this activity: global ascending pathways that connect posterior cortical columns with anterior cortical columns and the global descending pathways that connect anterior cortical columns with posterior cortical columns.

Finally, the authors analyzed the properties of the electromagnetic activity of the apical dendrite to determine what properties of conscious impressions they may support. Interestingly, they affirmed that clusters of neighbor apical dendrites developed electric field energy in their core structures from the combined electric field components parallel to the dendritic axes. The development of magnetic field energy between parallel electric field dipoles is much less significant compared to electric field energy coupling between pairs of parallel apical dendrites with internal collinear dipoles moving at the conduction velocity. According to the present theory, the electromagnetic field intensity would underlie our immediate impressions of consciousness.

3.2.2. Agnati et al.'s Proposal

Agnati et al. defined consciousness “as the global result of integrative processes taking place at different levels of miniaturization in plastic mosaics” [20].

The authors mainly referred to Cook and Sevush’s idea of a single-neuron theory of consciousness [19,84]. Accordingly, the synaptic connections within a network could represent the substrate of cognition, whilst the ion-flows during the action potential could represent the substrate of sentience. The authors claimed that both cognition and sensitivity could be present in the glial cells (other than neurons), which could represent in this way a prerequisite for consciousness and subjectivity. Furthermore, it is postulated the existence of the so-called functional modules representing microcircuits of neurons and astrocytes, which are well-organized in specific patterns to carry out a specific process. The functional modules do not have anatomical boundaries as they are determined by a functional structure. Furthermore, the functional modules are characterized by both vertical and horizontal organizations. The vertical organization includes 3 well-segregated levels, namely the molecular, local circuit, and cell network. Each level is then characterized by a horizontal organization; thus, the molecular level includes transmitters and receptors. The local circuit includes aggregates of synapses working as a unit. The cell network is composed of neurons and glia cells representing a network. A set of different functional modules could correspond to diverse sensory information that needs to be gathered in a high order fashion to give rise to conscious

processes. The functional modules perform the first-level integrative functions that transform sensation into perception. The sensory information deriving from the outside world is spread to the cortex by the thalamus, which receives and redirects both sensory (from the outside world to the cortex) and cortical (from the cortex to other structures/nuclei) afferences. The resting-state network (encompassing medial prefrontal, parietal, and cingulate cortices) activity has been considered by the authors as a background condition to bind different information together. However, what is the phenomenon that is directly responsible for such a bind is still under debate. The authors considered the synchronization of the neural activity (detected through the gamma range recording) as a possible explanation [20]. A similar hypothesis was advanced by Cook who affirmed that the contents of both cognition and subjective feeling were grounded in temporal synchronization of synaptic activity and firing of action potentials, respectively [19].

As for the anatomical structures involved in the emergence of consciousness, the authors focused on the key role of the thalamus in receiving and redirecting both sensory and cortical input [20]. Furthermore, they considered the role of the brainstem system as necessary for the emergence of consciousness. The locus coeruleus has a role in the arousal maintenance, the raphe nuclei (by means of its projections) has a role in slow-wave sleep maintenance, and the ventral tegmental area is crucial because of its projections to the limbic and prefrontal cortex (two areas implicated in reward, attention, working memory, and emotions). Lastly, the claustrum is considered as a region of integration between perceptual, sensory, and motor information. Its role is similar to the thalamus, the main difference being the integration mechanisms. The thalamus is assumed not to directly integrate the activity of all of its nuclei.

3.2.3. REM Sleep–Dream Protoconsciousness Hypothesis (AIM)

Hobson [21] took into account the definition of primary and secondary consciousness, as postulated by Edelman [85], along with rapid eye movement (REM) sleep and dream models, to hypothesize the existence of waking consciousness, defined as “the awareness of the external world, our bodies and ourselves (including the awareness of our awareness) that humans experience when awake” [21].

He analyzed the difference between the awake experience and the sleep experience in relation to Edelman’s definitions [86] before considering the evolutionary and developmental aspects of REM sleep.

Waking consciousness is richer than dream consciousness, especially considering the measurable evaluation between task and brain activity, or between background and foreground processing. This contrasts with dream consciousness, which appears richer than waking consciousness in creating a scenario with different mental images. For the author, dreaming abounds in features of primary consciousness, especially perceptions

and emotions, which are produced by the brain without external stimulation, whereas waking is fulfilled by secondary consciousness.

The hypothesis gives particular attention to the cortical-thalamic and limbic systems, which are active in waking and REM sleep for conscious experience. Moreover, this hypothesis considers dream consciousness and REM to start at the age of 5 years or as late as 8 years (there is an open debate on this point). This is when brain development has advanced sufficiently so that the narrative organization of subjective experience becomes possible. Considering that REM sleep appears before dreaming, the authors suggested that REM is one of the first signals that indicates that the brain is preparing itself for gradual development of integrative work among different functions, one of which is consciousness. The activation of the forebrain in the absence of external input during sleep could be important for the provision of an automatic, built-in, self-organized process that offers a spontaneous solution to the so-called binding problem, as demonstrated by the fact that REM sleep enhances sensorimotor integration.

The theory considered three factors that may be computerized: activation (A) (large parts of the brain are not only a collection of passive reflexive circuits but they also possess the means of regulating their own activation); input–output gating (I) and input–output gate control (factor I), which is mediated by a brainstem [87] that guarantees the coordination of factors A and I; modulation (M) (all modulatory brain cell populations in the brainstem during the move of being awake to nREM sleep to REM sleep, the so-called REM-off cells (which are active during waking and inactive during REM sleep [88], and REM-on cells (which are active during REM sleep and inactive during waking)). Joining the AIM model (which proposes that the wake–nREM–REM sleep cycle is the result of interactions between aminergic REM-off cells and cholinergic REM-on cells) to this ‘activation synthesis’ hypothesis of dreaming [89] (which posits that brain activation during REM sleep results in the synthesis of dream mentation), the authors affirmed that waking consciousness, with its impressive secondary features, might be present only in humans, who have the highly evolved and extensive cortical structures necessary to mediate the abstract aspects of conscious awareness.

3.2.4. Adaptive Resonance Theory (ART)

Grossberg [23] aimed to develop a theory of consciousness that is able to answer to the so-called “hard problem” of consciousness. The authors stated that “consciousness is not just a whirl of information-processing” [23], which purports to the already existing adaptive resonance theory (ART), where resonance consists of dynamical states represented by the amplification of neuronal firing and interactions between top-down and bottom-up processes [23]. This resonance is adaptive since it can trigger learning and memory traces. ART’s assumption is that “adaptive resonance refers to the fast learning that resonant states can drive” [22]. Resonant states bind together to build coherent world representations (i.e., learned bottom-up categories are bound with learned top-down expectations). The bottom-up categories and top-down expectations are bound only if there is a match between them. The matching between bottom-up

external input and the top-down expectation results in resonant states. In this framework, all conscious states are resonant states, but the reverse is not true. The authors stated that the rules of ART align with several sensory and cognitive processes, as attested to by different evidence.

A key process for producing conscious states is represented by attention. The authors enlisted three different types of attention in the framework of the spatial cognition: (i) boundary attention (throughout propagation, the spatial attention selects a certain object for inspection); (ii) surface attention (acting in a scene with eye movements that allow the learning of view-invariant categories); (iii) prototype attention (the kind of attention produced via the ART model, which enhances critical features of certain bottom-up patterns). The latter kind of attention represents the attentional process involved in the ART mode of functioning.

Furthermore, the authors introduced the concept of the stability–plasticity dilemma to explain the way by which the brain can quickly learn new information without falling into catastrophic forgetting. This principle, in the author's view, should be at the base of the way the brain unifies different sources of information into a coherent conscious experience. In order to explain the mechanism at the base of the stability–plasticity dilemma, the authors firstly considered the human beings as both intentional (meaning that the individuals learn expectations and make predictions about what is going to happen in the real world) and attentional (meaning that processing resources are focused on a restricted pool of information each time). If one has some expectations (i.e., searching for an object of a specific color among objects of different colors), then the attentional mechanism leads to the matching process between the bottom-up pattern and the expectations, amplifying the activation of the pattern resulting in a resonant state. In such a framework, only resonant states lead to rapid new learning. Hence, only experiences that attracted our attention and which have been learned can be considered conscious experiences.

As for the neural underpinnings of the learning process, a pivotal role is played by the thalamo-cortical circuit. Indeed, the first-order thalamic nuclei relay the sensory information towards cortical areas whilst the second-order thalamic nuclei receive inputs from low-order cortical areas. The learning of neural representations should correspond to the match between the first- and the second-order thalamic nuclei.

The authors explained the mechanisms underlying the ART model by using the visual system as an example. In this framework, the concept of attentional matching was introduced, meaning that matching between bottom-up and top-down processes were responsible for learning mechanisms. Specifically, whenever an individual receives an input, this can lead to the activation of target cells (bottom-up mechanism), while, at the same time, top-down mechanisms activate the on-center cells (inhibiting the off-surround of the same cells) and the attentional focus arises from the eventual match between top-down and bottom-up mechanisms. The attentional matching is

characterized by a reinforcement mechanism as the attentional focus boosts the bottom-up process, which, in turn, emphasizes the top-down process. Thus, the learning process occurs only if there is a good match between bottom-up and top-down processes. Conversely, when something new has to be learned, it is necessary to activate the orienting system. This situation happens with unfamiliar events that are related to a mismatch between bottom-up and top-down mechanisms. This “mismatched” situation elicits a memory search mechanism. Whilst the prefrontal cortex is supposed to be related to attentional matching, the hippocampus seems to be related to mismatch processing. Furthermore, evidence supports an association between attentional matching and faster gamma oscillations, while mismatch processing appears to be associated with slower beta oscillations.

Furthermore, the authors also considered the role of emotions because they are different from visual and auditory inputs since they are internally generated. However, even in this case, a cognitive-emotional resonance process has been hypothesized by using feedback interactions between invariant objects categories, value categories (representations), homeostatic representations, and object-value categories that receive object and value category inputs.

3.2.5. Attention Schema Theory (AST)/Graziano's Theory

The present hypothesis suggested that awareness is a perceptual reconstruction of the attentional state. The fundamental position of Graziano and Kastner [24] was that consciousness is not an emergent property but is itself information computed by an expert system.

According to previous results [90,91], the authors affirmed that “the machinery that computes information about other people’s awareness is the same machinery that computes information about our own awareness” [24]. In other words, when a subject constructs a perceptual model of someone else’s attention (e.g., “person A is aware of X”), he/she also computes secondary information such as “I am aware of X”. In this sense, social perception is the mechanism through which stimulus representation surpasses neural competition derived from other representations.

For this theory, awareness is a form of meta-social intelligence reconstructing someone else’s thoughts, beliefs, or emotions, which also determines the state of someone else’s attention. Consequently, when social perception is applied to oneself, it provides not only a description of one’s own inner thoughts and feelings but also “a description of one’s awareness of items in the outside environment” [24].

To support their hypothesis, the authors cited results from studies on neglect and out of body experience, discussing the role of the superior temporal sulcus, temporal-parietal junction, medial prefrontal cortex, and the frontal-parietal circuits as key areas for social perception. In relation to this point, the authors discussed interesting predictions derived

from their theory. For instance, they hypothesized the existence of different kinds of neglect associated with two neural networks—one for controlling attention (neglect caused by damage to parietal-frontal attentional mechanisms) and the other for perceptual attention (caused by the temporo-parietal junction and superior temporal sulcus impairment).

3.2.6. Bieberich's Theory

In the model proposed by Bieberich [25], “consciousness determines what is perceived as reality” [25]. It is composed by sentyons that belong to the physical world.

In the authors' view, the binding mechanism is at the base of the perception of reality, i.e., there are stimuli (visual, auditory etc.) within the space (i.e., environment) that are consciously perceived in an internal space (i.e., endospace); in doing so, the system needs to bind together different information deriving from the space to consciously experience the endospace. This binding mechanism should be operated at the single neuron level. However, according to the authors' view, in the neural space there would only be a summation process without any integration mechanism leading to conscious perception.

The dendritic trees are sites where the input signal starts and are summed with other signals: The sum of different signals, when surpassing the threshold level, provokes the firing of a client neuron. The latter sends the signal to other neurons in the network, and, in turn, it receives recurrent signals from the network's neurons, thus creating a so-called “psychic loop”. Furthermore, the dendritic trees in a neuron receive the synaptic activity of different frequencies. Since the signal flowing inside the dendritic trees is slower than the signals' propagation within the network, the signal looping times of a closed and distant network would be the same; this leads to an inverse connectivity due to the reverse relationship between the signal delay times within the dendritic tree and the recurrent signal delay times of the psychic loop. Thus, in the recurrent network with inverse connectivity, the farther the signal travels outside of the client neuron, the shorter it travels inside the neuron when the signal returns. This recurrent network is supposed to have a fractal architecture. The idea at the base is that the maps of a single part (e.g., dendritic tree inside the client neuron) are similar to the whole map (e.g., the neural network). To switch from the whole map to part of the map, a downscaling mechanism should be applied. Both mapping in a fractal way and downscaling are mathematically defined by the author who attempts to explain how these principles could describe how the entire brain functions. By assuming a refractal structure and by applying a downscale computation, it would be possible to describe both macro- and the micro-neural structures. These mechanisms define the recurrent fractal networks intended to underlie the emergence of consciousness as defined by the author.

Within this view, consciousness (derived from the activity of the recurrent fractal networks) resides at a single neuron level. However, because the networks connected between them are structurally fractal, consciousness is also diffused, as hypothesized by the global workspace theory [92].

In discussing the brain areas associated with consciousness, the authors suggested that the olfactory system were the more likely candidate to be the site of consciousness due to neuron features. It could at least represent a system with the minimum requirement for consciousness. Another site linked to consciousness is represented by both the hippocampus and the emotional centers of the brain (such as the amygdala). At a molecular level, calcium has been identified as responsible for the emergence of consciousness. Furthermore, by applying the downscaling, the fractal structure is supposed to present at the molecular level (at the level of ions channels and calcium).

3.2.7. The Cross-Order Integration (COI) Theory

The COI theory provides a hypothesis for the NCC, focusing on the following question: “Why would this particular neural feature, rather than another, correlate with consciousness?” [26].

One of the COI assumptions is that psychological events and states are representations in the brain, some of which can be conscious whilst others can remain non-conscious. Yet we can envisage both a conscious and non-conscious representation of the same stimulus.

Considering Rosenthal [93,94], Kriegel affirmed that “although representing something is not the mark of conscious states, it may well be that being represented is. This family of views is motivated by the basic idea that conscious states are states we are aware of” [26].

According to COI theory, conscious states arise from the integration, or unification (two representations become a single representation), of what is initially separate: (i) a first-order representation of an external stimulus and (ii) a higher-order representation of that first-order representation. Thus, when an individual has a higher-order representation that is unified with the first-order representation it represents, the conscious state consists of this representational unity. In this sense, the COI is different from the high order theories because high order representation is external to the conscious state, i.e., it is unconscious and does not take part in the subject’s overall phenomenology.

Therefore, to generate a testable hypothesis about the NCC using the COI theory, the authors identified three levels: (i) a floor-level representation that determines the specific contents of consciousness (e.g., visual and auditory inputs) once consciousness is present, rather than ensuring the presence of consciousness in the

first place; (ii) a higher-order representation of that floor-level representation; and (iii) the functional integration of these two representations. There is a neural triad of correlates that need to be pinned down before we can specify the neural correlate of a given conscious state.

According to COI, the second and third elements of the triad—the functional integration with the higher-order representation—constitutes the correlate of consciousness and not the content of consciousness.

The second element requires the ability to direct the cognitive system onto itself (involving prefrontal cortex), which is associated with executive and other forms of cognition functions (e.g., state-metacognition) [26]. Finally, the theory describes the third element as “the process of unifying disparate bits of information into a single representation in a functionally significant way, that is, in such a way that the functional role of the single representation is in some sense more than the sum of the functional roles of the different bits of information making it up” [26].

The COI affirmed that there are seven possible hypotheses about the neural correlates of higher-order representations. Three hypotheses suggest a single neuroanatomical site (the anterior cingulate cortex (ACC), the dorsolateral prefrontal cortex (dlPFC), the medial prefrontal cortex (mPFC)), whereas three other hypotheses suggest two neuroanatomical sites (ACC and dlPFC; ACC and mPFC; and dlPFC and mPFC). Lastly, one hypothesis suggests three sites (ACC, dlPFC, and mPFC).

As for the NCC associated with the third element (integration), the authors cited some studies regarding patients who exhibit “misbinding” or “illusory conjunctions” [95] due to different activities not coordinated by specific brain areas. Specifically, the authors considered the results deriving from research on the visual system to hypothesize that the synchronized activities in the visual and prefrontal cortices would be responsible for a fusion of the visual world- and self-awareness. Consequently, these areas can be considered neural correlates of visual consciousness.

3.2.8. The Centrencephalic Proposal (CP)

The so-called centrencephalic proposal, developed and revisited by Merker [27], is grounded in Penfield and Jasper’s assumption of a “highest integrative function which, while anatomically subcortical, is functionally supra-cortical” [27,96]. These authors claim that the brainstem structure plays a pivotal role in determining a certain level of consciousness. Indeed, lesions at the brainstem level can impair normal cortical functionality. Specifically, the main subcortical structures, which are part of this “brainstem system”, are the ventral tegmental area, substantia nigra, colliculus, thalamus, raphe nucleus, and pontine reticular formation. According to Merker’s view, consciousness can be defined as “the medium of any and all possible experience” [27] and, similarly to Penfield and Jasper’s assumption [96], it is determined by sub-cortical

activity that influences the cortical one. Within this framework, Merker's theory postulated the existence of the so-called "triangle selection", which encompassed the (i) target selection, (ii) action selection, and (iii) motivational ranking. Each component is stand-alone as they can be defined in their own way; however, they are interrelated as the action selection could depend on the target, and the motivational ranking could influence both the target and the action selection. The authors claimed that to interact with the environment and to take any kind of decision, this interaction must occur in real-time and, usually, it happens in the form of simulation. Using Merker's words, "The way this simulation is structured constitutes a conscious mode of function" [27]. Furthermore, when anyone is conscious, the above-mentioned interaction happens in an ego-centric frame of reference (where the contents of sensory consciousness are disclosed), which could be biased by the motivation. The neural structures involved in such interaction are represented by a system ranging from the colliculus to the periaqueductal gray. The colliculus, however, is considered the key structure in the present theory due to its bidirectional connections with the cortex and its laminar structure, which is strongly related to a multisensory topography. This multisensory layered structure (due to the afferences coming from different cortical areas) leads to a sort of multisensory integration mechanism. Furthermore, when the neural information crosses the collicular threshold, the cortical activity enters the mesodiencephalon, which is considered the first step for a process to be conscious.

3.2.9. The Consciousness State Space Model (CSS)

CSS represents a three-dimensional neurophenomenological model composed of three dimensions: time, awareness, and emotion.

According to the CSS, consciousness is fragmented into two different but interrelated categories [28,29]. The first is core consciousness (CC), which is related to the here and now. For instance, this category supports a sense of agency, ownership, and first-person content (by using the authors' words, "it encompasses the minimal self") [28]. The second category is represented by extended consciousness (EC), which is related to both episodic and prospective memory, as well as to verbal thoughts ("it encompasses the narrative self") [28]. These two categories are organized in an embedded way as the CC represents a 'low-order' category being strongly related to the body-sense, whilst the EC is far away from the body and is related to higher-order cognitive processes. Furthermore, each category is organized along a continuum of three dimensions. Specifically, the time dimension (strongly supported by the mnemonic functions) ranges from past to future (i.e., retrospective and prospective memory-related; EC) passing through the here and now (i.e., working memory-related; CC). Similarly, the awareness dimension extended from an absence of the awareness to a full-awareness state (included in the EC category), passing through a sensory and subliminal bodily awareness (included in the CC category). Moreover, the awareness dimension appears to be related to attentional processes that determine the degree of awareness itself. Finally, the emotional dimension is subdivided into two sub-components, namely the valence component (through which a positive or negative

value is ascribed to a specific event; EC) and the arousal component (which represents the bodily and visceral activations in relation to a specific event; CC).

As for the neural underpinning of CSS, the authors attributed the activity of those regions belonging to the default mode network to the EC, whilst the activity of regions implicated in multisensory integration processes would be at the base of the CC. This implies a widespread set of regions underlying the CSS, which encompass both cortical and subcortical areas. Since CC and EC are assumed to be interrelated, the responsible network of such interaction could be represented by the frontoparietal network, which encompass both the executive control network (i.e., dorsal attention network) and the salience network.

3.2.10. Damasio's Theory

The aim of the study by Bosse et al. [30] was to test a simulation model related to Damasio's theory of consciousness [97].

The authors considered three components as conceptualized by Damasio: (i) emotion as a neural representation (deriving from neural activation after a certain change in the environment); (ii) feeling as body change perception (the emotion causes body changes that represent observable emotional states); and (iii) core consciousness (or feeling a feeling) as the detection of the body change after a stimulus has occurred.

To create a simulation model of Damasio's theory, the authors referred to the temporal trace language (TTL; [98]), which allows computing a change of the state if given temporal parameters. For instance, the change of state A into B is computed by considering the duration of state A, the delay from state A to state B, and the duration of state B.

By using this model, an emotion is the result of external stimulation (e.g., auditory), which is detected and induces an internal representation (which is supposed to be multidimensional). This, in turn, prepares the body to act (each representational dimension corresponds to a component of body preparation to action).

The feeling is related to two different dynamics in the model: the "body loop" and the "as if body loop". The first represents real change in the body and the second represents a cognitive change related to the body instead of a real change in the body.

The core consciousness (i.e., feeling a feeling) occurs when changes caused by the external event/stimulus are detected and attributed to a specific event/stimulus. In this context, the temporal specification plays a pivotal role (see the TTL principles as specified before). Moreover, the authors stated that a false core consciousness could

occur when, for instance, two events occur even if only one of these is detected. The body change may occur due to the event not being detected; however, the subject, in this case, would erroneously associate the occurred body change with the only detected event.

For neural correlate concerns, the authors referred to Damasio's theory, which considers the pivotal role of the cingulate gyrus, thalamus, and superior colliculus [97].

In this framework, subjectivity is outlined as a possible explanation for different feelings related to the same event/stimulus in different individuals.

3.2.11. The Thalamic Dynamic Core Theory (DCT)

In 2011, Ward proposed the thalamic DCT, which is grounded in similar mechanisms explained by other authors. Further, it consists of the synchronous firing of large cortical neuronal population [86,99]. However, different from previous theories, the authors affirmed that neither a specific cerebral structure nor a single specific kind of neural activity is necessary nor sufficient for explaining conscious states [31]. In other words, the theory assumes the existence of consciousness thanks to a distributed neural activity with a central role of the synchronized neural activity in cortico-thalamic circuits. The central role of the thalamus is supported by evidence on both patients with disorders of consciousness and from the functional data during the REM and NREM sleep patterns in healthy subjects. A novelty in this theory is represented by the distinction of two different neuronal populations within the thalamus, although this was already postulated by Jones [100]. Specifically, the "core neurons" are typical of sensory and motor thalamic nuclei and represent relay stations for sensory and motor pathways, whilst "matrix neurons" are typical of non-sensory thalamic nuclei and project mainly to frontal areas bound by the activity of the thalamus and cortex. Furthermore, the authors distinguished the burst mode of functioning as a promotion of sleep, and the tonic mode of functioning as a promotion of wakeful consciousness. In the latter condition, there is both thalamo-cortical and cortico-cortical synchronization, which are promote matrix neurons and are mediated by the thalamo reticular nucleus and brainstem. Within this framework, the cortex represents the site where computation takes place, thus creating the contents of consciousness. The thalamus integrates information derived from the cortical computation.

3.2.12. The Electromagnetic Field Theories

The electromagnetic field theories consider neurons densely packed in the brain, with about 10^4 neurons/mm². The basic idea of this group of theories is that when any neuron receives a signal from upstream neurons, synaptic transmitters stimulate ion pumps that cause the membrane to become negatively polarized. This massive membrane depolarization generates an electromagnetic field perturbation that influences the probability of firing of adjacent neurons, so the electromagnetic fields of

adjacent neurons are not discrete but form a complex overlapping field composed of the superposition of the fields of millions of neurons [101].

The key feature of this superposition field is its ability to integrate vast quantities of information into a single physical system. Therefore, it can account for the binding of information for consciousness.

The consciousness electromagnetic information field theory (CEMI). CEMI [32] is one of the theories that studied the above-mentioned process. Specifically, it affirmed that an electromagnetic field generated by brain activity inevitably influences the membrane's dynamics and the action of ion pumps, thus creating a self-referred "feedback loop" that can represent the physical substrate of consciousness.

The authors suggested that, normally, this superposition field is not influenced by the external electromagnetic field at all, because the high conductivity of the cerebral fluid creates an effective 'faraday cage' that insulates the brain from most natural exogenous electric fields.

In contrast to other viewpoints, the CEMI's proposal is that consciousness corresponds to the component of the brain's electromagnetic field that impacts motor activity in the widest sense [101]. As McFadden wrote: "Digital information within neurons is pooled and integrated to form an electromagnetic information field. Consciousness is the component of the brain's electromagnetic information field that is downloaded to motor neurons and is thereby capable of communicating its state to the outside world" [32]. This process can unify information and account for the unity of consciousness because a field is a continuum of information within space and time dimensions.

The authors criticized the need to use quantum mechanics to explain consciousness because the brain is not an optimal place for quantum coherence, considering the infinite number of information that should be stored in a qubit (unit of quantum information). In this sense, he suggested that our brain may be capable of implementing quantum algorithms to perform quantum computing without the need for physically unrealistic quantum states in the brain [102]. The CEMI accounts for all essential features of consciousness rather than for direct consciousness generation.

The electromagnetic field hypothesis (EM). The hypothesis of the electromagnetic field theory of consciousness by Susan Pockett [33] postulated that conscious experience is identical with certain spatiotemporal patterns in the electromagnetic field.

The theory was based on the idea that consciousness is a local, brain-generated, configuration of patterns, and the hypothesis was also confirmed by the evidence that localized electromagnetic fields are known to be capable of causing neurons to fire,

which in principle offer a mechanism by which consciousness can cause a “behavior”. Consequently, for this theory, what is important is not individual neurons but the spatiotemporal pattern of neural firing over large groups of neurons, which generates a conscious field-configuration in some circumstances but not in others (explanation of why some processes could be unconscious).

The explanation offered by the present theory is that only some configurations of the electromagnetic field have the property of consciousness, although the definitions of these characteristics have not yet been investigated. The implications of the EM theory should be analyzed considering different perspectives. Specifically, the EM supports the existence of different types of consciousness: the human consciousness probably differs from the consciousness of bats [103] because the spatiotemporal electromagnetic field-configurations generated by a bat’s brains differs from those generated by a human brain. Furthermore, it would be theoretically possible to generate specific electromagnetic fields in the total absence of neurons that could be an experimental test of the theory.

The resonator. Lewis and MacGregor [34] proposed a model of consciousness based on the exchange of energy between the mental and the physical realms. Considering the limitation of natural science, they worked on nine hypotheses, two of which are particularly important [34]. The two most important hypotheses are:

- (i)

Hypothesis 7. The neural foundations and the nature, actions, and properties of consciousness are all describable in terms of a system of resonator elements by which consciousness participates in energy exchanges with the brain, which mediate both the generation of conscious awareness and the active modulation of volitional processes in the brain. All resonator elements produce $P \rightarrow M$ transformation and engage in collective partially-free $M-M$ selections, but some (perceptual) may produce lesser or no $M \rightarrow P$ transformation”;

- (ii)

Hypothesis 8. (a) The capacity for consciousness is undergirded by the physical matter of resonator elements, which are target structures of a consciousness gene. (b) Consciousness itself is a dispositional form of energy (DE) that may relate to physical forms of energy, similar to how the phase state of a gas relates in a material way to its liquid phase. Conscious awareness and its selective release are parts of DE and under the partially-free regulation of home resonator DE advised by collective systemic DE. (c) The functional identity of conscious perceptions and feelings is intrinsically determined by the anatomical locations of neurons which either house resonators or are connected directly to non-neural cells which house resonators (the labeled line hypothesis).

Following these hypotheses, the authors introduced the concept of an energy-based resonator, intended as a brain element, that “may receive and project metabolic, traditional neural, and perhaps electromagnetic energy with traditional neural and brain elements”. The consequence of the introduction of the resonator is that the receiving and projecting connections of resonators with neural structures allows for the generation of a composite conscious–subconscious cooperative action [34].

The neural candidates for energy transformation have different locations according to resonator elements, and the theory presumes a special role for the astroglial systems and their modulators of metabolite transmission (e.g., glutamate receptors) to pyramidal cells as the conscious modulators of attention, which is in line with hypothesis number 9 (i.e., consciousness consists of dispositional energy generated by metabolic biomolecules in resonator regions of astrocytes, which undergird attention in the human brain). By merging the electromagnetic models with an astroglial effector, the total resonator’s influence on pyramidal neurons and their apical dendrites can be coupled with modulation of traditional metabolic and neural actions, as well as possible EM effects. This is the core of the theory that also offers an upgrading of consciousness levels citing the Damasio’s model [104].

3.2.13. Gelepithis’s Theory

The theory described by Gelepithis analyzes different approaches of consciousness study, highlighting the problems related to all of them [105], especially for the definition of consciousness derived by these theories. Then, Gelepithis formulated a theory based on seven definitions that compose a definitional system [35]:

Definition 1.

For a human, H , I call neural formation, N , a structure of interacting sub-cellular components across nerve cells able to influence the survival or reproduction of H .

Definition 2.

For a human, H , a neural formation is meaningful (symbol N_m), if and only if it is an N that influences the attention of that H .

Definition 3.

The meaning of a novel Sc (i.e., stimulus within its context), for the human H at time t , is whatever N_m is created by the interaction of Sc and H at time t .

Definition 4.

The meaning of a previously encountered Sc , for the human H , at time t is the prevailed N_p of N .

Definition 5.

A human, H , is conscious of an external Sc , at time t , if and only if, there are Nm structures that correspond to Sc and these structures are activated by H 's attention at that time.

Definition 6.

A human, H , is conscious of an internal Sc , at time t , if and only if, the Nm structures identified with the internal Sc are activated by H 's attention at that time.

Definition 7.

A human, H , is reflectively conscious of an internal Sc , at time t , if and only if, the N structures identified with the internal Sc are activated by H 's attention at that time and they have already been modified by H 's thinking processes activated by primary consciousness at least once.

3.2.14. The Global Workspace Theory

The Global Workspace Theory (GWT). The core idea grounding the GWT is the presence of an interaction between bottom-up and top-down attentional modulation mechanisms which, throughout a broadcasting process, allows for a specific percept/event to become conscious [106]. This idea has been developed by several authors with slight differences. Baars et al. [38,40,44] stated that several inputs enter the so-called workspace and only the more salient ones are selected and broadcasted, thus becoming conscious. Since the selection of the more salient inputs can affect the already stored internal and external representations, this process has been considered particularly important for the interaction with the environment, thus influencing selection [40,44]. Similarly, Prakash et al. [39] affirmed that some information reaches a sort of buffer and, by means of the top-down attentional amplification, they become conscious. Successively, Raffone and Pantani [41] supported the existence of two different forms of top-down attentional modulation hypothesized in the GWT, namely attention for perception and attention for access. The first is supposed to affect both the phenomenal consciousness (i.e., related to the sensory information; phenomenally conscious states that are not cognitively accessible; [107]) and the access consciousness (i.e., content information broadcasted in the global workspace). Conversely, the attention for access is supposed to only affect access consciousness. Despite the segmentation of the top-down attentional modulation for access perception and attention, the layered model hypothesized by the authors is similar to models postulated in the above-mentioned works [38,39,40]. The authors claimed that the sensory maps generate sensory representations that pass through intermediate levels (i.e., the phenomenal registry and global workspace gate) before reaching the workspace where, thanks to the so-called

consumer system, broadcasting process takes place, thus enabling a conscious experience.

As for the neural correlates of consciousness, it has been hypothesized that the key role of an extended cortico-thalamic system is the responsibility of different sensory areas [108] to bind and integrate sensory information. The prefrontal cortex has a pivotal role in the broadcast process and for access to consciousness [41]. Within this system, the thalamus has a key role in redirecting both sensory and cortical input [20]. Furthermore, the mediotemporal lobe plays a crucial role in organizing the subjective experience and episodic memory, following the theory presented by Baars et al. [44].

Besides the anatomical description of the neural correlates of consciousness, some authors supported the pivotal role of the neural firing patterns at the base of consciousness within the GWT. Specifically, Prakash et al. [39] suggested that the spatiotemporal patterns of the electromagnetic fields generated by neuronal firing can be indicative of the emergence of consciousness. Experimental evidence attested to more complex neural electromagnetic dynamics in conscious states/events compared to unconscious state/events. Similarly, Bartolomei et al. [46] considered the neural network synchronization as a key feature at the base of consciousness. It is determined by slow oscillations (theta, alpha, and beta bands) in a long-ray, as well as by fast gamma-band activity at the local level and supported by clinical evidence derived from epileptic patients with alterations of consciousness. Indeed, in this population, alterations of consciousness seem to be mainly caused by slow electrical activity in the associative cortices, which appear deactivated along with a subcortical activity during epileptic discharges. From this evidence, it is hypothesized that hyper synchronization (impairing both local and distant networks) underlies an alteration of consciousness. The neural structures mainly implicated in such hyper synchronization are represented by the corticothalamic system with cortical areas mainly involving the prefrontal and the parietal cortices [46].

Finally, among the retrieved works on the GWT, Sergent and Naccache [43] were the only ones to explicitly reference subjectivity. For the authors, this had great importance in determining what is consciously perceived. Indeed, the authors stated that if someone was conscious of something, he/she must be able to report (either verbally or not) the content of their consciousness. However, subjectivity has not been integrated within the GWT into the author's article.

The Global Neuronal Workspace hypothesis (GNW). The GNW hypothesis [45] was born from Baars idea of a model where current conscious content is represented within a distinct mental space called a global workspace, which is a structure shared and updated by many specialized modules [45]. The GNW hypothesis proposes associative perceptual, motor, attention, and memory areas interconnect to create a unified space where information is shared in a global blackboard and is sent back to lower-level processors [42]. Dehaene and Changeaux proposed, in their simulations, that this

neuronal workspace is anatomically formed by cortical pyramidal cells in layers II/III with long-range excitatory axons, particularly dense in prefrontal, cingulate, and parietal regions, together with relevant thalamocortical circuits [109].

For GNW, what we subjectively experience as consciousness is the selection, amplification, and global broadcasting of a single piece of information selected for its salience or relevance to current goals, while the rest of the neurons composing the GNW are inhibited. In other words, the hypothesis postulated that the availability of information is what we subjectively experience as a conscious state [110].

This global availability of information, which is the main core of conscious access, was hypothesized through two mechanisms. The first is the feed-forward propagation of information, during which input progresses through high-order areas in a feed-forward manner involving all probabilistic representations of the stimuli. All signals converge in high-order areas and then return to lower sensory representations that favor an increase in the power of representations compatible with current goals. The second mechanism postulates that the NMDA-mediated feedback connections (composed of GABAergic inhibitory interneurons) inhibit other neurons from amplifying the input that cross the threshold for the “ignition” and cause a self-amplification of the global state of the selected representation. According to Dehaene and Changeaux [109], “In GNW simulations, ignition manifests itself, at the cortical level, as a depolarization of layer II/III apical dendrites of pyramidal dendrites in a subset of activated GNW neurons defining the conscious contents, the rest being inhibited”. This is in order to prevent multiple ignitions.

The ignition mechanism has an important role in GNW. Some simulations demonstrated that it could fail to be triggered for very brief or low amplitude stimuli because it is not able to achieve sufficient reverberant activation during the feedforward process [111]. Moreover, several experiments with visual stimuli seem to provide robust results about the role of ignition. Indeed, some evidence seems to support that the early stages of non-conscious processing show a linear variation in activation, whereas conscious access is often characterized by a late non-linear amplification of activation through the cortex areas reported above [112,113,114]. The sharing of information creates a global workspace formed by high-level interconnected areas that work differently with respect to non-conscious stimuli, which are processed in parallel by specialized cortical processors. The global workspace allows for the explanation of when a novel operation is required, especially when a behavioral decision requires a subsequential serial strategy [115,116].

In conclusion, the GNW focuses on the well-delimited issue of how an external or internal piece of information goes beyond nonconscious processing and gains access to conscious processing, a transition characterized by the existence of a reportable subjective experience.

3.2.15. Gurwitsch's Theory

Gurwitsch accounted for the totality of conscious experience, delineating its structure and dynamics by using the phenomenological method. Using his words, Gurwitsch proposed a theory of the “articulation of the total field of consciousness and the patterns and forms in which co-present data are organized with respect to each other” [117]. According to the theory, any field of consciousness is organized according to the following three domains: theme, thematic field, and marginal consciousness.

Gurwitsch defined the theme as “that which engrosses the mind of the experiencing subject, or as it is often expressed, which stands in the ‘focus of his attention’” [117]. The thematic field has been defined as a domain of relevance. It comprises all data co-present with the theme and experienced as materially belonging and related to the theme. Finally, for the marginal consciousness, Gurwitsch claims that data in the margin are irrelevant to the theme. The only relation between marginal data and the theme is simultaneous occurrence. These are “data which, though co-present with, have no relevancy to, the theme” [117]. The author elaborated upon three distinct categories in marginal consciousness: (1) Self-awareness, which consists of awareness of one’s own inner thoughts, a stream of ideas, concepts, and inner speech running through the mind. Gurwitsch states that “self-awareness permanently and necessarily pervades all of our conscious life” [118]. The author referred to the self we are aware of in the sense of our “psychic self”. In addition to the sense of our psychic selves, we also have an ongoing sense of our bodies, corresponding to what Gurwitsch calls the corporeal or “somatic” self: “There is no moment in conscious life when we are completely unaware of our bodily posture, of the fact that we are walking, standing, sitting, lying down, etc.” [118]. Finally, insofar as we are aware of our body, we are aware of it as positioned at some location in the physical world, according to Gurwitsch: “Our body appears in experience under the perspective of the perceptual world and derives its positional index... from this horizon” [118]. Thus, we always have at least some marginal consciousness of where we are in the perceptual world: “The perceptual world has...the privilege of omnipresence” [118]. According to Gurwitsch, we pass from the theme to an item in the thematic field along a “line of relevancy”. Regarding relevance, the author wrote: “Two co-present items in the field of consciousness are relevant if they are felt to be ‘intrinsically related’ due to their ‘material contents’ transition”. In contradistinction to marginal data, items belonging to the thematic field not only present themselves along with the theme but are also experienced as intrinsically related to the theme due to the material content involved.

Yoshimi and Vinson, in 2015, proposed an extension of this theory, focusing on three points [37]:

1. Considering some results about inattentional blindness and changes in blindness, subjects sometimes fail to report seeing anything, so they hypothesized that there also exists a peripheral experience.

2. They distinguished several types of relevance, showing how one of these concepts (i.e., “causal relevance”) can be empirically tested. They introduced a new type of relevance called “predictive relevance”.
3. They developed the idea that the theme has a “variable size”, expanding and contracting, and sometimes disappearing.

Furthermore, Yoshimi [36] showed phenomenological and connectionism approaches as having the same mathematical solution for some problems, e.g., giving an interpretation of the space of possible brain state and the relationship with a possible conscious state using connectionist dynamical system approach. He described some axioms in his approach, such as the idea that the base state of a system at a time is sufficient for determining the main state of that system at that time.

3.2.16. The Representational Theories: High-Order (HOT) and First-Order (FOR) Models

The representational theories affirmed that consciousness is directly linked to “mental representations” rather than to a physical state. The main technical question related to this group of theories could be expressed as followed: What makes a mental state a conscious mental state?

The representational theories were mainly divided into two groups: FOR and HOR representationalism.

There are several varieties of HOR yet they all share common neural correlates at the level of the prefrontal (PFC) and parietal cortex, which assumed a crucial role for the creation of high representation that link different information. Conversely, the FOR performs a distinction between post-sensory structures that correspond to the same neural correlates of consciousness in the HOR, which determine the so-called “general consciousness”. Further, sensory structures correspond to specific neural sensory areas and sensory thalamic nuclei, deputing to the so-called “specific consciousness”, which corresponds to the contents of consciousness [50]. HORs, based on the thinking of Locke and Kant [119,120], contrast with FOR, which affirmed that perceptual representations formed in the sensory regions and are able to determine a conscious mental state, as reported by Dretske [121] and Michael Tye [122,123]. In other words, the HORs affirmed that consciousness depends on the supra-mental representation, while the FOR argued that the supra-mental representation must be determined by a specific content so that the sensory structures have a pivotal role in determining specific consciousness [50].

The high-order thoughts (HOT) theory, one of the most cited HOR theories, posits that consciousness awareness depends mainly on mental representation in the terms of mental states that represent oneself as being in a representational character of a first-

order representation [48]. In other words, as Rosenthal wrote, “What makes a mental state M conscious is that it is the object of some kind of higher-order mental state directed at M” [124]. To empirically sustain this position, Lau and Rosenthal [48] considered the results from different studies, showing how transcranial magnetic stimulation (TMS) applied to PFC can affect the awareness in healthy subjects without impairing their task performance [125].

At the moment, the debate about HOR and FOR is sizable, although it is important to observe the critical consequence derived from HOR, including the fact that “if PFC activity is necessary for all conscious experience, and if there is little or no PFC activity in infants and most animals, then either infants and most animals do not have conscious experience”, as was noted by Gennaro [126].

3.2.17. The Integrated Information Theory (IIT)

The IIT tries to explain the enigmas of conscious experience. Within this framework, Tononi [52] offered a mathematical definition to measure whether a system is conscious, to what degree it is conscious, and to explain consciousness itself. The IIT claims that consciousness is determined by its causal properties and it is, therefore, an intrinsic, fundamental property of any physical system. Consciousness is linked to integrated information (identified by the symbol ϕ), which represents the information generated by a system that goes beyond what can be explained by its parts working independently [51]. The IIT identified the main properties of consciousness, namely information and integration. Information is defined as the reduction of uncertainty; indeed, it is the ability to discriminate among many alternatives. Thus, the more numerous alternatives to be excluded, the greater the reduction of uncertainty, and thus the greater the information [52]. Integration allows the unity of experience, which is due to causal interaction between the elements of a system. Thus, if they are disconnected, their performance breaks down [52].

More precisely, Tononi explains both the quantity and the quality of consciousness. The first corresponds to the amount of integrated information generated by a system and it corresponds to its irreducibility, $\phi_{\max}$ (i.e., the highest value of ϕ). The second is identified by the set of informational relationships generated within that complex system [52,54]. The quality of consciousness could be geometrically represented within a space, called qualia space, in which it is possible to see the complexity of informational relationship that determines it [52]. Citing Tononi: “Qualia space (Q) is a space where each axis represents a possible state of the complex, each point is a probability distribution of its states, and arrows between points represent the informational relationships among its elements generated by causal mechanisms (connections)” [52].

The cortico-thalamic system is one of the most important NCCs for IIT since cortical damage leads to the permanent loss of consciousness. On the contrary, the cerebellum removal, even if it is full of neurons, does not considerably hurt consciousness [52]. The

IIT addresses the hard problem of consciousness by proposing a set of phenomenological axioms, ontological postulates, and identities. The axioms are as follows [15,55]:

- Existence: Consciousness exists—it is an undeniable aspect of reality. Paraphrasing Descartes: “I think, therefore, I am”.
- Composition: Consciousness is compositional (structured)—each experience consists of multiple aspects in various combinations. Within the same experience, one can see, for example, left and right, red and blue, a triangle and a square, a red triangle on the left, a blue square on the right, and so on.
- Information: consciousness is informative—each experience differs from other possible experiences. Thus, an experience of pure darkness is what it is by particularly differing from an immense number of other possible experiences. A small subset of these possible experiences includes, for example, all the frames of all possible movies.
- Integration: consciousness is integrated—each experience is (strongly) irreducible to non-interdependent components. Thus, experiencing the Italian word “SONO” (i.e., I am) written in the middle of a blank page is irreducible to an experience of the Italian word “SO” (i.e., I know) at the right border of a half-page, plus an experience of the Italian word “NO” (i.e., no) on the left border of another half page—the experience is whole. Similarly, seeing a red triangle is irreducible to seeing a triangle but no red color, plus a red patch but no triangle.
- Exclusion: consciousness is exclusive—each experience excludes all others. At any given time there is only one experience having its full content, rather than a superposition of multiple partial experiences; each experience has definite borders—certain things can be experienced while others cannot; each experience has a particular spatial and temporal grain—it flows at a particular speed and has a certain resolution such that some distinctions are possible, whereas finer or coarser distinctions are not.

The postulates are sufficient conditions concerning the requirement that each system has to satisfy to account for experience (reported from [55]):

- Existence: mechanisms in a state exist. A system is a set of mechanisms.
- Composition: elementary mechanisms can be combined into higher-order ones.
- Information: a mechanism can contribute to consciousness only if it specifies “differences that make a difference” within a system. That is, a mechanism in a state generates information only if it constrains the states of a system that can be its possible cause and effect repertoire. The more selective the possible causes and effects, the higher the cause–effect information specified by the mechanism.
- Integration: a mechanism can contribute to consciousness only if it specifies a cause–effect repertoire (information) that is irreducible to independent components. Integration/irreducibility ϕ is assessed by partitioning the mechanism and measuring what difference this makes to its cause–effect repertoire.
- Exclusion: a mechanism can contribute to consciousness at most one cause–effect repertoire, the one having the maximum value of integration/irreducibility ϕ_{Max} . This is its maximally irreducible cause–effect repertoire (MICE, or quale

sensu stricto (in the narrow sense of the word)). If MICE exist, the mechanism constitutes a concept.

Finally, identities are posited between phenomenological properties and informational/causal aspects of systems. IIT identifies the central identity as follows: “An experience is a maximally integrated conceptual information structure” [53].

The maximally irreducible conceptual structure (MICS) generated by a complex of elements is identical to its experience. The constellation of concepts of the MICS specifies the quality of the experience (its quale “sensu lato”; i.e., in the broad sense of the term). Its irreducibility Φ_{Max} specifies its quantity. The maximally irreducible cause–effect repertoire (MICE) of each concept within a MICS specifies what the concept is about (how it contributes to the quality of the experience, i.e., its quale sensu stricto, in the narrow sense of the term), while its value of irreducibility (ϕ_{Max}) specifies how much the concept is present in the experience. The experience is thus an intrinsic property of a complex of mechanisms in a state. In other words, the maximally irreducible conceptual structure specified by a complex exists intrinsically (from its own intrinsic perspective), without the need for an external observer.

3.2.18. Layered Reference Model of the Brain (LRMB)

Wang et al. developed the LRMB theory [127], which affirms the hierarchical life functions of the brain as being divided into two categories. The first one is the subconscious life functions encompassing the layers of sensation, action, memory, and perception. These functions are provided by layers 1 to 4, which are inherited, fixed, and relatively matured when a person is born. The conscious functions include the layers of metacognition, inference, and cognitive functions sustained by layer 5–7. These functions are acquired, highly plastic, and intentionally controlled based on willingness, goals, and motivations [58]. In the model, the thalamus and cerebellum are considered key structures for consciousness. The role of the cerebellum in the cognitive and logical approach to consciousness was analyzed by the author who declared that it has a functional role of conscious status memory (CSM), which is a “new type of memory identified in the brain, it is supplementary to other types of memories such as the short-term, long-term, sensory buffer, and action-buffer memory” [127,128,129,130]. The thalamus, instead, was considered as a processor, empathizing its role of switching center through the number of connections to almost all parts of the brain such as the cerebral cortex, eyes, and visual cortex [131,132,133,134].

Analyzing the mathematical, cognitive, behavioral, computational, and structural models of consciousness, Wang et al. described the cognitive process of consciousness using denotational mathematics. They affirmed that a conscious state is a function (fc) that maps a Cartesian product of entire events (E) and the current status of conscious status memory (CSM) into an updated state of the memory in CSM. Entire events in an individual’s brain can be classified as external stimuli (S) and/or internal motivations (M),

$E = S \cup M$. Thus, external and internal sub-functions of consciousness (fc-e and fc-i) can be expressed using individual or combinatorial expressions.

3.2.19. The Memory Consciousness and Temporality Theory (MCTT)

Dalla Barba and Boissè [59] redefined the concept of consciousness in the framework of the MCTT. Specifically, they claimed that consciousness is always related to a certain object. In other words, when someone is conscious, he/she is always conscious of something. Furthermore, they distinguished between temporal and the knowing consciousness. The latter refers to specific features of an object; consequently, there are many knowing consciousnesses that refer to the same object. Temporal consciousness, instead, encompasses three other different sub-categories of consciousness. The first one is represented by past consciousness, which corresponds to the remembering of something. This is strictly related to the self; it is not a piece of generic semantic information related to the past. The second category is represented by present consciousness, which is different from other constructs, such as the “objective present” and the “psychological present”, as postulated by William James. Present consciousness encompasses all the present events, perceptions, and sensations related to the self. The third component is represented by future consciousness, which encompasses the future possibility of being (taking into consideration the past and the present). The three components altogether (i.e., the temporal consciousness) make it possible for the temporal existence for the subject. Furthermore, the temporal consciousness allows one to perceive the uniqueness of an object against its multiplicity, which is, in contrast, related to knowing consciousness (e.g., I perceive one specific pen as mine rather than a generic object belonging to the category “pen”). For what concerns the neural correlates of consciousness, the authors claimed a key role of the mediotemporal lobe and its related structures in determining the temporal consciousness. However, their hypothesis was developed to better frame the confabulatory behavior of patients suffering from brain damages.

3.2.20. The Mesocircuit Hypothesis

In 2010, Schiff formally described the so-called “mesocircuit hypothesis”, which considers the thalamus (and globus pallidus) as the key structure for conscious states due to its widespread direct and indirect connections with both brainstem and frontal areas [60]. The author notes that in normal conditions medium spiny neurons (MSN) in the striatum inhibit the globus pallidus interna, thus allowing the thalamus to communicate with the frontal brain areas, which, in turn, project to the striatum. If the inhibitory mechanism of the striatum ceased, the globus pallidus interna inhibit the thalamus, thus shutting down the necessary connections and activity needed to sustain a conscious state. Consequently, the entire network would appear dysregulated, and, without thalamic input, the frontal areas would not be able to communicate and regulate the striatum activity anymore. This theory accounted for both evidence from patients with disorders of consciousness following thalamic damages (see for example [135]), from neuromodulation studies [136,137,138] and the mechanisms underlying the

paradoxical effect of the Zolpidem in restoring the level of consciousness, thanks to its inhibitory action on the globus pallidus interna.

3.2.21. The Min's Model

In 2010, Min defined consciousness as “a mental state embodied through thalamic reticular nucleus (TRN)-modulated synchronization of thalamocortical networks” [61]. More specifically, it “consists of each mental unit, which is an individual thalamocortical looping mechanism, no matter what cognitive stages it involves” [61]. According to the present theory, the TRN could act as a mediator for the signals derived from the other thalamic nuclei to the cortex and vice versa, exercising in this way a sort of “control” over the thalamic activity in a feed-forward fashion. Specifically, the neural synchronization is considered as a prerequisite to neural signal association, and it seems to be at the base of conscious events. In physiological terms, the synchronization takes place thanks to the inhibitory GABAergic neurons, which allow for the detection of classic gamma range signal (>30 Hz) (in general, the gamma range signal indicates the presence of information processing within a network). Because the TRN is equipped with GABAergic neurons, it is plausible that it is responsible for the synchronization of the cortical and thalamic neural activities given the rise of information processing and, hence, consciousness. In other words, the author's hypothesis considered that the emergence of consciousness due to the thalamocortical activation above a specific threshold level could determine the synchronization of neural activity. Thus, it is predicted that when the system does not reach the threshold level of activation, the sensory stimuli simply pass through the thalamus without becoming conscious. Furthermore, the contents of consciousness are supposed to be determined by the degree and the distribution of the cortical activity. Indeed, given the connections between the TRN and the cortex, the presence of an attentional top-down modulation operated by the prefrontal cortex can be postulated. Specifically, the attention could act as a signal amplifier whilst it eliminates unnecessary information. At the same time, the working memory retains the elements which are processed during the so-called “attended conscious awareness” [61]. Even if the attentional process is strictly linked to the prefrontal cortex' activity, this theory assumes the pivotal role of TRN in triggering prefrontal cortex activity. However, the attentional process is considered optional to the so-called “conscious awareness” and insufficient for the emergence of consciousness. The theory takes into account a background condition to be conscious, which is represented by arousal. Arousal is indeed considered a necessary condition for consciousness and, even in this case, it is assumed that it serves pivotal role for TRN due to its connections with the brainstem [61]. Finally, given the TRN' modular structure (i.e., different sensory-modality maps exist within the TRN), the emergence of a unitary consciousness awareness is explained via a characterization of shaggyiness across the boundaries of the receptive fields of the TRN neurons [61].

3.2.22. The Network Inhibition Hypothesis (NIH)

The NIH aims to explain consciousness from the evidence reported in epilepsy studies [5]. The idea is that the so-called “conscious system” is analogous to other systems

(e.g., motor, limbic, etc.) by involving different cortical and sub-cortical structures and that consciousness system must include neural areas that govern functions such as attention (i.e., a prerequisite for consciousness), awareness, and an alert state. In this sense, the medial thalamus, upper brain stem, hypothalamus, interhemispheric regions (including medial frontal cortex, cingulate cortex, and precuneus), lateral frontal, and temporal-parietal associative cortices seem to be important structures involved in the loss of consciousness due to the absence of seizure. Therefore, they could have a role in the appearance of consciousness [5]. The authors observed that the high-frequency neural activities in the temporal lobe during seizures reach the midline sub-cortical neurons and are correlated to an inhibition of frontal and parietal activities with a decrease in cerebral blood flow (CBF), as shown by studies that used single photon emission computed tomography (SPECT; [139,140]). The NIH hypothesized that loss of consciousness is a secondary effect due to the dysregulation of the networks described above and that focal limbic seizures propagate subcortical structures involved in the arousal systems. This could be correlated due to decreased activity in the fronto-parietal association cortices [141] caused by an increase in effective inhibition, or a decrease in effective excitation, within these networks [141].

The NIH affirmed that the upper brainstem interacts with the cerebral cortex to maintain wakefulness. However, during seizures, if they remain confined in the mesial temporal lobe, then a simple-partial seizure occurs without impairment of consciousness. In the case of seizure activity from the temporal lobe to midline subcortical structures, the propagation of the contralateral temporal lobe is possible. Finally, the NIH hypothesized that temporal lobe seizures inhibit the subcortical arousal systems leading to activity decrement in the bilateral fronto-parietal associative cortex, causing a deep sleep-like activity that is followed by the loss of consciousness.

3.2.23. O'Doherty's Theory

In O'Doherty's view, "Consciousness represents the storage of past events for use in future situations and it is altered by external experience of the organism", resulting "from the gradual evolutionary development of the human information processing function" [62].

The framework adopted by the author encompasses evolution, learning, and behavioral theories. Hence, consciousness can be conceptualized as the product of an interaction between an individual and the environment, rather than something located inside the individual himself/herself. The author's attempt is to overtake both dualistic and monistic notions of consciousness. Furthermore, language is considered a precursor of consciousness in an evolutionary fashion.

In other words, consciousness could be the result of an evolutionary process involving both learning mechanisms and information processing, which must be conscious (different learning and information processing exist that take place without awareness).

Furthermore, the language is supposed to be involved in coherently connecting the experiences with their outcomes, so that language allows the categorization of past events and the access to them. Within this view, memory is considered a prerequisite for language development, as it allows for information storage. Finally, qualia are considered as a key feature of consciousness since an individual cannot be conscious in the absence of an experience/stimulation and its storage. Furthermore, the experience of qualia can be modified through language (which builds links between different stored information) and the interaction of an individual with other individuals and the environment (as it happens, for instance, in sensory illusions).

3.2.24. Passive-Frame Theory (PFT)

The PFT, proposed in 2015 by Morsella et al., represents a framework within which consciousness can be viewed as a phenomenon strictly related to voluntary actions. Indeed, the authors defined consciousness as “a phenomenon associated with perceptual-like processing and interfacing with the somatic nervous system” [63]. In other words, consciousness determines the skeletal muscle outputs to produce adaptive behavior when different motor plans are available and in conflict with each other. This conflict arises due to the awareness of different contents (i.e., what is experienced in a given moment), which requires specific actions to reach the goal. The sum of the conscious contents represents the conscious field, which is continuously updated (through the frame-check process) to select the most adaptive skeletomotor plan in that moment.

As for the neural correlates of consciousness, the authors considered different evidence, concluding that both the sensorium hypothesis [142,143] supporting the pivotal role of perceptual regions of the brain along with subcortical structures, and the hypothesis supporting the pivotal role of cortical circuits, are valid. To disentangle the role of cortical regions and perceptual regions, they suggest future research study simple neural systems such as the olfactory one.

The theory did not analyze the subjectivity dimension which corresponds to the first-person perspective in the authors' view.

3.2.25. A Psychological Theory of Consciousness (PToC)

Shannon proposed a theory from a psychological and phenomenological point of view [64]. He defines consciousness as a multifaceted phenomenon that includes different forms that be sorted into clusters. Originally, Shannon identified three types of consciousness (identifying consciousness as the region of the internal subjective experience), hierarchically ordered and interdependent among them. They form a well-integrated system, which is defined as a “tripartite system” [64].

The first type of consciousness is the Cons1, the “sensed being or sentience”. It concerns the primitive and elementary aspect of consciousness that distinguishes living from non-living organisms. It has no specific context or structure, it is pervasive, and it is present all our life [144,145]. The Cons2, called “mental awareness”, is typical of higher-order mammals and is related to subjective experience, as well as to the contents and forms of those experiences, such as mental images, ideations, flows of consciousness, and internal verbal monologues. The Cons3, called “meta-mentation”, is the mental ability to take its own productions as object for further reflection. This type of ability has different manifestation, such as meta-observation, reflection, monitoring, and control. The Cons2 and Cons3 are interrelated and someone can go up a level (from 1 to 2) without stepping down onto level 3 [64].

The system of consciousness encloses different regions and each one presents different levels: There is the self (self 1, 2, 3) concerning personal identity, the world (world 1, 2, 3), which is inherent in the relationship between knowledge and world and, finally, the temporality (temp 1, 2, 3) that concerns the temporal part of the experience.

Shannon [64] has identified other non-ordinary levels of consciousness, which are considered “mystical experiences”, namely Cons 4 and 5. Cons 4 involves internal subjective experiences that the person does not feel like the product of his/her own mind (i.e., hallucinations); Cons 5 is considered by Shannon as “the highest state of mind”, according to which the inner subjective experience leads the subject to experience things as coming from something supreme. It is a phenomenon very similar to nirvana and to the bliss that implicates the transcendence of the self and thus becomes ineffable [64].

This theory also provides a structural analysis of consciousness, through the identification of a series of parameters. The first parameter concerns the sense of perception, as well as sensory information recording and perception and how they are considered pertinent to reality; the parameter of the sense of meaningfulness concerns the meaning attributed to the perceptual experience, while the aesthetic sense parameter allows giving organization and composition to experienced things.

Moreover, Shannon [64] described five main features of consciousness: the quality of sentience; having experiences; making experiences the object of reflections and mentations; orientation and disposition towards others; the eventual capacity for transcendence.

The author specified that consciousness is a dynamic system subjected to different operations. The first processes are differentiation and crystallization (from which the different levels of consciousness specified above arise). The process of distancing allows separation from the sentient being between one’s inner world and the surrounding environment and, therefore, permits a distinction between subject and

object. This process is more evident towards the highest level of consciousness, even if it tends to decrease in Cons 4 and 5.

For what concerns the internalization, Shannon [64] stated that consciousness creates a mental domain in which the cognizer may operate even when actual action in the real world is not feasible. Moreover, Shannon described the reflexivization as a dynamic mental operation through which the products of mentation, in turn, generate other mentations; externalization, instead, concerns the ability of the mind to create objects that are not part of reality, even if they are experienced as real, e.g., hallucinations. Finally, PToC reported the semantic operations, which are defined as “semantic composition or narrativization, the poesis, metaphoricity, and enhanced creativity encountered in the psychedelic state, the interplay between losing oneself and concentrated control achieved in masterly skilled performance, role playing, and the adoption of modified identity as in the metamorphoses” [64].

3.2.26. Q-Theories

In this section, we summarize the different theories based on hypotheses that implied quantum mechanisms to explain the appearance of consciousness.

The four-dimensional Einstein. Sieb [74] affirmed that conscious experience oriented in space and time is fundamental for a coherent conscious experience. Considering that there are three dimensions of space (length, width, height; when one person analyzes the experience of a room where one is reading this paper, there are many events like light, walls, windows, etc. separated in time) and one dimension of time (e.g., moving the mouse of a PC we observe a series of events separated in time and having cause–effect and past–future relationships), conscious experience can be said to have four dimensions.

Taking into account that space–time is any mathematical model that combines space and time into a single continuum [146], the author considered the notion of space cells [147,148] and time cells of the hippocampus, following the MacDonald idea according to where place cells and time cells reflect fundamental mechanisms by which hippocampal neural networks parse any spatiotemporal context into quantal units of where and when important events occur, thereby organizing elements in a conceptual manner [149].

The authors suggested that the posterior parietal cortex (PPC) and prefrontal cortex (PFC) may extract space-time interval inequality information from the hippocampus in a process of relational integration [74,150]. The authors considered that the sensory neurons involved in this representation of the information, whenever present in the environment, are the same that maintain the information in working memory. In this case, the role of PFC not only contributes to working memory but, instead, manages attentional focus to select information and allows for the executive functioning to control cognitive processes.

In this macro system, quantum mechanical description is close to the classical description, and the authors suggested that conscious experience is when quantum systems and classical systems agree. In other words, when we represent a quantum system in conscious experience, we represent the system as a large (macroscopic) system wherein it is possible to use classical physics to describe it [74]. For example, the authors reported that our visual system transforms a quantum mechanical feature into the classical physics of conscious experience, so that wave-particle duality is reduced to classical physics in conscious experience.

Koehler's mathematical approach. Koehler's work was devoted to analyzing the possibility of checking experimentally regarding whether the perceptual process can lead to the collapse of the wave function. The author demonstrated that it is possible to describe the quantum collapse using a new Clifford algebra, obtained by the assignment of a numerical value in $A(Si)$ by our cognition, which determines new commutation rules [68].

The idea is that a mathematical proof of the collapse of the wave function could be given every time we observed a direct cognitive attribution of a numerical value to the basic element of the $A(Si)$. In other words, the author affirmed that the quantum collapse is a transition from a superimposed linear dynamic to a new dynamic in which a semantic act (the direct involvement of information, relating mental entities whose consciousness is the basics representative) is involved. For example, the act of seeing starts when the image is formed on the retina, an image which is not considered as a simple photodetector but, according to this new approach, as part of the brain that allows the perceptual and cognitive processes to be directly involved in the analysis of information following the role of quantum mechanics [68].

A Holoinformational Model of Consciousness. Di Biase [151] proposed a quantum-informational holographic model of brain–consciousness–universe interactions based on the holonomic neural networks of Karl Pribram [152] (found in the holographic quantum theory developed by David Bohm [153]). Pribram developed a neural wave equation, demonstrating that the cerebral cortex is the site of a holographic information process called a multiplex neural hologram, which is dependent on local circuits of neurons without long fibers that do not transmit ordinary nervous impulses. “These neurons function in the undulatory mode and are above all responsible by the horizontal layer connections of the neural tissue where holographic interference patterns can be built” [154]. Therefore, the author's model affirmed that the “quantum holographic brain dynamic patterns are conceived as an active part of the universal quantum-holographic informational field, and capable of generating an informational field interconnection that is simultaneously nonlocal (quantum-holistic) and local (Newtonian-mechanistic), i.e., holoinformational”. In this conception is the holoinformational non-local flux that permits the interaction of holonomic informational quantum brain dynamics—this theory considers the viewpoint of Eccles [155] extended dendrons with the quantum-holographic nature of the universe.

The orchestrated objective reduction (Orch OR) theory. The Orch OR theory [71] is grounded on quantum physics. Penrose noted that one of the fundamental questions in the study of consciousness is “how it is that any physical structure whatsoever can give rise to the puzzling phenomenon of awareness?” [156]. The theory, in contrast to strong artificial intelligence positions, which argue that all mental processes can be reduced to computational models, affirmed that consciousness depends on processes of the general nature of quantum computations that occur in the brain.

Using the argument of Gödel’s theorem [157], Penrose supports the view that consciousness is not a feature of computational activity, thus taking into consideration the mathematical possibility that non-computational activity could be the base for the appearance of consciousness [158].

Consequently, Stuart Hameroff and Sir Roger Penrose developed a new theory called “orchestrated objective reduction” (‘Orch OR’), in which the main actors are (i) the “objective reduction” (OR) and (ii) the brain microtubules as anatomical structures in which the quantum computation occurred.

The idea starts from the measurement problem derived from the two fundamental procedures in quantum physics. The first concerns the deterministic evolution of a quantum state linked to the Schrödinger equation [159] whilst the second is related to the collapse of the wave function, where probabilistic rules seem to be involved when a measure of a quantum state in a system is needed (measurement problem). Penrose supposed that an objective form of quantum state reduction (OR) could be a solution of this problem, hypothesizing that consciousness could be dependent by quantum computation present in the brain that terminated OR in some form.

As reported by the author, the “orchestrated objective reduction is a theory which proposes that consciousness consists of a sequence of discrete events, each being a moment of ‘objective reduction’ of a quantum state” (according to the Diósi–Penrose objective reduction scheme, where it is taken that these quantum states exist as parts of a quantum computations) [160]. This is primarily carried out on in neuronal structures. In Orch OR, consciousness is a collapsed, self-organized process on the edge between quantum and classical realms as reported by the authors [71].

The authors sought out structures in the brain that support such putative non-computational actions at the borderline between quantum and classical physics, which could plausibly have important influence on brain activity. In this sense, Hameroff suggested the role of microtubules (i.e., polymers with cylindrical forms of around 25 nanometers of diameters, a length of a few hundred nanometers, probably up to meters in long nerve axons) as one of the major components of the structural skeleton prevalent in neurons.

The theory affirmed that for OR to be operative in the brain, a super position of sufficient amounts of material accounting for gravitational self-energy is needed, undisturbed by environmental entanglement, where this reduction occurs following the above OR scheme in a specific time scale in line with conscious experience. Pragmatically, authors affirmed that a coherent quantum super position/computation evolves during the integration phases in integrate-and-fire brain neurons (the most logical place is supposed to be the post-synaptic dendrites and somata). This causes an increase in quantum super position until the threshold is met at time $\tau \approx \hbar/EG$ (where τ is the time until OR occurs and $\hbar$ is the reduced Planck constant), when a conscious moment occurs. Each OR selects microtubules' output states that govern axonal firings and regulate synapses. Synaptic inputs 'orchestrate' these quantum computations, from which they derived the name "orchestrated objective reduction" (Orch OR), a process that, in a progressive large scale, could be the base for consciousness.

The Quantum no go theorems and consciousness. Georgiev's viewpoint is related to two problems intrinsically linked to theories of consciousness [6]. The first consists of the fact that the material world in classical deterministic physical theories is causally closed [161]. In other words, if brain states produce conscious experiences, then these experiences cannot possibly have an effect upon brain dynamics, which is already fully determined by the fundamental physical quantities of the brain such as mass, charge, length, and time.

The second problem is represented by the illusion that consciousness is able to make a choice between alternatives as a "free" system, since the deterministic frameworks make it an illusion produced by our subjective brain [162].

The author considered that consciousness should work in a real short time window, and that decoherence time for quantum states within cytoskeletal proteins was estimated to be around 0.1 picosecond [163]. This could be coherent with consciousness process, considering that minds operate on a millisecond timescale and that a theory of consciousness consider this issue.

Moreover, the author provided mathematical proofs on the first two no-go theorems [164,165], positing that they invalidate Frankfurt's celebrated argument on the principle of alternate possibilities [166,167]. Georgiev's main idea is related to the fact that "because there is no measurement of q-bit A that would allow one to distinguish unequivocally between any pair of non-orthogonal states, it follows that unambiguous determination of an individual quantum state jWA_i of a q-bit A is impossible by only measuring the q-bit A" and subsequently that "an unknown quantum state jWA_i of a q-bit A cannot be cloned to another q-bit B". This implies that quantum theories of mind are the best candidates to explain consciousness, rather than the classical determinist theories of mind, as they avoid some paradoxes, such as the possibility to create perfect copies of one's mind (and brain), the non-existence of self-identity, and Frankfurt's thesis stating that free will exists in cases where the subject could not have

done otherwise, guaranteeing the privacy of the mental state and our free will in relation to the presence of multiple choices [6].

The single particle consciousness hypothesis. Argonov [69] studied the unity of consciousness as one of the most important problems when approaching a consciousness theory. How spatiality and time unities to form stream of consciousness is a main topic. The author does not support the idea of macroscopic coherency as a collective quantum phenomenon working like a single super-particle [158], as well as the idea that microscopic intracellular collective quantum effects are more realistic than the macroscopic ones. Specifically, the author suggested that each electron has its consciousness, following the idea of Bohm and Hiley who said that “in some sense a rudimentary mind-like quality is present even at the level of particle physics. As we go to subtler levels this mind-like quality becomes stronger and more developed” [153].

The fundamental issues of this hypothesis are as follows: “Each electron (or fermions) is the subjective ‘observer’ of its quantum dynamics (energy, momentum, ‘shape’ of wave function). Each electron ‘feels’ its quantum dynamics as ‘own’ subjective sensations and volition” [69]. Most of the electrons (or fermions) in the universe have a primitive consciousness. However, some particles in biological cells (especially in the brain) have complex consciousness due to complex dynamics in complex organic environment. Animals are hierarchical structures of particles. Some chemically active electrons in animal brains are “on the top” of the hierarchy. Their dynamics are directly influenced by sensory data and have a direct influence over animal behavior. The author called them “pontifical particles (PPs)”.

These electrons work following synchronized dynamics in one brain and they “feel” the same performance and volitional act. The human brain might have many observers that share a similar “human” mind.

Argonov supposed that PPs or pontifical syncytium (i.e., networks of neuron connected, with common cytoplasm and many nuclei) [168] may be placed in voltage-gated ion channels (chemically active electrons in carboxyl groups of amino-acid molecules), which play a functional role in axonal signal transmission involved in different activities, such as thinking.

The three-layer model. The idea of this model is that there are different kinds of consciousness sustained by different neural correlates [73]. The proposed model considers communication among preconscious states, from maps to a single consciousness [169], postulating independent and distributed neural structures at three hierarchical levels: “The lowest Level 1, which may include subliminal and preconscious nodes. The nodes in the next higher Levels, 2 and 3, do not possess specific neuronal correlates and there is aggregation of several of the lowest consciousness and the unitary consciousness is defined at the highest Level 3.” Authors suggested that the

“nature of the conscious experience would depend on what other nodes are accessible to the node at Level 3” [170]. The mind selects memories that emerge from any conscious state and are based on the stream of previous conscious information. Thus, states of consciousness depend on the degree to which preconscious and memory states are accessible to awareness [73]. The mediated variable of this hierarchical model are languages and metalanguages that work at Levels 1 and 2, whereas the highest node is not located in a physical space, but still allows for the unity of consciousness. Nonlocality, entanglement, and coherent behavior are required to model memories as quantum objects [171], which is an explanation for long range correlations inside the brain and in behavior. It has been proposed that the cognitive system, through its evolution, veils the nonlocal processes of the quantum interactions amongst consciousness states by creating classical narrative for them, suggesting that the entropy associated with the consciousness states is a projection of three-dimensional reality.

Timeless and spaceless. Li’s theory [70] focuses on a pre-space/time dimension starting from 4 principles. (1) “The quantum superposition principle is assumed to be universally correct; (2) any physical property, that is accessible by physical means, is extrinsic, and it is nothing more and nothing less than the relations of the entity to the rest of the world; (3) consciousness can feel or memorize the past largely determined the configuration of the emergent time; (4) equivalent principle is applicable to consciousness”.

The author introduced a super quantum state called Dao, which “contains everything but tells nothing” (similar to other theories of Bohm [172], Mensky [173] and Wheller-Dewitt [174]) but in which time and space are absent. Moreover, if the Dao is a composite number, then it is decomposed as an entangled state of a composite system, generating a smaller quantum state. When Dao is separated into two subsystems (consciousness and universe), time emerged, whereas space is defined in the further creation of subsystems from universe.

The four principles guided the mathematical demonstration of the theory, suggesting that the universe is completely determined and that the relative distance and mass (a negative mass for consciousness was hypothesized) could be defined by the entanglement entropies among physical entities separated from a pure quantum state.

The Conscious Agent Thesis. Hoffman and Prakash [72] focused their attention on the relationship between perception and reality, citing the interface theory of perception [175]. They presented several premises to introduce their model. The authors stated that (i) in evolutionary biology and psychological perspectives, the natural selection has shaped our perception to be an accurate depiction of reality especially of those aspects useful for our survival; (ii) veridical perceptions strategies tuned to the true structure of the world are routinely dominated by non-veridical strategies tuned to fitness. Therefore, strategy favored by selection is best thought as a windows interface of a PC instead of

as a window on truth. Consequently, the color and shape of an icon for a text file do not entail that the text file itself has a color or shape, so our perceptions of space–time and objects do not entail that objective reality has the structure of space–time and objects [176,177,178]. Considering these perspectives, authors proposed a rigorous formalism called a conscious agent thesis.

According to the authors' perspective, consciousness involves three processes: perception, decision, and action. As for perception, "A conscious agent interacts with the world (W) and, in consequence, has conscious experiences. In the process of decision, a conscious agent chooses what actions to take based on the conscious experiences it has. In the process of action, the conscious agent interacts with the world in light of the decision it has taken and affects the state of the world".

They claim that every property of consciousness can be represented by some property of a conscious agent or system of interacting conscious agents. Then, the authors defined a conscious agent as a six tuple $C = ((X,X), (G,G), P,D, A,N)$, where: (X, X) and (G, G) are measurable spaces; $P: W \times X \rightarrow [0, 1]$, $D: X \times G \rightarrow [0, 1]$, $A: G \times W \rightarrow [0, 1]$ are mathematical formalism of Markovian kernels, and N is an integer.

Regarding the explanation of the generation of a macro-subject in terms of assembling micro-subjects [179], i.e., combination problems, the theory of the conscious agent affirmed that it is necessary to discuss conditional probability of having a specific conscious experience, considering that most of our mental processes are unconscious processes [180]. Then, the theory of the conscious agent provides two ways to combine conscious agents, i.e., undirected combinations and directed combinations, giving a formal solution for the combination problem of conscious experiences explained through theorems. Finally, regarding the theorization of the object (the previous parts were described as a theory of the subject), the authors described the idea that "space–time and objects are among the symbols that conscious agents employ to represent the properties and interactions of conscious agents" [181]. Specifically, they observed that "the harmonic functions of the space–time chain that is associated with the dynamics of a system of conscious agents are identical to the wave function of a free particle", identifying a formal system and how it works. This part characterizes the conscious agent thesis. The authors affirmed that "one particular object, the quantum free particle, has a wave function that is identical in form to the harmonic functions that characterize the asymptotic dynamics of conscious agents; particles are vibrations not of strings but of interacting conscious agents". Consequently, this allows one to reinterpret physical properties such as position, momentum, and energy as properties of interacting conscious agents, rather than as pre-existing physical truths.

3.2.27. Reji Kumar's Theory

Reji Kumar suggested a model where consciousness is the result of the information processing taking place in the mind which, in turn, consists of information accepting,

processing, and generating [76]. The information process was assumed to be similar among individuals (even if some variations can exist in internal processing); however, what subjectively changes is the information accepted. Furthermore, the author assumes that if the information's entry into the mind is blocked, consciousness does not exist [76].

The present theory is grounded on four different axioms: (i) A model represents a piece of information; (ii) the mind compares and classifies different models; (iii) new models can be created starting from the existing models; (iv) the mind attributes a value/preference to each model.

The above-mentioned models consist of three different parameters. The α -models concern the sensory information deriving from the outside world and processed by the mind (e.g., an object in the external environment). They consist in pieces of sensory information that can be processed even without consciousness, and they are also called "pre-linguistic models" [76]. The β -models allow one to express one's own experience of the external world (e.g., the name of the object in the external environment) and they consist of a one-to-one correspondence [76]. The γ -models are more complex and are useful to explain the experience of mind (e.g., a story or a poem; [76]). In other words, the α -models are supposed to contain raw data, whilst the β - and γ -models are deputed to raw data processing [78]. Furthermore, the models are characterized by a time component that is apparent in models happening in different instances (this means that simultaneous models have the same time component), and a space component which concerns difference in corresponding models [79].

In 2016, the author enlarged his theory by distinguishing the information from the knowledge. Whilst the information corresponds to data deriving from the outside world, the knowledge consists of forms that such information assumes [78]. Furthermore, the author attributed a pivotal role to memory by assuming the existence of temporary memory, which receives information from both the outside world and an individual's inside world, as well as a permanent memory where information is stored after their processing and integration according to the α -, β -, and γ - models [78].

Moreover, in the first conceptualization of this theory, the author outlined subjectivity as a core component of consciousness [76]. The concept has been further developed by the author. The model could be associated with different realities, thus producing ambiguity and confusion [79]. In these cases, it is necessary to perform a choice by selecting some items among many others in an arbitrary way. This idea has been mathematically developed by the author as follows [79]. Given the individual's models of consciousness ($M \in CX$), the subjectivity is a function ($f: CX \rightarrow [-1, 1]$) according to which the individual can like ($f(M) > 0$) or dislike a model ($f(M) < 0$). If the function value is equal to 0, the model must be considered objective. Furthermore, subjectivity changes over time and can be computed in different ways by considering the minimum/maximum of the values of the subjectivity of component models ($f_{\min}(M) = \min(f(M_i))$), and by

averaging the subjectivity of all component models ($fave(M) = \sum_i M_i/N$, where N represents the number of sub-models of M). Similarly, the stability of subjectivity over time can be computed via the following formula ($S = (x + b) - (x - b)$), which can range from 0 (maximum stability) to 1 (maximum instability).

The central role of subjectivity in this theory has been supported by other authors who stated that subjectivity arises when preferences are associated with a specific model [77]; in other words, subjectivity acts when individuals choose among models. The authors assumed that, within the α -models, there is no subjectivity, as these models represent an information registry where stimuli processing is unconscious. When information is associated with an entity within β -models, the perception is subjective and therefore conscious. Similarly, in the γ -models, additional subjective attention is required to perform information-entity association. This is different than the β -models, which allow the association of single information with multiple entities [77].

3.2.28. The Radical Plasticity Thesis (RPT)

In 2007, Cleeremans defined consciousness as “the brain’s theory about itself, gained through experience interacting with the world, and, crucially, with itself” [81].

The author proposed the radical plasticity thesis of consciousness by considering subjective experience as a core component. Experience is defined by the author as “something that takes place not in any physical entity but rather only in special physical entities, namely cognitive agents” [81]. Indeed, the author stated that whenever experience is experienced, something happens in the organism (such as experiencing emotions, retrieving memories, learning something new, and so on) and the organism itself registers its state. Metarepresentations play a pivotal role in this. The individual creates some representations based on the external input, and the metarepresentations inform the individual of his internal status by considering the previously created representations. Subjective experience occurs only if a certain system (i.e., individual) learns its representations. Consequently, consciousness is the awareness of both external and internal status.

Representations are always associated with the activation of neural networks, which can vary across different dimensions, such as the stability of time (i.e., how long a representation can be maintained active; prefrontal cortex involvement), strength (i.e., how many units are involved in a certain representation and how strong is their activation), and distinctiveness (i.e., how different a certain representation is from others; it is a measure inversely related to the degree of overlap among representations, and it characterizes the hippocampal representations compared to the cortical ones). Moreover, the representations can be different considering the extent to which they can influence behavior, the availability to control, and the availability to subjective experience. As such, three different forms of representation are described. The implicit representation is characterized by weakness so that they are out of the

individual awareness and, as a consequence, out of control. However, this kind of representation can influence behavior. The second kind of representation is the explicit one characterized by enough strength to be available to individual awareness and thus to the control. They can influence behavior as long as the individual is involved in controlling them. The subjective experience is typical of this kind of representation. The third kind of representation is the automatic one characterized by high strength, which impedes these representations to be controlled. However, differently from the implicit representations, the individual is fully aware of this kind of representation. Within this framework, the availability of the representations to consciousness strongly depends on the representations' quality; furthermore, the learning process can produce higher quality representations in time (this mechanism is at the base of the development of adaptive behavior). In keeping with this, the role of consciousness is to control those representations strongly enough to affect the behavior but not strongly enough to be considered automatic adaptive representations. From here it derives that (i) consciousness involves the subjective experience along with the control dimension; (ii) the availability to consciousness correlates with the quality of representations; (iii) the development of high-quality representations takes time; and (iv) consciousness allows flexible and adaptive control over behavior.

The metarepresentations have the same features as the representations (strength, stability, and distinctiveness). Furthermore, they have specific functions, allowing for individuals to communicate their status and to make predictions. Cleeremans et al. provided empirical evidence by means of simulation experiments attesting how a higher-order network can be trained to observe the internal states of another network and use this information to perform tasks required to know the internal structure of such internal states [80]. In this way, the core idea of the radical plasticity thesis is that “the brain continuously and unconsciously learns not only about the external world, but about its own representations of it. The result of this unconscious learning is conscious experience, in virtue of the fact that each representational state is now accompanied by (unconscious learned) metarepresentations that convey the mental attitude with which these first-order representations are held” [81].

3.2.29. The Semantic Pointer Competition Theory of Consciousness (SPC)

The phenomenon of consciousness has been studied by the SPC through three different hypotheses. The first one supposes that consciousness is the result of a process within the brain that springs from neural mechanisms [82]. Thus, this hypothesis starts from the assumption that all conscious organisms possess a brain, and therefore complexes of neural processes represent a basic condition for consciousness to arise. Accordingly, consciousness is present in all entities capable of introspection, self-evaluation, and implementing behaviors that are indicative of consciousness itself.

The second hypothesis identifies representation by patterns of firing in neural populations, binding of these representations into semantic pointers, and competition among semantic pointers” [82]. The representations are created thanks to neuronal patterns of activation following the interaction within the environment. Consequently, representations of different natures (e.g., perceptive, motor, verbal) can be created. These kinds of representations are called “semantic pointers” [182,183,184], which, after a binding mechanism, can be combined into a single semantic indicator with symbolic function [185]. As for the neural correlates of consciousness, the present theory does not refer to specific brain areas, but rather considers the pivotal role of cellular systems activation in consciousness.

The third hypothesis assumes that consciousness derives from an interactive competition between semantic pointers after which the winning indicator determines the qualitative aspects of experience (i.e., qualia). The SPC addresses the subjectivity by assuming the existence of different qualitative experiences due to the linkage between different neural representations and semantic indicators. According to the authors, “Semantic pointers bind together neural representations of a situation, physiological changes, and cognitive appraisals to produce a combined representation” [82].

In addition to this qualitative experience, the SPC specifically addresses other issues concerning the main aspects of consciousness, namely onset and cessation, shifts in experience, different kinds of consciousness, unity and disunity of consciousness, and, finally, storage and retrieval. All these aspects are mathematically explained through computer simulations. Starting from the first aspect (i.e., the cessation of consciousness, such as in the case of sleep, anesthesia, stroke, etc.), it is linked to three mechanisms, which depend on the balance between neural excitation and inhibition. Neural inhibition involves a decrease in the number of semantic pointers. Conversely, in the case of epilepsy and convulsions in which neural hyperactivity occurs, the interruption of consciousness is due to the saturation of the semantic pointers network, which interferes with the mechanism of competition that no single semantic pointer can overcome, especially considering the threshold necessary to generate consciousness [82].

Regarding these kinds of consciousness, we must distinguish between a minimal form of consciousness linked to the “here and now” and a form of consciousness characterized by greater complexity that allows for self-representation [186]. The different kinds of consciousness depend on differences in complexity and the extent of the three mechanisms that generate experiences.

The unity of consciousness concerns the reasons why we experience in an integrated way experience and is made possible by the combination of the different mental representations and, therefore, by the association of the semantic pointers. On the contrary, this disunity occurs during the mechanism of competition, where there is no

semantic pointer that prevails over the others other than different “winning” semantic pointers [82].

The SPC emphasizes the interactions between consciousness and memory since, during an experience of an event, similar experiences are recalled. This is determined by the storage and retrieval capacity of each semantic pointer [82].

4. Discussion

From the 68 analyzed articles, the present systematic scoping review found 29 theories of consciousness that show heterogeneous perspectives. The theories with the highest number of articles focused on them were quantum theories of consciousness (which represent a category including theories that refer to a quantum mechanism to explain consciousness), the IIT, and the GWT. The articles describing these theories were found in almost all of the considered years. This result was predictable considering the number of focus reviews published in the last five years on these theories (Hinterberger, 2015; Owen, 2019; Zhao et al., 2019). However, the total number of articles found is not directly correlated to the level of analysis of each theory on the themes we focused on. Indeed, different theories obtained the maximum scores in at least one dimension analyzed (in alphabetical order: ADT, Agnati et al.’s theory, ART, Bieberich’s theory, COI, CP, CSS, DCT, GWT, HOT/FOR, IIT, LRMB, Min’s model, Rejikumar’s theory, RPT) independently from the number of articles found (among the above-mentioned theories, ADT, Bieberich’s theory, COI, CP, DCT, LRMB, and Min’s model were represented by a single article).

4.1. Descriptive Analysis of the Results for Each Dimension

Subjectivity, NCC, and consciousness and cognitive functions were the most debated in the articles we found (i.e., dimensions with the lowest number of 0 scores during the assessment), showing the highest grade of variability.

The NCC dimension was the most studied among the others (the highest number of articles ($n = 30$; see [supplementary materials](#) for details) obtained a score of 4 or 5 in the NCC dimension during the assessment). Among the neural correlates of consciousness reported in [Table 3](#), the thalamus, basal ganglia, and the hippocampus were the most reported sub-cortical structures, whereas the cingulate, prefrontal, and temporal areas were the most reported cortical ones. In general, the thalamo-cortical system was the macro system most described in the literature as the system able to sustain consciousness (intended as the neuronal mechanisms jointly sufficient for conscious percept), especially when consciousness is related to representation. However, our results highlighted the widest range of proposals. Consciousness was associated with different structures from a single electron to different brain areas. This is an important point that should be analyzed in the future, especially when studying the relationship between NCC and the “type” of consciousness they refer to. As recently pointed out by other authors [187], there could be a potential convergence among

theories on the neural structures involved in the mechanisms that can explain consciousness and its alterations. Therefore, macro-perspectives are needed in the next future.

The dimensions called “consciousness and translation into clinical practice” and “subjectivity” were the two dimensions that obtained the lowest number of 4 and 5 scores during the assessment (3 and 6 articles respectively); in other words, they represent the dimensions with the fewest number of articles focused on them. In the first dimension, we analyzed the translation of theoretical models into procedures able to help patients with disorders of consciousness. The lack of findings in this dimension is likely due to the difficulty in translating the hypothesis made by the authors into clinical practice (i.e., definition of prognostic or diagnostic markers, ad hoc treatments, etc.). An example of these difficulties is represented by the challenges in developing technical systems able to help researchers in testing their hypothesis. Indeed, considering the hypothesis made by authors of the quantum theories of consciousness, it sounds difficult to test how a quantistic phenomenon can be observed/reproduced in a “wet” material such as the brain, or how we can experimentally modify the activity of specific cell populations (e.g., the glia), or to analyze the effect of vibrations present in some neural microstructures on human consciousness. However, this does not imply that this perspective should not be studied by a new generation of researchers able to match neurological and quantistic viewpoints. This reflects a real challenge for researchers and clinicians. Future research programs should be developed to foster new approaches that can help patients with disorders of consciousness through new therapeutic tools (e.g., neural stimulation programs) in parallel to new measures of consciousness (13 articles obtained a score of 4 or 5 in the Quantitative measures of consciousness dimension). This would allow for quantitative and qualitative evaluations of the outcomes.

Finally, we found 11 articles that obtained a score of 4 or 5 for the association between consciousness and other cognitive functions. As in the NCC dimension, we found interesting perspectives that ranged from the complete overlap of consciousness with other functions (e.g., memory and attention in the learning process) [22,23] to the definition of consciousness as a sort of metacognition in layered models (see [58]), which is debated in some articles that considered metacognition equivalent to consciousness (published after the time range of this review) [188,189].

4.2. A Definition of Consciousness: Problems and Perspectives

The number of articles that propose a definition of consciousness was quite high and almost all theories reported a main definition, offering several arguments and interpretation of the results that should be analyzed in the future. Indeed, the sentences analyzed, such as “consciousness is [...]”, were supported by many interpretations—i.e., defining consciousness as “being aware of” something or as the “subject experience”. We also found definitions of “awareness” and “experience”, and, of course, some clear statements on what “awareness” and “experience” is not. However, we

found that four terms—“(subjective) experience”, “information”, “processes”, and “(mental) state”—were used in most of the analyzed articles ([Table 2](#)) to define consciousness. These terms imply several reasons regarding their nature/structure and the ontology behind them. For example, “information” is a term that has been debated in the scientific literature over the last few decades [[190,191](#)]. New researchers can find many references sustaining a specific ontology linked to this term [[192](#)]. Although everyone can affirm to have an idea of what “experience” is, it is difficult to find a clear definition of what it is not. What experience is it? What is the structure of an experience? Is consciousness synonymous with experience or do we have to think about a tailored association between the structure of experience and the structure of consciousness? Some theories attempted to answer these questions, but the debate is still open and thus these questions still represent issues for the science of consciousness that researchers will face in the future.

We also found many new definitions and new subcategorizations of consciousness, such as “knowing consciousness” [[59](#)], “unreflected consciousness” [[65](#)], or “meta-mentation” [[64](#)], as reported in [Table 1](#). These definitions constitute a new way to develop ad hoc perspectives and measurable concepts. Our impression is that authors are developing new models in which consciousness is not a monolithic meaning and can be sub-divided into different components (working hierarchically or in parallel), offering a test in experimental settings whenever possible.

4.3. Limitations

This scoping review does have limitations. Firstly, during the full-text analysis, we opted for a very strict selection of the articles to be included. Consequently, one of the articles’ criteria for selection was consciousness as the main object of an article. This rigid criterion implied that raters excluded records on topics related to other arguments such as body-representation, bodily-self, etc., or describing the elaboration of conscious vs. unconscious stimuli using fMRI or EEG paradigms (e.g., the existence of unconscious working memory, memory processes linked to subliminal stimuli only, etc.). Moreover, we excluded general theories that identified basic rules for life generation that explain consciousness indirectly, as well as papers that did not report clearly the original underlying theory.

We acknowledge that all of these arguments are linked to consciousness models and that our article selection process could have excluded works that contributed to the study of consciousness but, considering the difficulty of analyzing each collected model, we decided to limit our work to articles completely related to a consciousness model itself as the main topic, analyzing papers on these arguments only if they were linked specifically to a general theory of consciousness. The choice to include articles using strict inclusion criteria could have caused a loss of information, but we preferred to make this choice in order to offer a well-structured overview of the topic without diverging from the main aim of this scoping review. Another limitation was that we analyzed only the text of the articles included in the review found in the databases

reported in the methodological section. As a matter of fact, details on some theories could be described in other papers published before the years considered in the present work or published in other online databases. Consequently, some missing information about either each dimension or definitions of consciousness may be due to the adopted methodology rather than to a real lack of information (this is the main reason why we inserted the [Table 1](#) note “nf” (i.e., not found)). Moreover, the choice to search articles in “medical” literature databases could have caused the loss of important works marked by the conceptual (philosophical) perspectives on consciousness simply because they are not contained in those repositories. For this reason, we avoided distinguishing articles included as “conceptual” or “empirical” because many works contained both parts, as shown by the dimensional model analysis. Therefore, the limitation on the medical dataset and time range deployed in the search strategy determined that several articles that we knew to be important for a debate on consciousness models were not included in our analysis. There is an increasing amount of literature in computer science on the relationships between intelligence and consciousness (e.g., Drigas et al. presented a new layered model of consciousness that combines the dominant theories of intelligence adapted to recognized models of telecommunications and computer science [[188,189](#)]), affirming that cognitive processes contribute to the representation of knowledge. Some important metacognitive components are transformed into superior levels of consciousness (see [[193](#)]). Another example is provided by other studies that sought to define a precise formal definition of a conscious Turing machine, also called a conscious AI, [[194,195](#)]. We recognize that the analysis of a limited set of databases represents a limitation but we hope that this paper can be a useful tool for researchers to increase the debate on the frameworks underlying the various theories of consciousness, as noted by other authors (e.g., [[187](#)]). It could also be considered as a general framework from which the readers can start to analyze some arguments that are only cited in this review.

Finally, considering the different areas of study of the authors who wrote the analyzed articles, as well as the difficulty derived from the use of different technical languages, the probability that there are some theoretical aspects insufficiently treated in the summaries is high. To comply with the formatting guidelines, we had to limit the amount of information in the description of each theory, and this may have resulted in the above-mentioned limitations. Moreover, the Likert score attributed to each article in relation to the evaluation on how each dimension was debated in the text should not be intended as a qualitative score of the theories. For example, an article can debate many dimensions, such as the neural correlates that offer less “clear” messages on the structures involved in consciousness rather than an article that explicitly debated this point. The aim of this review was not to compare the different theories found or to discuss them in a critical manner (debating for one hypothesis or another one) but to synthesize them to provide a general overview on models that explain one of the most difficult topics known to humanity. Considering the limits of our work, this overview, overall, can be useful for readers and future researchers both to reflect on the variabilities that exist on the various topics related to consciousness in the scientific literature and to develop new ideas and collaborative research groups starting from the theories presented.

5. Conclusions

In conclusion, we found 29 theoretical models that described different perspectives on consciousness. Among the dimensions analyzed, the NCC was the most reported in the full-text collected and we found a very heterogeneous set of neural correlates of consciousness, ranging from the single electron to the whole brain. Moreover, we found fewer articles that focused on the development or implementation of new tools/strategy/markers for clinical practice in settings where consciousness levels were altered. Regarding the definition of consciousness, the review highlighted that the words “information”, “experience”, and “state” were often reported within definitions, and that several new terms/labels related to sub-categorization of consciousness (more than twenty terms) were also proposed. The results of this scoping review allows us to present the state-of-the-art research on consciousness studies and can offer a global vision on this issue for researchers all over the world. Because consciousness is one of the most ancient questions considered by human beings, this accomplishment is no small feat.

Acknowledgments

The present publication was submitted on behalf of the Coma Research Centre (CRC) multidisciplinary team and the Insubria University PhD program. The authors thank all raters involved in the study. A special thanks goes to Rosalind Hendricks for the language revision of the manuscript and to Saba Motta and Alessandra Atterrato for their help in the development of the search strategies for each database, and for the document delivery service.

Supplementary Materials

The following are available online at <https://www.mdpi.com/article/10.3390/brainsci11050535/s1>. Table S1: Search strategy; Table S2: Maximum score obtained in each dimension by the theories analyzed; Table S3: Secondary definitions and definitions of related terms and features of consciousness. Table S4: Abstracts of each theory analyzed. [Click here for additional data file.](#) (401KB, zip)

Author Contributions

Conceptualization: D.S., M.P., and M.L.; methodology: D.S. and A.V.F.; formal analyses: F.G.M., M.C. (Milena Caputo), and L.B.; writing—original draft preparation: D.S. and F.G.M.; writing—review and editing: M.P., M.L., A.V.F., M.C. (Milena Caputo), L.B., and M.C. (Martina Cacciatore); supervision: D.S., M.L., and M.P. All authors have read and agreed to the published version of the manuscript.

Funding

This study derived from the project CRC, “Coma Research Centre”, supported by Grant No. IX/000407—05/08/2010 awarded by Regione Lombardia (Italy), from the PhD program supported by Insubria University (Varese, Lombardia) and the Italian Ministry of Health (STRIVE project-RF/GR-2016-02365049).

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

This is a review article. Data available at <https://doi.org/10.5281/zenodo.4671955>.

Conflicts of Interest

The authors declare no conflict of interest.

Footnotes

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- 1. Sommerhoff G. Consciousness explained as an internal integrating system. *J. Conscious. Stud.* 1996;3:139–157. [[Google Scholar](#)]
- 2. Monzavi M., Murad M.H.S.A., Rahn timer M., Shamshirband S. Historical path of traditional and modern idea of ‘conscious universe’. *Qual. Quant.* 2016;51:1183–1195. doi: 10.1007/s11135-016-0324-3. [[DOI](#)] [[Google Scholar](#)]
- 3. Capra F. *The Tao of Physics*. Shambhala Publications; Boston, MA, USA: 1974. [[Google Scholar](#)]
- 4. Clark A. *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford University Press; Oxford, UK: 2008. [[Google Scholar](#)]
- 5. Yu L., Blumenfeld H. Theories of Impaired Consciousness in Epilepsy. *Ann. N. Y. Acad. Sci.* 2009;1157:48–60. doi: 10.1111/j.1749-6632.2009.04472.x. [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 6. Georgiev D. Quantum No-Go Theorems and Consciousness. *Axiomathes.* 2013;23:683–695. doi: 10.1007/s10516-012-9204-1. [[DOI](#)] [[Google Scholar](#)]

- 7.Hardie W.F.R. Concepts of Consciousness in Aristotle. *Mind*. 1976;85:388–411. doi: 10.1093/mind/LXXXV.339.388. [\[DOI\]](#) [\[Google Scholar\]](#)
- 8.Crick F., Koch C. Towards a neurobiological theory of consciousness. *Semin. Neurosci*. 1990;2:263–275. [\[Google Scholar\]](#)
- 9.Koch C. The Quest for Consciousness: A Neurobiological Approach. *Eng. Sci*. 2004;67:28–34. [\[Google Scholar\]](#)
- 10.Moher D., Liberati A., Tetzlaff J., Altman D.G., The PRISMA Group Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med*. 2009;6:e1000097. doi: 10.1371/journal.pmed.1000097. [\[DOI\]](#) [\[PMC free article\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
- 11.Aru J., Bachmann T., Singer W., Melloni L. Distilling the neural correlates of consciousness. *Neurosci. Biobehav. Rev*. 2012;36:737–746. doi: 10.1016/j.neubiorev.2011.12.003. [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
- 12.Facco E., Lucangeli D., Tressoldi P. On the Science of Consciousness: Epistemological Reflections and Clinical Implications. *Explore*. 2017;13:163–180. doi: 10.1016/j.explore.2017.02.007. [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
- 13.Fekete T., van Leeuwen C., Edelman S. System, Subsystem, Hive: Boundary Problems in Computational Theories of Consciousness. *Front. Psychol*. 2016;7:1041. doi: 10.3389/fpsyg.2016.01041. [\[DOI\]](#) [\[PMC free article\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
- 14.Lee U., Blain-Moraes S., Mashour G.A. Assessing levels of consciousness with symbolic analysis. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci*. 2015;373:20140117. doi: 10.1098/rsta.2014.0117. [\[DOI\]](#) [\[PMC free article\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
- 15.Tononi G., Koch C. Consciousness: Here, there and everywhere? *Philos. Trans. R. Soc. B Biol. Sci*. 2015;370:20140167. doi: 10.1098/rstb.2014.0167. [\[DOI\]](#) [\[PMC free article\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
- 16.Koch C., Massimini M., Boly M., Tononi G. Neural correlates of consciousness: Progress and problems. *Nat. Rev. Neurosci*. 2016;17:307–321. doi: 10.1038/nrn.2016.22. [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
- 17.Cohen M.A., Dennett D.C. Consciousness cannot be separated from function. *Trends Cogn. Sci*. 2011;15:358–364. doi: 10.1016/j.tics.2011.06.008. [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)

- 18.LaBerge D., Kasevich R. The apical dendrite theory of consciousness. *Neural Netw.* 2007;20:1004–1020. doi: 10.1016/j.neunet.2007.09.006. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 19.Cook N. The neuron-level phenomena underlying cognition and consciousness: Synaptic activity and the action potential. *Neuroscience.* 2008;153:556–570. doi: 10.1016/j.neuroscience.2008.02.042. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 20.Agnati L.F., Guidolin D., Cortelli P., Genedani S., Cela-Conde C., Fuxe K. Neuronal correlates to consciousness. The “Hall of Mirrors” metaphor describing consciousness as an epiphenomenon of multiple dynamic mosaics of cortical functional modules. *Brain Res.* 2012;1476:3–21. doi: 10.1016/j.brainres.2012.01.003. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 21.Hobson J.A. REM sleep and dreaming: Towards a theory of protoconsciousness. *Nat. Rev. Neurosci.* 2009;10:803–814. doi: 10.1038/nrn2716. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 22.Grossberg S. Consciousness CLEARS the mind. *Neural Netw.* 2007;20:1040–1053. doi: 10.1016/j.neunet.2007.09.014. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 23.Grossberg S. Towards solving the hard problem of consciousness: The varieties of brain resonances and the conscious experiences that they support. *Neural Netw.* 2017;87:38–95. doi: 10.1016/j.neunet.2016.11.003. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 24.Graziano M.S.A., Kastner S. Awareness as a perceptual model of attention. *Cogn. Neurosci.* 2011;2:125–127. doi: 10.1080/17588928.2011.585237. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 25.Bieberich E. Introduction to the Fractality Principle of Consciousness and the Sentyon Postulate. *Cogn. Comput.* 2012;4:13–28. doi: 10.1007/s12559-011-9104-5. [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 26.Kriegel U. A cross-order integration hypothesis for the neural correlate of consciousness. *Conscious. Cogn.* 2007;16:897–912. doi: 10.1016/j.concog.2007.02.001. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 27.Merker B. Consciousness without a cerebral cortex: A challenge for neuroscience and medicine. *Behav. Brain Sci.* 2007;30:63–81. doi: 10.1017/S0140525X07000891. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]

- 28.Eberkovich-Ohana A., Eglicksohn J. The consciousness state space (CSS)—A unifying model for consciousness and self. *Front. Psychol.* 2014;5:341. doi: 10.3389/fpsyg.2014.00341. [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 29.Berkovich-Ohana A., Wittmann M. A typology of altered states according to the consciousness state space (CSS) model: A special reference to subjective time. *J. Conscious. Stud.* 2017;24:37–61. [[Google Scholar](#)]
- 30.Bosse T., Jonker C.M., Treur J. Formalisation of Damasio's theory of emotion, feeling and core consciousness. *Conscious. Cogn.* 2008;17:94–113. doi: 10.1016/j.concog.2007.06.006. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 31.Ward L.M. The thalamic dynamic core theory of conscious experience. *Conscious. Cogn.* 2011;20:464–486. doi: 10.1016/j.concog.2011.01.007. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 32.McFadden J. Conscious Electromagnetic (CEMI) Field Theory. *NeuroQuantology.* 2007;5:262–270. doi: 10.14704/nq.2007.5.3.135. [[DOI](#)] [[Google Scholar](#)]
- 33.Pockett S. Difficulties with the Electromagnetic Field Theory of Consciousness: An Update. *NeuroQuantology.* 2007;5:271–275. doi: 10.14704/nq.2007.5.3.136. [[DOI](#)] [[Google Scholar](#)]
- 34.Lewis E.R., MacGregor R.J. A Natural Science Approach to Consciousness. *J. Integr. Neurosci.* 2010;9:153–191. doi: 10.1142/S0219635210002202. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 35.Gelepithis P.A.M. A Novel Theory of Consciousness. *Int. J. Mach. Conscious.* 2014;6:125–139. doi: 10.1142/S1793843014400150. [[DOI](#)] [[Google Scholar](#)]
- 36.Yoshimi J.K. Phenomenology and Connectionism. *Front. Psychol.* 2011;2:288. doi: 10.3389/fpsyg.2011.00288. [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 37.Yoshimi J., Vinson D.W. Extending Gurwitsch's field theory of consciousness. *Conscious. Cogn.* 2015;34:104–123. doi: 10.1016/j.concog.2015.03.017. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 38.Baars B.J., Franklin S. An architectural model of conscious and unconscious brain functions: Global Workspace Theory and IDA. *Neural Netw.* 2007;20:955–961. doi: 10.1016/j.neunet.2007.09.013. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]

- 39. Prakash R., Prakash O., Prakash S., Abhishek P., Gandotra S. Global workspace model of consciousness and its electromagnetic correlates. *Ann. Indian Acad. Neurol.* 2008;11:146–153. doi: 10.4103/0972-2327.42933. [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 40. Baars B.J., Franklin S. Consciousness is Computational: The Lida Model of Global Workspace Theory. *Int. J. Mach. Conscious.* 2009;1:23–32. doi: 10.1142/S1793843009000050. [[DOI](#)] [[Google Scholar](#)]
- 41. Raffone A., Pantani M. A global workspace model for phenomenal and access consciousness. *Conscious. Cogn.* 2010;19:580–596. doi: 10.1016/j.concog.2010.03.013. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 42. Dehaene S., Changeux J.-P., Naccache L. The global neuronal workspace model of conscious access: From neuronal architectures to clinical applications. *Res. Perspect. Neurosci.* 2011 doi: 10.1007/978-3-642-18015-6_4. [[DOI](#)] [[Google Scholar](#)]
- 43. Sergent C., Naccache L. Imaging neural signatures of consciousness: “What”, “When”, “Where” and “How” does it work? *Arch. Ital. Biol.* 2012;150:91–106. doi: 10.4449/aib.v150i2.1270. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 44. Baars B.J., Franklin S., Ramsøy T.Z. Global Workspace Dynamics: Cortical “Binding and Propagation” Enables Conscious Contents. *Front. Psychol.* 2013;4:200. doi: 10.3389/fpsyg.2013.00200. [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 45. Dehaene S., Charles L., King J.-R., Marti S. Toward a computational theory of conscious processing. *Curr. Opin. Neurobiol.* 2014;25:76–84. doi: 10.1016/j.conb.2013.12.005. [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 46. Bartolomei F., McGonigal A., Naccache L. Alteration of consciousness in focal epilepsy: The global workspace alteration theory. *Epilepsy Behav.* 2014;30:17–23. doi: 10.1016/j.yebeh.2013.09.012. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 47. Lau H.C. A higher order Bayesian decision theory of consciousness. *Prog. Brain Res.* 2007;168:35–48. doi: 10.1016/S0079-6123(07)68004-2. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 48. Lau H., Rosenthal D. Empirical support for higher-order theories of conscious awareness. *Trends Cogn. Sci.* 2011;15:365–373. doi: 10.1016/j.tics.2011.05.009. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]

- 49.Friesen L. Higher-Order Thoughts and the Unity of Consciousness. *J. Mind Behav.* 2014;35:201–224. [[Google Scholar](#)]
- 50.Mehta N., Mashour G.A. General and specific consciousness: A first-order representationalist approach. *Front. Psychol.* 2013;4:407. doi: 10.3389/fpsyg.2013.00407. [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 51.Balduzzi D., Tononi G. Integrated Information in Discrete Dynamical Systems: Motivation and Theoretical Framework. *PLoS Comput. Biol.* 2008;4:e1000091. doi: 10.1371/journal.pcbi.1000091. [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 52.Tononi G. Consciousness as Integrated Information: A Provisional Manifesto. *Biol. Bull.* 2008;215:216–242. doi: 10.2307/25470707. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 53.Tononi G. Integrated information theory of consciousness: An updated account. *Arch. Ital. Biol.* 2012;150:293–329. [[PubMed](#)] [[Google Scholar](#)]
- 54.Casali A.G., Gosseries O., Rosanova M., Boly M., Sarasso S., Casali K.R., Casarotto S., Bruno M.-A., Laureys S., Tononi G., et al. A theoretically based index of consciousness independent of sensory processing and behavior. *Sci. Transl. Med.* 2013;5 doi: 10.1126/scitranslmed.3006294. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 55.Oizumi M., Albantakis L., Tononi G. From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0. *PLoS Comput. Biol.* 2014;10:e1003588. doi: 10.1371/journal.pcbi.1003588. [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 56.Tononi G., Boly M., Massimini M., Koch C. Integrated information theory: From consciousness to its physical substrate. *Nat. Rev. Neurosci.* 2016;17:450–461. doi: 10.1038/nrn.2016.44. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 57.Tsuchiya N., Taguchi S., Saigo H. Using category theory to assess the relationship between consciousness and integrated information theory. *Neurosci. Res.* 2016 doi: 10.1016/j.neures.2015.12.007. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 58.Wang Y. The Cognitive Mechanisms and Formal Models of Consciousness. *Int. J. Cogn. Inform. Nat. Intell.* 2012;6:23–40. doi: 10.4018/jcini.2012040102. [[DOI](#)] [[Google Scholar](#)]

- 59.Barba G.D., Boisse M.-F. Temporal consciousness and confabulation: Is the medial temporal lobe “temporal”? *Cogn. Neuropsychiatry*. 2010;15:95–117. doi: 10.1080/13546800902758017. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 60.Schiff N.D. Recovery of consciousness after brain injury: A mesocircuit hypothesis. *Trends Neurosci*. 2010;33:1–9. doi: 10.1016/j.tins.2009.11.002. [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 61.Min B.-K. A thalamic reticular networking model of consciousness. *Theor. Biol. Med. Model*. 2010;7:10. doi: 10.1186/1742-4682-7-10. [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 62.O’Doherty F. A Contribution to Understanding Consciousness: Qualia as Phenotype. *Biosemiotics*. 2012;6:191–203. doi: 10.1007/s12304-012-9140-x. [[DOI](#)] [[Google Scholar](#)]
- 63.Morsella E., Godwin C.A., Jantz T.K., Krieger S.C., Gazzaley A. Homing in on consciousness in the nervous system: An action-based synthesis. *Behav. Brain Sci*. 2015;39:e168. doi: 10.1017/S0140525X15000643. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 64.Shannon B. A Psychological Theory of Consciousness. *J. Conscious. Stud*. 2008;15:5–47. [[Google Scholar](#)]
- 65.Klemm D.E., Klink W.H. Consciousness and Quantum Mechanics: Opting from Alternatives. *Zygon*. 2008;43:307–327. doi: 10.1111/j.1467-9744.2008.00919.x. [[DOI](#)] [[Google Scholar](#)]
- 66.Das T. Theory of Consciousness. *NeuroQuantology*. 2009;7 doi: 10.14704/nq.2009.7.2.237. [[DOI](#)] [[Google Scholar](#)]
- 67.Di Biase F. Quantum-Holographic Informational Consciousness. *NeuroQuantology*. 2009;7:657–664. doi: 10.14704/nq.2009.7.4.259. [[DOI](#)] [[Google Scholar](#)]
- 68.Koehler G. Q-consciousness: Where is the flow? *Nonlinear Dyn. Psychol. Life Sci*. 2011;15:335–357. [[PubMed](#)] [[Google Scholar](#)]
- 69.Argonov V.Y. Neural correlate of consciousness in a single electron: Radical answer to “quantum theories of consciousness”. *NeuroQuantology*. 2012;10:276–285. doi: 10.14704/nq.2012.10.2.548. [[DOI](#)] [[Google Scholar](#)]
- 70.Li J. A Timeless and Spaceless Quantum Theory of Consciousness. *NeuroQuantology*. 2013;11 doi: 10.14704/nq.2013.11.3.684. [[DOI](#)] [[Google Scholar](#)]

- 71.Hameroff S., Penrose R. Consciousness in the universe: A review of the “Orch OR” theory. *Phys. Life Rev.* 2014;11:39–78. doi: 10.1016/j.plrev.2013.08.002. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 72.Hoffman D.D., Prakash C. Objects of consciousness. *Front. Psychol.* 2014;5:577. doi: 10.3389/fpsyg.2014.00577. [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 73.Kak A., Gautam A., Kak S. A Three-Layered Model for Consciousness States. *NeuroQuantology.* 2016;14:166–174. doi: 10.14704/nq.2016.14.2.935. [[DOI](#)] [[Google Scholar](#)]
- 74.Sieb R.A. Human Conscious Experience is Four-Dimensional and has a Neural Correlate Modeled by Einstein’s Special Theory of Relativity. *NeuroQuantology.* 2016;14:630–644. doi: 10.14704/nq.2016.14.3.983. [[DOI](#)] [[Google Scholar](#)]
- 75.Brabant O. More Than Meets the Eye: Toward a Post-Materialist Model of Consciousness. *Explore.* 2016;12:347–354. doi: 10.1016/j.explore.2016.06.006. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 76.Reji Kumar K. Modeling of Consciousness: Classification of Models. *Adv. Stud. Biol.* 2010;2:141–146. [[Google Scholar](#)]
- 77.Ahmad F., Khan Q. Can mathematical cognition formulate consciousness?; Proceedings of the ICOSST 2012—International Conference on Open Source Systems and Technologies; Lahore, Pakistan. 20–22 December 2012; pp. 7–11. [[Google Scholar](#)]
- 78.Reji Kumar K. Mathematical modeling of consciousness: A foundation for information processing; Proceedings of the IEEE International Conference on Emerging Technological Trends in Computing, Communications and Electrical Engineering, ICETT 2016; Turku, Finland. 24–26 August 2016; Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers Inc.; 2016. [[Google Scholar](#)]
- 79.Reji Kumar K. Mathematical modeling of consciousness: Subjectivity of mind; Proceedings of the IEEE International Conference on Circuit, Power and Computing Technologies, ICCPCT 2016; Nagercoil, India. 18–19 March 2016; Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers Inc.; 2016. [[Google Scholar](#)]
- 80.Cleeremans A., Timmermans B., Pasquali A. Consciousness and metarepresentation: A computational sketch. *Neural Netw.* 2007;20:1032–1039. doi: 10.1016/j.neunet.2007.09.011. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]

- 81.Cleeremans A. Consciousness: The radical plasticity thesis. *Prog. Brain Res.* 2007;168:19–33. doi: 10.1016/s0079-6123(07)68003-0. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 82.Thagard P., Stewart T.C. Two theories of consciousness: Semantic pointer competition vs. information integration. *Conscious. Cogn.* 2014;30:73–90. doi: 10.1016/j.concog.2014.07.001. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 83.Mountcastle V.B. Modality and Topographic Properties of Single Neurons of Cat's Somatic Sensory Cortex. *J. Neurophysiol.* 1957;20:408–434. doi: 10.1152/jn.1957.20.4.408. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 84.Sevush S. Single-neuron theory of consciousness. *J. Theor. Biol.* 2006;238:704–725. doi: 10.1016/j.jtbi.2005.06.018. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 85.Edelman G.M. *Bright Air, Brilliant Fire: On The Matter Of The Mind.* Basic Books; New York, NY, USA: 1992. [[Google Scholar](#)]
- 86.Tononi G., Edelman G. Consciousness and the integration of information in the brain. In: Jasper H.H., Descarries L., Castellucci V.F., Rossignol S., editors. *Consciousness: At the Frontiers of Neuroscience.* Lippencott-Raven; Philadelphia, PA, USA: 1998. pp. 245–279. [[PubMed](#)] [[Google Scholar](#)]
- 87.Pompeiano O. The neurophysiological mechanisms of the postural and motor events during desynchronized sleep. *Res. Publ. Assoc. Res. Nerv. Ment. Dis.* 1967;45:351–423. [[PubMed](#)] [[Google Scholar](#)]
- 88.Aston-Jones G., Bloom F. Activity of norepinephrine-containing locus coeruleus neurons in behaving rats anticipates fluctuations in the sleep-waking cycle. *J. Neurosci.* 1981;1:876–886. doi: 10.1523/JNEUROSCI.01-08-00876.1981. [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 89.Hobson J.A., McCarley R.W. The brain as a dream state generator: An activation-synthesis hypothesis of the dream process. *Am. J. Psychiatry.* 1977;134:1335–1348. doi: 10.1176/ajp.134.12.1335. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 90.Baumeister R.F., Masicampo E.J. Conscious thought is for facilitating social and cultural interactions: How mental simulations serve the animal–culture interface. *Psychol. Rev.* 2010;117:945–971. doi: 10.1037/a0019393. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]

- 91.Carruthers P. How we know our own minds: The relationship between mindreading and metacognition. *Behav. Brain Sci.* 2009;32:121–138. doi: 10.1017/S0140525X09000545. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 92.Baars B.J. A Cognitive Theory of Consciousness. Cambridge University Press; Cambridge, UK: 1993. [[Google Scholar](#)]
- 93.Rosenthal D.M. A Theory of consciousness. In: Block N., Flanagan O.J., Guzeldere G., editors. *The Nature of Consciousness*. MIT Press; Cambridge, MA, USA: 1997. ZIF Technical Report 40. [[Google Scholar](#)]
- 94.Rosenthal D.M. Explaining consciousness. In: Chalmers D.J., editor. *Philosophy of Mind: Classical and Contemporary Readings*. Oxford University Press; Oxford, UK: 2002. pp. 406–421. [[Google Scholar](#)]
- 95.Treisman A., Schmidt H. Illusory conjunctions in the perception of objects. *Cogn. Psychol.* 1982;14:107–141. doi: 10.1016/0010-0285(82)90006-8. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 96.Penfield W., Jasper H. *Epilepsy and the Functional Anatomy of the Human Brain*. Little Brown & Co; Boston, MA, USA: 1954. [[Google Scholar](#)]
- 97.Damasio A. *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. Houghton Mifflin Harcourt; New York, NY, USA: 1999. [[Google Scholar](#)]
- 98.Jonker C.M., Treur J. Compositional Verification of Multi-Agent Systems: A Formal Analysis of Pro-activeness and Reactiveness. *Int. J. Coop. Inf. Syst.* 1998;11:51–91. doi: 10.1142/S0218843002000480. [[DOI](#)] [[Google Scholar](#)]
- 99.Edelman G., Tononi G. *A Universe of Consciousness*. Basic Books; New York, NY, USA: 2000. [[Google Scholar](#)]
- 100.Jones E.G. Thalamic circuitry and thalamocortical synchrony. *Philos. Trans. R. Soc. B Biol. Sci.* 2002;357:1659–1673. doi: 10.1098/rstb.2002.1168. [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 101.McFadden J. The Conscious Electromagnetic Information (Cemi) Field Theory: The Hard Problem Made Easy? *J. Conscious. Stud.* 2002;9:45–60. [[Google Scholar](#)]
- 102.McFadden J. Synchronous Firing and Its Influence on the Brain's Electromagnetic Field. *J. Conscious. Stud.* 2002;9:23–50. [[Google Scholar](#)]
- 103.Nagel T. What Is It Like to Be a Bat? *Philos. Rev.* 1974;83:435. doi: 10.2307/2183914. [[DOI](#)] [[Google Scholar](#)]

- 104.Damasio A.R. The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Philos. Trans. R. Soc. B Biol. Sci.* 1996;351:1413–1420. doi: 10.1098/rstb.1996.0125. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 105.Gelepithis P.A.M. A concise comparison of selected studies of consciousness. *Cogn. Syst.* 2001;5:373–392. [[Google Scholar](#)]
- 106.Baars B.J. In the theatre of consciousness. Global workspace theory, a rigorous scientific theory of consciousness. *J. Conscious. Stud.* 1997;4:292–309. [[Google Scholar](#)]
- 107.Block N. Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behav. Brain Sci.* 2007;30:481–499. doi: 10.1017/S0140525X07002786. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 108.Baars B.J., Ramsøy T.Z., Laureys S. Brain, conscious experience and the observing self. *Trends Neurosci.* 2003;26:671–675. doi: 10.1016/j.tins.2003.09.015. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 109.Dehaene S., Changeux J.-P. Experimental and Theoretical Approaches to Conscious Processing. *Neuron.* 2011;70:200–227. doi: 10.1016/j.neuron.2011.03.018. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 110.Dehaene S. Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition.* 2001;79:1–37. doi: 10.1016/S0010-0277(00)00123-2. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 111.Dehaene S., Changeux J.-P. Ongoing Spontaneous Activity Controls Access to Consciousness: A Neuronal Model for Inattentional Blindness. *PLoS Biol.* 2005;3:e141. doi: 10.1371/journal.pbio.0030141. [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 112.Del Cul A., Baillet S., Dehaene S. Brain Dynamics Underlying the Nonlinear Threshold for Access to Consciousness. *PLoS Biol.* 2007;5:e260. doi: 10.1371/journal.pbio.0050260. [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 113.De Lange F.P., van Gaal S., Lamme V.A.F., Dehaene S. How Awareness Changes the Relative Weights of Evidence during Human Decision-Making. *PLoS Biol.* 2011;9:e1001203. doi: 10.1371/journal.pbio.1001203. [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 114.Sergent C., Baillet S., Dehaene S. Timing of the brain events underlying access to consciousness during the attentional blink. *Nat. Neurosci.* 2005;8:1391–1400. doi: 10.1038/nn1549. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]

- 115.De Lange F.P., Jensen O., Dehaene S. Accumulation of Evidence during Sequential Decision Making: The Importance of Top-Down Factors. *J. Neurosci.* 2010;30:731–738. doi: 10.1523/JNEUROSCI.4080-09.2010. [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 116.Schwiedrzik C.M., Singer W., Melloni L. Subjective and objective learning effects dissociate in space and in time. *Proc. Natl. Acad. Sci. USA.* 2011;108:4506–4511. doi: 10.1073/pnas.1009147108. [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 117.Gurwitsch A. *The Field of Consciousness*. Duquesne University Press; Pittsburgh, PA, USA: 1964. [[Google Scholar](#)]
- 118.Gurwitsch A. *Marginal Consciousness*. Lester Emb. Ohio University Press; Athens, OH, USA: 1985. [[Google Scholar](#)]
- 119.Kant I. In: *Critique of Pure Reason*. Guyer P., Wood A., editors. Cambridge University Press; Cambridge, UK: 1787. [[Google Scholar](#)]
- 120.Nidditch P. *John Locke: An Essay Concerning Human Understanding*. Oxford University Press; Oxford, UK: 1975. [[Google Scholar](#)]
- 121.Dretske F. *Naturalizing the Mind*. Bradford/MIT Press; Cambridge, MA, USA: 1995. [[Google Scholar](#)]
- 122.Tye M. *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. MIT Press; Cambridge, MA, USA: 1995. [[Google Scholar](#)]
- 123.Tye M. *Consciousness, Color, and Content*. MIT Press; Cambridge, MA, USA: 2000. [[Google Scholar](#)]
- 124.Rosenthal D.M. *Consciousness and Mind*. Clarendon Press; Wotton-under-Edge, UK: 2005. [[Google Scholar](#)]
- 125.Rounis E., Maniscalco B., Rothwell J.C., Passingham R.E., Lau H. Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness. *Cogn. Neurosci.* 2010;1:165–175. doi: 10.1080/17588921003632529. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 126.Gennaro R. *The Consciousness Paradox: Consciousness, Concepts, and Higher-Order Thoughts (Representation and Mind)* MIT Press; Cambridge, MA, USA: 2011. [[Google Scholar](#)]
- 127.Wang Y., Patel S., Patel D. A layered reference model of the brain (LRMB) *IEEE Trans Syst. Man Cybern. Part C Appl. Rev.* 2006;36:124–133. doi: 10.1109/TSMCC.2006.871126. [[DOI](#)] [[Google Scholar](#)]

- 128.Wang Y. Transactions on Computational Science V. Springer; Berlin/Heidelberg, Germany: 2009. Formal description of the cognitive process of memorization; pp. 81–98. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) [[Google Scholar](#)]
- 129.Wang Y. On Cognitive Informatics. Brain Mind. 2003;4:151–167. doi: 10.1023/A:1025401527570. [[DOI](#)] [[Google Scholar](#)]
- 130.Wang Y. The Theoretical Framework of Cognitive Informatics. Int. J. Cogn. Inform. Nat. Intell. 2007;1:1–27. doi: 10.4018/jcini.2007010101. [[DOI](#)] [[Google Scholar](#)]
- 131.Payne D., Wenger M. Cognitive Psychology. Houghton Mifflin College Division; Boston, MA, USA: 1998. [[Google Scholar](#)]
- 132.Smith R. Psychology. West Publishing Company; St. Paul, MN, USA: 1993. [[Google Scholar](#)]
- 133.Sternberg R.J. In Search of the Human Mind. 2nd ed. Harcourt Brace College Publishers; Cambridge, MA, USA: 1998. [[Google Scholar](#)]
- 134.Wang Y. On abstract intelligence and brain informatics: Mapping the cognitive functions onto the neural architectures; Proceedings of the IEEE 11th International Conference on Cognitive Informatics and Cognitive Computing; Kyoto, Japan. 22–24 August 2012; Piscataway, NJ, USA: IEEE; 2012. pp. 5–6. [[Google Scholar](#)]
- 135.Castaigne P., Lhermitte F., Buge A., Escourolle R., Hauw J.J., Lyon-Caen O. Paramedian thalamic and midbrain infarcts: Clinical and neuropathological study. Ann. Neurol. 1981;10:127–148. doi: 10.1002/ana.410100204. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 136.Fridman E.A., Schiff N.D. Neuromodulation of the conscious state following severe brain injuries. Curr. Opin. Neurobiol. 2014;29:172–177. doi: 10.1016/j.conb.2014.09.008. [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 137.Schiff N.D. Central thalamic deep brain stimulation to support anterior forebrain mesocircuit function in the severely injured brain. J. Neural Transm. 2016;123:797–806. doi: 10.1007/s00702-016-1547-0. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 138.Schiff N.D. Handbook of Clinical Neurology. Elsevier B.V.; Amsterdam, The Netherlands: 2013. Central thalamic deep brain stimulation for support of

forebrain arousal regulation in the minimally conscious state; pp. 295–306. [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)

- 139. Lee K.H., Meador K.J., Park Y.D., King D.W., Murro A.M., Pillai J.J., Kaminski R.J. Pathophysiology of altered consciousness during seizures: Subtraction SPECT study. *Neurology*. 2002;59:841–846. doi: 10.1212/WNL.59.6.841. [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
- 140. Blumenfeld H., McNally K.A., Vanderhill S.D., Paige A.L., Chung R., Davis K., Norden A.D., Stokking R., Studholme C., Novotny E.J., et al. Positive and Negative Network Correlations in Temporal Lobe Epilepsy. *Cereb. Cortex*. 2004;14:892–902. doi: 10.1093/cercor/bhh048. [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
- 141. Blumenfeld H. Impaired consciousness in epilepsy. *Lancet Neurol*. 2012;11:814–826. doi: 10.1016/S1474-4422(12)70188-6. [\[DOI\]](#) [\[PMC free article\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
- 142. Gray J. *Consciousness: Creeping Up on the Hard Problem*. Oxford University Press; Oxford, UK: 2004. [\[Google Scholar\]](#)
- 143. Godwin C.A., Gazzaley A., Morsella E. *The Unity of Mind, Brain and World: Current Perspectives on a Science of Consciousness*. Cambridge University Press; Cambridge, UK: 2013. Homing in on the brain mechanisms linked to consciousness: Buffer of the perception-and-action interface; pp. 43–76. [\[Google Scholar\]](#)
- 144. Husserl E. In: *The Phenomenology of Internal Time-Consciousness*. Churchill J., editor. Martinus Nijhoff; The Hague, The Netherlands: 1964. [\[Google Scholar\]](#)
- 145. Natsoulas T. Basic problems of consciousness. *J. Pers. Soc. Psychol*. 1981;41:132–178. doi: 10.1037/0022-3514.41.1.132. [\[DOI\]](#) [\[Google Scholar\]](#)
- 146. Petkov V. *Minkowski Spacetime: A hundred Years Later*. Springer; Berlin, Germany: 2010. [\[Google Scholar\]](#)
- 147. O'Keefe J., Dostrovsky J. The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Res*. 1971;34:171–175. doi: 10.1016/0006-8993(71)90358-1. [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
- 148. O'Keefe J. Place units in the hippocampus of the freely moving rat. *Exp. Neurol*. 1976;51:78–109. doi: 10.1016/0014-4886(76)90055-8. [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)

- 149. MacDonald C.J., Lepage K.Q., Eden U.T., Eichenbaum H. Hippocampal “Time Cells” Bridge the Gap in Memory for Discontiguous Events. *Neuron*. 2011;71:737–749. doi: 10.1016/j.neuron.2011.07.012. [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 150. Sieb R. Four-Dimensional Consciousness. *Act. Nerv. Super.* 2017;59:43–60. doi: 10.1007/s41470-017-0008-x. [[DOI](#)] [[Google Scholar](#)]
- 151. Di Biase F. A holoinformational model of consciousness. *Quantum Biosyst.* 2009;3:207–220. [[Google Scholar](#)]
- 152. Pribram K., Yasue K., Jibu M. *Brain and Perception: Holonomy and Structure in Figural Processing*. Psychology Press; London, UK: 1991. [[Google Scholar](#)]
- 153. Bohm D., Hiley B. *The Undivided Universe—An Ontological Interpretation of Quantum Theory*. Routledge; London, UK: 1993. [[Google Scholar](#)]
- 154. Pribram K.H. *Rethinking Neural Networks: Quantum Fields and Biological Data*. Lawrence Erlbaum Associates; Mahwah, NJ, USA: 1993. [[Google Scholar](#)]
- 155. Eccles J. A unitary hypothesis of mind-brain interaction in the cerebral cortex. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* 1990;240:433–451. doi: 10.1098/rspb.1990.0047. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 156. Hameroff S. Consciousness, the brain, and spacetime geometry. *Ann. N. Y. Acad. Sci.* 2006;929:74–104. doi: 10.1111/j.1749-6632.2001.tb05709.x. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 157. Gödel K. Über formal unentscheidbare sätze der principia mathematica und verwandter systeme I. *Mon. Math. Phys.* 1931;38:173–198. doi: 10.1007/BF01700692. [[DOI](#)] [[Google Scholar](#)]
- 158. Penrose R. *Shadows of the Mind: A Search for the Missing Science of Consciousness*. Oxford University Press; Oxford, UK: 1994. [[Google Scholar](#)]
- 159. Schrödinger E. An Undulatory Theory of the Mechanics of Atoms and Molecules. *Phys. Rev.* 1926;28:1049–1070. doi: 10.1103/PhysRev.28.1049. [[DOI](#)] [[Google Scholar](#)]
- 160. Penrose R. On Gravity’s role in Quantum State Reduction. *Gen. Relativ. Gravit.* 1996;28:581–600. doi: 10.1007/BF02105068. [[DOI](#)] [[Google Scholar](#)]

- 161. Robinson W. Epiphenomenalism. In: Zalta E., Nodelman U., Allen C., editors. Stanford Encyclopedia of Philosophy. Stanford University; Stanford, CA, USA: 2012. [[Google Scholar](#)]
- 162. Wegner D. The Illusion of Conscious Will. The MIT Press; Cambridge, MA, USA: 2002. [[Google Scholar](#)]
- 163. Tegmark M. Importance of quantum decoherence in brain processes. Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top. 2000;61:4194–4206. doi: 10.1103/PhysRevE.61.4194. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 164. Bell J.S. On the Problem of Hidden Variables in Quantum Mechanics. Rev. Mod. Phys. 1966;38:447–452. doi: 10.1103/RevModPhys.38.447. [[DOI](#)] [[Google Scholar](#)]
- 165. Kochen S., Specker E. The problem of hidden variables in quantum mechanics. J. Math. Mech. 1967;17:59–87. doi: 10.1512/iumj.1968.17.17004. [[DOI](#)] [[Google Scholar](#)]
- 166. Frankfurt H. Equality as a Moral Ideal. Ethics. 1987;98:21–43. doi: 10.1086/292913. [[DOI](#)] [[Google Scholar](#)]
- 167. Widerker D., Frankfurt H. Alternate possibilities and moral responsibility. In: Frankfurt H., editor. Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities. 1st ed. Taylor and Francis; Abingdon, UK: 2018. pp. 17–25. [[Google Scholar](#)]
- 168. Draguhn A., Traub R.D., Schmitz D., Jefferys J.G.R. Electrical coupling underlies high-frequency oscillations in the hippocampus in vitro. Nature. 1998;394:189–192. doi: 10.1038/28184. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 169. Gilbert C.D., Sigman M. Brain States: Top-Down Influences in Sensory Processing. Neuron. 2007;54:677–696. doi: 10.1016/j.neuron.2007.05.019. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 170. Lambert N., Chen Y.-N., Cheng Y.-C., Li C.-M., Chen G.-Y., Nori F. Quantum biology. Nat. Phys. 2012;9:10–18. doi: 10.1038/nphys2474. [[DOI](#)] [[Google Scholar](#)]
- 171. Kak S. Biological Memories and Agents as Quantum Collectives. NeuroQuantology. 2013;11:391–398. doi: 10.14704/nq.2013.11.3.682. [[DOI](#)] [[Google Scholar](#)]
- 172. Bohm D. Wholeness and the Implicate Order. Psychology Press; London, UK: 2002. [[Google Scholar](#)]

- 173.Mensky M.B. Mathematical Models of Subjective Preferences in Quantum Concept of Consciousness. *NeuroQuantology*. 2011;9:614–620. doi: 10.14704/nq.2011.9.4.488. [[DOI](#)] [[Google Scholar](#)]
- 174.DeWitt B. *The Many Worlds Interpretation of Quantum Mechanics*. Princeton University Press; Princeton, NJ, USA: 1973. [[Google Scholar](#)]
- 175.Hoffman D.D., Singh M., Prakash C. The Interface Theory of Perception. *Psychon. Bull. Rev.* 2015;22:1480–1506. doi: 10.3758/s13423-015-0890-8. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 176.Hoffman D.D. *Hallucinations: Research and Practice*. Springer; New York, NY, USA: 2012. The construction of visual reality; pp. 7–15. [[Google Scholar](#)]
- 177.Hoffman D.D. The Interface Theory of Perception (To appear in *The Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience*) In: Dickinson S., Tarr M., Leonardis A., Schiele B., editors. *Object Categorization: Computer and Human Vision Perspectives*. Cambridge University Press; New York, NY, USA: 2009. pp. 148–165. [[Google Scholar](#)]
- 178.Hoffman D.D. “The sensory desktop”. In: Brockman J., editor. *This Will Make You Smarter: New Scientific Concepts to Improve Your Thinking*. Harper Perennial; New York, NY, USA: 2012. pp. 135–138. [[Google Scholar](#)]
- 179.Coleman S. The Real Combination Problem: Panpsychism, Micro-Subjects, and Emergence. *Erkenntnis*. 2013;79:19–44. doi: 10.1007/s10670-013-9431-x. [[DOI](#)] [[Google Scholar](#)]
- 180.Bargh J.A., Morsella E. The Unconscious Mind. *Perspect. Psychol. Sci.* 2008;3:73–79. doi: 10.1111/j.1745-6916.2008.00064.x. [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 181.Hoffman D.D. Public Objects and Private Qualia. In: Albertazzi L., editor. *The Wiley-Blackwell Handbook of Experimental Phenomenology*. Wiley-Blackwell; New York, NY, USA: 2013. pp. 71–89. [[Google Scholar](#)]
- 182.Dayan P., Abbott L., Abbott L. *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. MIT Press; Cambridge, MA, USA: 2001. [[Google Scholar](#)]
- 183.Eliasmith C., Anderson C. *Neural Engineering: Computation, Representation, and Dynamics in Neurobiological Systems*. MIT Press; Cambridge, MA, USA: 2004. [[Google Scholar](#)]

- 184.O'Reilly R., Munakata Y. Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain. MIT Press; Cambridge, MA, USA: 2000. [[Google Scholar](#)]
- 185.Blouw P., Solodkin E., Thagard P., Eliasmith C. Concepts as Semantic Pointers: A Framework and Computational Model. *Cogn. Sci.* 2015;40:1128–1162. doi: 10.1111/cogs.12265. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 186.Damasio A. *Self Comes to Mind: Constructing the Conscious Brain*. Vintage; New York, NY, USA: 2012. [[Google Scholar](#)]
- 187.Northoff G., Lamme V. Neural signs and mechanisms of consciousness: Is there a potential convergence of theories of consciousness in sight? *Neurosci. Biobehav. Rev.* 2020;118:568–587. doi: 10.1016/j.neubiorev.2020.07.019. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 188.Drigas A.S., Pappas M.A. The Consciousness-Intelligence-Knowledge Pyramid: An 8 × 8 Layer Model. *Int. J. Recent Contrib. Eng. Sci. IT.* 2017;5:14–25. doi: 10.3991/ijes.v5i3.7680. [[DOI](#)] [[Google Scholar](#)]
- 189.Drigas A., Mitsea E. The Triangle of Spiritual Intelligence, Metacognition and Consciousness. *Int. J. Recent Contrib. Eng. Sci. IT.* 2020;8:4–23. doi: 10.3991/ijes.v8i1.12503. [[DOI](#)] [[Google Scholar](#)]
- 190.Shannon C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* 1948;27:379–423. doi: 10.1002/j.1538-7305.1948.tb01338.x. [[DOI](#)] [[Google Scholar](#)]
- 191.Madden A.D. A definition of information. *Aslib Proc.* 2000;52:343–349. doi: 10.1108/EUM0000000007027. [[DOI](#)] [[Google Scholar](#)]
- 192.Capurro R., Hjørland B. The concept of information. *Annu. Rev. Inf. Sci. Technol.* 2005;37:343–411. doi: 10.1002/aris.1440370109. [[DOI](#)] [[Google Scholar](#)]
- 193.Drigas A., Mitsea E. The 8 Pillars of Metacognition. *Int. J. Emerg. Technol. Learn.* 2020;15:162–178. doi: 10.3991/ijet.v15i21.14907. [[DOI](#)] [[Google Scholar](#)]
- 194.Blum M., Blum L. A Theoretical Computer Science Perspective on Consciousness. *J. Artif. Intell. Conscious.* 2021;8:1–42. doi: 10.1142/S2705078521500028. [[DOI](#)] [[Google Scholar](#)]
- 195.Rudrauf D., Bennequin D., Granic I., Landini G., Friston K., Williford K. A mathematical model of embodied consciousness. *J. Theor. Biol.* 2017;428:106–131. doi: 10.1016/j.jtbi.2017.05.032. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]

What if consciousness is not an emergent property of the brain? Observational and empirical challenges to materialistic models

[Helané Wahbeh](#)^{1*}, Dean Radin, ¹[Cedric Cannard](#), and ¹[Arnaud Delorme](#)^{1,2}

- ¹Research Department, Institute of Noetic Sciences, Petaluma, CA, United States
- ²Swartz Center for Computational Neuroscience, Institute of Neural Computation, University of California, San Diego, San Diego, CA, United States

The nature of consciousness is considered one of science's most perplexing and persistent mysteries. We all know the subjective experience of consciousness, but where does it arise? What is its purpose? What are its full capacities? The assumption within today's neuroscience is that all aspects of consciousness arise solely from interactions among neurons in the brain. However, the origin and mechanisms of *qualia* (i.e., subjective or phenomenological experience) are not understood. David Chalmers coined the term "the hard problem" to describe the difficulties in elucidating the origins of subjectivity from the point of view of reductive materialism. We propose that the hard problem arises because one or more assumptions within a materialistic worldview are either wrong or incomplete. If consciousness entails more than the activity of neurons, then we can contemplate new ways of thinking about the hard problem. This review examines phenomena that apparently contradict the notion that consciousness is exclusively dependent on brain activity, including phenomena where consciousness appears to extend beyond the physical brain and body in both space and time. The mechanisms underlying these "non-local" properties are vaguely suggestive of quantum entanglement in physics, but how such effects might manifest remains highly speculative. The existence of these non-local effects appears to support the proposal that post-materialistic models of consciousness may be required to break the conceptual impasse presented by the hard problem of consciousness.

What is consciousness?

The term "consciousness" means different things to different audiences. From a lay perspective, the fact of consciousness (here meaning *awareness*) is so self-evident that the only question that may arise is why anyone would consider consciousness to be mysterious in the first place, akin to asking a fish, "What is water?" From a scientific perspective, the unsolved mystery is how consciousness emerges from brain activity. How does a three-pound lump of tissue inside the skull give rise to a mind that is self-aware and enjoys subjective experience? In philosophy, which has debated the "mind-body problem" for millennia, many sets of assumptions have been proposed, ranging from materialism (i.e., all is dependent on or reducible to the physical) to idealism (i.e., ideas or thoughts make up fundamental reality). For mystics and others entranced by the esoteric traditions, the problem is not so much about the mind but how the physical world emerges from a non-physical "substance."

Over the past few decades, most scientifically oriented research on consciousness has studied consciousness as a controlled variable. They have reduced consciousness to

simpler constructs, such as perception, and focused on comparisons in brain processes during conscious and unconscious conditions, the so-called “contrastive approach.” In this approach, differences in brain activity are examined when the same stimulus is subjectively perceived versus when it is not ([Baars, 2005](#)). This search for the neural correlates of consciousness (NCC) is defined as the “minimum neuronal mechanisms jointly sufficient for any one specific conscious experience” ([Koch et al., 2016](#)). Experimental designs have used methods like stimulus masking ([Shapiro et al., 1997](#); [Simons and Chabris, 1999](#); [Dehaene et al., 2001](#)) or binocular rivalry ([Leopold and Logothetis, 1999](#)) to examine the brain activity associated with subliminal perception where information is not consciously perceived yet processed by the brain. Other approaches have relied on brain lesions ([Hebb and Penfield, 1940](#)), or, more recently, by artificially modulating brain activity in specific regions and networks intracranially, during neurosurgery, or non-invasively with transcranial magnetic/electric/ultrasound stimulation ([Selimbeyoglu and Parvizi, 2010](#)). Other authors have focused solely on states of unconsciousness during general anesthesia ([Hudetz, 2012](#)), epilepsy ([Blumenfeld and Taylor, 2003](#)), or sleep ([Steriade et al., 2001](#)). While a front line of research, the NCC approach has conceptual limitations. Mainly, the terms “consciousness” and “correlates” are enmeshed because brain events that co-vary with conscious experience can either be the experience’s neural substrates, the prerequisites, or even the experience’s neural consequences ([Aru et al., 2012](#); [de Graaf et al., 2012](#)).

Several scientific theories of consciousness are primed to be tested experimentally instead of merely identifying correlations between conscious/unconscious events and brain activity. In the sections below, we briefly summarize the main neuroscience theories, referred to as physicalist or reductionist theories, most of which assume that consciousness emerges from the brain. While most physicalist theories aim to explain different aspects of consciousness, they often share similarities and have been recently grouped into four categories: higher-order theories (HOTs), global work-space theories (GWTs), integrated information theory (IIT), and re-entry and predictive processing theories ([Seth and Bayne, 2022](#)). Here, we follow this categorization and present a brief summary. However, a full review of each theory is beyond this article’s scope (see [Seth and Bayne, 2022](#), for a full review).

Physicalist theories of consciousness

Global work-space theories

Cognitive scientist Bernard J. Baars first proposed the GWT in 1983. GWT is a cognitive architecture inspired by artificial intelligence where a centralized resource is available through which specialized processors share and receive information ([Baars, 2005](#); [Baars et al., 2021](#)). The theory is based on the observation that there are highly specialized brain regions that process information locally and unconsciously, such as the visual cortex. Conscious experience occurs once there is a distributed activity in other brain areas, that is, “broadcasting” to the system as a whole ([Baars, 2005](#)). The widespread access, operation, and coordination of specialized neural networks, which

would otherwise operate autonomously, is coordinated by consciousness, involving mainly the frontoparietal network and high-frequency oscillatory rhythms ([Baars, 2005](#)).

Dehaene and colleagues have adjusted the GWT to account for new knowledge about the brain, the so-called “neuronal global work-space” ([Dehaene et al., 2003](#)). For example, global activity among work-space neurons is generated by excitatory neurons responding to sensory stimuli with long-range cortico-cortical connections. In turn, this global activity inhibits alternative activity patterns among the work-space neurons to prevent the conscious experience of other stimuli (e.g., attentional blink). This inhibition mechanism would allow the unified experience of consciousness (the composition and exclusion axioms in IIT). Supported by experimental findings in the context of the search for NCCs, this popular model suggests that (1) most of the brain’s computations are performed in a non-conscious operation mode and that (2) conscious access must be distinguished from selective attention, (3) conscious perception may be characterized by a non-linear function that ‘ignites’ a network of distributed areas (a gradual increase in stimulus visibility accompanied by a sudden transition of the neuronal work-space into a pattern, i.e., the broadcasting), and (4) the selected information gains access to additional computations for conscious perception such as temporary maintenance, global sharing, flexible routing ([Dehaene et al., 2014](#)). The model predicts that measures of complexity, long-distance correlation, and integration of brain signals should provide reliable indices of conscious processing and has clinical applications (e.g., sleep, coma, anesthesia; [Dehaene et al., 2014](#); [Mashour et al., 2020](#)). Some investigators have recently attempted to generalize the GWT to brain-inspired artificial architectures by implementing the GWT into deep learning algorithms ([VanRullen and Kanai, 2021](#)). GWT is promising as a model. However, it is unclear what determines when information is broadcasted to the whole (e.g., threshold) and what discriminates different types of subjective experiences within the GWT theory.

Higher-order theories

The HOTs family includes, for example, the self-organizing meta-representational account theory (SOMA; [Cleeremans et al., 2020](#)), the adversarial framework for probabilistic computation ([Gershman, 2019](#)), and the perceptual reality monitoring theory ([Lau, 2019](#)). In this view, consciousness is defined as a higher-order representation of lower-order representations. In other words, subjective experiences reflect higher brain orders like meta-representations, which have learned to describe and interpret the lower-order functions such as local modules specialized in processing specific information. In this sense, consciousness is the brain’s unconscious, embodied, enactive, non-conceptual theory about itself ([Cleeremans et al., 2020](#)). While Gehrman’s model interprets this view in terms of computations and algorithms, Lau’s view is in terms of belief and epistemic justification on a subjective level. The NCCs for the HOTs generally consist in anterior regions of the brain, like the prefrontal cortex, reflecting their involvement in complex cognitive functions ([Lau and Rosenthal, 2011](#)).

The primary limitation of both GWTs and HOTs is that they do not account for the phenomenal differences between distinct subjective experiences ([Seth and Bayne, 2022](#)). Furthermore, neither GWTs nor HOTs have addressed the adaptive and

evolutionary role of conscious experience (i.e., embodiment and environmental embeddedness; [Seth and Bayne, 2022](#)).

Integrated information theory

Integrated information theory is a mathematical approach based on phenomenology by first identifying the essential properties of consciousness, the so-called axioms: *intrinsic information* – each experience is specific, there is intrinsic information in the system that is associated with that experience that differs from alternative experiences, *information* – consciousness is composed of a specific set of specific phenomenal distinctions and is different from other possible experiences, *integration* – consciousness is unified with each experience being irreducible to non-interdependent components, and *exclusion* – consciousness is unique in content and spatio-temporal context ([Tononi, 2015](#); [Tononi et al., 2016](#)). IIT infers the postulates or requirements for a physical system to be a physical substrate of consciousness from the axioms.

Integrated information refers to a system's constituents that are discriminated by their respective information. The whole cannot be reduced to the information of each part, called the “causal-effect power” ([Oizumi et al., 2014](#)). These irreducible maxima of additional integrated information generated by the system as a whole compared to its parts are termed and quantified as Φ (“phi”) and affect the probability of its past and future states. The larger the Φ value, the more intrinsic cause-effect power the system has and the more conscious it is ([Koch, 2018](#)). Thus, any complex and interconnected physical system with these properties will have some level or quantity of consciousness, corresponding to the amount of intrinsic cause-effect power the substrate has. The content of a conscious experience is predicted to be structurally identical to the cause-effect structure of its physical substrate ([Albantakis, 2020](#)). So the more structurally complex the system is, the more structurally complex the experience is. IIT, therefore, provides a potential method to (1) assess whether a physical system constitutes a physical substrate of consciousness through its compliance with the postulates, (2) quantify the level of consciousness of that system, and (3) estimate its phenomenological structure in causal terms ([Albantakis, 2020](#)).

Unlike HOTs and GWT, progress has been made to assess the relevance of environmental embodiment for consciousness (i.e., consciously capturing the causal structure of a rich environment). For example, [Haun and Tononi \(2019\)](#) showed how this is useful for successfully navigating spacetime ([Haun and Tononi, 2019](#)). [Albantakis et al. \(2014\)](#) showed that this is an important driving force for organisms to develop highly integrated networks (“brains”), leading to an increase in their internal complexity ([Albantakis et al., 2014](#)). These concepts bring evolutionary context to the development of consciousness and the complexity of the brain.

Re-entry and predictive processing approaches

Re-entry theories were developed from the idea that “we are fooled into thinking that we know what we are conscious of” ([Lamme, 2010](#), p. 204). Therefore, introspective or behavioral observations make understanding the mind-brain relationship impossible.

Thus, this approach removes intuitive or psychological notions of conscious experience from the study of consciousness ([Lamme, 2006](#)). In the local recurrency theory, conscious perception corresponds to top-down signaling ([Lamme and Roelfsema, 2000](#); [Lamme, 2006](#); [Seth and Bayne, 2022](#)). Consciousness emerges from simple localized recurrent, top-down processing within perceptual cortices, and frontal and parietal regions would be crucial to perceptual experience content, reasoning, and decision making ([Lamme and Roelfsema, 2000](#); [Lamme, 2006](#), [2010](#)). Local recurrency theory is similar to GWTs, except that we do not know what we are conscious of and that it is about perception. In contrast, GWT is about access ([Lamme, 2010](#)).

Predictive processing approaches in computational neuroscience, such as the hierarchical generative model, consider the brain as a machine that matches bottom-up inputs with top-down expectations through cortical processing, aiming to minimize the error in these predictions ([Clark, 2013](#)). In this bi-directional model, top-down connections from higher levels encode the predictions in the lower levels. This fully explains the bottom-up signal and leaves only residual prediction errors propagating information forward in the system to update the following predictions ([Huang and Rao, 2011](#); [Clark, 2013](#)). This predictive control function is termed “active inference” ([Seth and Bayne, 2022](#)).

The adaptive resonance theory (ART) was developed by Stephen Grossberg and Gail Carpenter to address the “stability-plasticity dilemma,” or how the brain learns so quickly and stably without forgetting past knowledge ([Grossberg, 2013a](#)). For ART, various brain processes are required, namely Consciousness, Learning, Expectation, Attention, Resonance, and Synchrony (the CLEARS processes). Top-down expectations (E) direct the focus of attention (A) across competitive features. When a match occurs between the expectation and what is perceived, a resonant synchronization (RS) occurs and generates attentional focus driving fast learning (L) of bottom-up, called “many-to-one maps,” and top-down, called “one-to-many maps” representations. This whole process is called “adaptive resonance.” Here, the focus of attention corresponds to the minimization of error in the prediction function. There is growing experimental data supporting these predictions, and some ART models are thought to explain and predict behavioral, anatomical, neurophysiological, and biochemical data ([Grossberg, 2013b](#)).

In brief, for this family of theories, perceptual experience is the brain’s best guess of its cause (minimization of the prediction error) through the exchange of top-down predictions and bottom-up prediction errors ([Rao and Ballard, 1999](#); [Friston, 2010](#); [Hohwy, 2013](#); [Seth and Bayne, 2022](#)). For example, subjective emotions are considered to emerge from cognitive evaluations of physiological changes in the body and their causes (“constructed emotion” and “interoceptive inference”; [Seth, 2013](#); [Barrett, 2016](#)).

Summary of physicalist models

One common element across physicalist theories is the uncertainty reduction that results from allocating mechanisms to consciousness. The system must settle into one unified and highly informative representational state ([Hohwy and Seth, 2020](#)). This point

of uncertainty reduction often corresponds to a threshold at which the contents become conscious (e.g., broadcasting for GWT, optimization of signal-to-noise ratio for HOTS, ϕ for IIT, information integration, and learning for re-entry and predictive theories). The second common element is the high importance of top-down signaling (e.g., a system with no top-down dimension has no ϕ in IIT; [Oizumi et al., 2014](#)).

The first disagreement between these theories regards the distinction between consciousness and cognition. Cognitive access relies on consciousness in GWT, consciousness is cognitively accessible in HOTS, whereas cognition is possible without consciousness for IIT and vice versa for predictive processing and re-entry theories. The second, more critical, disagreement between physicalist theories is in regards to the unity of consciousness, i.e., the subjective experience of awareness that fully captures what it is like to be an agent at any time. It is required by IIT and may be supported by the broadcast in GWTs but is ignored by HOTS, re-entry, and predictive theories that do not consider this concept necessary. These various theories may, in fact, address different aspects of consciousness.

In the well-known parable, “The Blind Men and the Elephant,” each person attempts to describe the elephant but only touches one small part ([Saxe, 2016](#)). Thus, they arrive at very different conclusions about what an elephant is like: a tree, a fan, a rope, a spear. They commence arguing, each one convinced that they are correct in their conclusions. The story’s moral is that we must step back to observe broader perspectives to describe and fully understand the elephant’s nature. Similarly, these physicalist theories may describe some aspects of consciousness very well, but they likely do not describe it completely.

In conclusion, while extensive and rigorous efforts have attempted to find, test, and validate physicalist theories or NCCs ([Reardon, 2019](#); [Templeton World Charity Foundation, 2022](#)), the field is far from consensus about which theories are valid and could potentially explain consciousness or neural differences between different phenomenological experiences. We suggest that the gaps in physicalist theories in explaining consciousness may arise because the debate is framed around how the brain *generates* consciousness. The theories discussed so far attempt to explain phenomenological experience or qualia through reductionist brain mechanisms or correlations, often equated to computational or information processing systems. Some of these theories (e.g., re-entry or predictive processing theories) even consider subjective reports and introspection to be unreliable. Consequently, none of these theories have completely and convincingly explained the nature of consciousness.

A different approach: Non-local consciousness theories

Alternative non-physicalist theories may inform other aspects of consciousness that are not completely explained by physicalist theories. Physicalist theories usually assume that consciousness is generated solely and purely from the brain and is only local to the brain. Alternatively, non-physicalist theories do not make these assumptions, even though both types of theories attempt to explain the underlying brain mechanisms of

consciousness. Physicalist theories purport that consciousness originates from physical substrates like neurons that have evolved to be more and more complex over time through adaptation, leading to the emergence of consciousness. Non-physical models do not assume a physical substrate generates consciousness, and many even propose that consciousness is, in fact, more fundamental than matter and spacetime. In this view, that is the natural view for most ancient and eastern cultures, matter and spacetime arise from consciousness rather than the other way around. Perhaps a non-physicalist framework where consciousness is considered fundamental and has non-local properties (such as at the quantum scale) would better explain the full range of reported human phenomenology. For example, there are well-documented experiences of people perceiving information from distant locations, the future, and mental impressions from other people without the use of rationale or traditional means ([Cardeña, 2018](#)). In addition, there are verified cases of cognitive function when the neural substrate is severely degenerated, precluding normal brain function. These experiences, most of which are currently regarded as anomalous, will be described in the Section “Phenomena suggested by a model of nonlocal consciousness” as cases of what would be observed should non-physicalist theories of consciousness be valid.

These and other documented phenomenological experiences suggest a different nature of consciousness: one that may not be exclusively generated by neuronal activity and exhibits properties that transcend the conventional constraints of spacetime and, therefore, the physical body. The term “non-local consciousness” has been proposed to denote these purported transcendent properties of consciousness ([Dossey, 1994](#)). Physicalist scientists typically consider such experiences anomalous because they challenge prevailing assumptions about the nature and role of consciousness in physical reality. The term non-local is also referenced as a central idea within physics as an aspect of the physical world. For example, the brain operating, even to a small extent, in a quantum fashion might be a valid explanation for these non-local phenomena. However, this idea is not yet widely accepted because while there is evidence for quantum biology, quantum coherence in brain processing is so short-lived that it appears irrelevant to understanding consciousness. Neuroscience today says consciousness is generated by and localized in the brain because it emerges from brain activity. Alternatively, we propose that consciousness may not originate in the brain, although some aspects of human perception of consciousness may be dependent on the brain. We also suggest that awareness also extends beyond the brain. These non-physical, non-local properties of consciousness may be due to a non-local material effect, to consciousness being fundamental, or something else we have not yet discovered.

To begin an exploration of some of these non-physical theories, we present theoretical frameworks proposed by scientists from multiple disciplines, most of which include the idea that consciousness is fundamental, meaning that consciousness precedes the physical substrates ([Chalmers, 1996](#); [Curriuan, 2017](#); [Kastrup, 2017, 2021](#); [Goff, 2019](#); [Faggini, 2021a](#)). Traditional materialists envision a world in which mathematics is more fundamental than physics, which is more fundamental than chemistry, which is, in turn, more fundamental than biology. Thus, in this way, physical processes are foundational to the generation of our biology. However, suppose we envision that

consciousness is actually more foundational than physics. In that case, we can imagine that these other physical disciplines can arise from consciousness. In other words, if biology emerges from chemistry, chemistry from physics, and physics emerges from consciousness, then from this perspective, non-local consciousness phenomena would no longer be regarded as anomalous because consciousness can transcend some physical laws. Theories proposing this idea have been offered by Federico Faggin, Donald Hoffman, Bernardo Kastrup, Vernon Neppe, and numerous others. Most of these theories are speculative, while others are supported through mathematical arguments or empirical data ([Hoffman et al., 2015](#); [Neppe and Close, 2020](#); [Faggin, 2021b](#)). We briefly review a sample of non-local consciousness theories.

Operational probabilistic theory

Federico Faggin starts with the assumption that reality emerges from the free-will communications of a vast number of conscious entities ([Faggin, 2021a](#)). Faggin calls the totality of what potentially and actually exists, *One*. Any self-knowing within this *one* is a transformation from potential existence into actual existence, where potential existence is the “reservoir” of self-knowing that has not yet manifested. Each new self-knowing brings rise to a *consciousness unit (CU)*. The CU reflects the whole of *One* and is also part of *One* because *One* is never complete in its self-knowing process. Thus there must be continued self-knowing and continual generation of CUs, which explains an apparently growing number of conscious entities ([Faggin, 2021a](#), p. 294). Faggin describes the CUs characteristics and how they combine into *self*, in which an entity with identity, awareness, and agency is dynamic, holistic, and self-knowing. Faggin views the physical world as a virtual reality metaphor, in which sophisticated avatars controlled by conscious beings interact with each other, where the body that controls the avatar exists outside the computer and is not part of the program. Similarly, the conscious entities that control physical bodies exist beyond the physical world that contains the body ([Faggin, 2021b](#), p. 286).

Interface theory of perception

Donald D. Hoffman proposes a model based on a mathematical structure called “conscious agents.” Space and time emerge from conscious agents’ exchanges ([Hoffman, 2014](#)). Hoffman proposes that our perceptions (i.e., the conscious agents) are not views of a grounded truth but are more like a personal computer’s operating system and interface ([Hoffman, 2014, 2019](#)). Perceptions allow us to interact dynamically with the world and survive and evolve in this environment but not be aware of its actual structure. Space-time and physical objects do not represent a universal objective reality but are species-specific components that provide an evolutionary advantage. Hoffman highlights that evolutionarily, perception of spacetime and the physical world are shaped by natural selection in such a way that obfuscates the truth that we are experiencing an interface rather than a universal objective reality and thus influences adaptive behaviors. He further claims that the equations of quantum mechanics can be derived from formalized descriptions of the interactions between conscious agents ([Hoffman et al., 2015](#)).

Analytic idealism

Bernardo Kastrup proposes “analytic idealism” as a model for reality, in which the ground of existence is universal phenomenal consciousness ([Kastrup, 2021](#)). Analytic idealism is a metaphysics that postulates consciousness as Nature’s sole fundamental ground and that all natural phenomena are ultimately reducible to universal consciousness. He describes phenomenal consciousness as a raw subjective experience of awareness that differs from cognition, meta-cognition, self-awareness, or other higher mental functions. Meta-cognition allows humans to know that they are having an experience and also supports cognitive properties like reasoning and planning. Experiential consciousness or pure awareness can also occur without meta-cognition, as reported in classical mystical states. Because there is only one universal consciousness, individuated living beings are described as dissociated mental complexes of the “fundamentally unitary universal mind” ([Kastrup, 2021](#), p. 267). This dissociation creates a subjective private inner world that can perceive itself as interacting with the transpersonal world. Matter in this model is described as the outward appearance of the inner experience as observed from across the dissociative boundary. Put another way,

As experienced from the inside—that is, from the first-person perspective—each living being, plus the inanimate universe as a whole, is a conscious entity. But as experienced from outside—that is, from a[n illusory] second- or third-person perspective—our respective inner lives present themselves in the form of what we call matter, or physicality...all matter—is merely the name we give to what conscious inner life looks like from across its dissociative boundary. ([Kastrup, 2021](#), pp. 267–268)

Triadic dimensional vortical paradigm

Vernon Neppe and Ed Close propose that the standard 4-dimensional model of physics (three dimensions of space and one of time) results in many contradictions or unexplained discrepancies (see [Neppe and Close, 2020](#) for examples of apparent discrepancies). For example, using the Diophantine equation (a polynomial equation involving two or more unknowns and in which only integer solutions are allowed), Neppe states that the mass/energy of up-quarks and down-quarks produces an inequality that is unstable ([Neppe and Close, 2015](#)). To address these discrepancies, Neppe and Close describe a mathematical model in which we exist in a 9-dimensional finite, quantized, volumetric, spinning reality embedded in an infinite continuity (9D+). The model requires an extra component that they dub “gimmel,” which is mass and energy less. Close expresses that “gimmel is the connection between consciousness, life, and the atomic structure and that the potential for conscious life existed in the mutable mass and energy of quarks even before they became the first protium atoms of physical reality” ([Close, 2018](#)). The model proposes that the 4D world we ordinarily experience is the physical component of this 9D+ existence. Neppe and Close believe that the model has been empirically demonstrated with correspondences to normalized data for the mass-energy equivalence volumetric data for measured particles. They also claim that their model is mathematically valid at the micro, macro, and cosmological scales.

Mathematically, gimmel necessarily has to exist in union with any particle in the universe for that particle to be stable. Without gimmel, the spinning (vortical) atoms would be unstable and asymmetrical about their axes and would, in effect, fly apart: Our world and the physical universe could not exist. ([Neppe and Close, 2020](#), p. 4)

Zero-point field

Joachim [Keppler \(2018\)](#) proposes a theory where the energy of the vacuum is the basis for consciousness, the so-called “zero-point field” ([Keppler, 2018](#)). This is a theory of panpsychism where consciousness permeates the universe yet is only concentrated and apparent in certain circumstances. Unlike other panpsychism theories, it is not the “matter” that is conscious but empty space. However, the idea that matter is conscious may be incompatible with theoretical physics. If matter is conscious, particles may have the yet unknown property “consciousness.” Mathematically, particles are elementary because they cannot be assigned additional parameters than those currently assigned (e.g., a charge, spin, mass). Therefore, the idea that they are conscious is challenging to reconcile with physics. The zero-point field does not have the same problem. Keppler hypothesizes that the human brain is one of the physical mediums which can interact directly with the zero-point field by concentrating on it and thus experiencing consciousness. The details of this putative interaction are not currently known. However, the interesting element of this theory is that it leads to testable predictions, e.g., interactions between the brain (maybe through quantum phenomena as in the Orch OR theory), and the zero-point field could possibly be observed and measured. For example, there might be specific types of photon exchanges that would reveal this interaction.

Orchestrated objective reduction theory

The Orch OR theory was developed by Stuart Hameroff and Sir Roger Penrose ([Hameroff, 2021](#); [Hameroff and Penrose, 2014](#)). While the Copenhagen interpretation posits that the collapse of quantum states into a single state (the so-called “collapse of the wave function”) is determined by an observation (i.e., subjective reduction), Penrose’s objective reduction (OR) posits that it occurs when the energy difference (measured by spacetime curvature and mediated by gravity) of these states reaches an objective threshold (called the “Diósi–Penrose criterion”). Random proto-conscious moments of experience occur at each OR moment ([Hameroff, 2021](#), p. 74). At the biological level, this OR would be orchestrated (Orch) by connective proteins (e.g., microtubule-associated proteins; MAPs) that influence this spacetime-separation of the qubits’ superimposed states. These quantum processes are performed by qubits formed on cellular microtubules by oscillating dipoles (the microtubule condensate), forming superposed resonance rings in helical pathways throughout the microtubule lattices. These oscillations are either electric or magnetic and are then amplified by neurons, leading to consciousness. This collective process corresponds to the orchestration of the objective reduction of quantum states in the brain (Orch-OR). The microtubules both influence and are influenced by the conventional synaptic activity of neurons. Hameroff later added that the condensates might travel across more considerable distances in the

brain through dendritic-dendritic gap junctions (connections that allow much faster transfer of action potentials than synapses), generating gamma oscillations (high-frequency brain rhythms) associated with conscious perception, for example. This theory provides a straightforward mechanism that can be tested more easily than others. Experiments are underway to test the theory by evaluating if the proposed quantum interference is, in fact, present in microtubules and dampened by anesthesia ([Kalra et al., 2020](#)).

Schooler hypothesis of subjective time

Psychologist Jonathan Schooler proposes subjective time as a new dimension of physics that would allow us to have a causal effect on the world ([Schooler, 2014](#)). This model proposes that one could conceive of the possibility of alternative dimensions of meta-perspective where each of us could move across time and raises the possibility that consciousness itself could have some causal role. In his model, a hierarchical cascade of conscious elements would have synchronization happening essentially like carrier waves. The lower level of waves has a particular rhythm. They are also synchronized, or cross-coupled, with the higher levels. In the same way that you can have very high-frequency waves or vibrations synced in with lower-frequency ones, in a sort of cross-coupling manner, you could also have the rhythms of the lower-level ones connected up to the higher-level ones. Through cross-frequency coupling, there potentially exists both top-down and bottom-up paths, explaining consciousness at a macroscopical level.

Theory of double causality

Philippe Guillemant, a theoretical physicist, has proposed that trajectories between two spacetimes are not fixed within the block universe ([Guillemant and Medale, 2019](#)). The block universe is a model where the future is already realized and is implied by general relativity. Within this framework, Guillemant proposes a non-deterministic model of the block universe where consciousness and free will are mechanisms by which the exact path between two spacetime points is decided. He shows that this does not contradict the equations of physics. He also suggests that the irreversibility of time as we experience it might not be a fundamental property of the world but a statistical one. Statistically, time moves forward, but there might be rare instances where it could move backward. Similarly, he suggests that there might be future traces in the present. Although statistically, we will mostly see causal traces of the past, future traces may be experienced as observations of coordinated systems that past observations cannot explain. For example, one might observe an organized pattern that is not due to a specific causal effect in the past. In his model, he argues that the organization must come from the future as it has no causal past reason to exist. He states, “We can carry on doing physics, but we must be absolutely logical about it, considering our intentions as physical realities, with the added ingredient that they do not appear to depend solely on our brains but also on an information system outside spacetime” ([Guillemant, 2016](#), p. 9).

Summary of non-local consciousness models

Most of these theories assume that consciousness is fundamental and primary to all else. Our subjective intersection with this fundamental consciousness is described in different ways, such as being an interface, a dissociative boundary, or a consciousness unit. Moreover, the mechanistic structure of our world with consciousness as fundamental is explained in various ways (e.g., dimensions, conscious agents, *gimmel*).

However, it is important to note that physicalist theories still have a place in this framework. Even if consciousness is fundamental, these theories will inform on the mechanisms for the embodiment of consciousness into this materialistic reality (e.g., how the interface works). If we can perceive non-local information (as observed at the quantum scale), we likely still need to filter out the noise from the environment through uncertainty reduction, broadcast, and top-down processes for that information to become conscious. Predictive processes and updating the error prediction might be a crucial process to allow the perception of non-local information.

Another important point is that the IIT model could be a tool to study both physicalist and non-local theories of consciousness by including non-local properties into the spacetime postulates. In Section “Physicalist theories of consciousness,” we placed IIT as a physicalist theory of consciousness in the sense that it excludes non-local spacetime properties into the spacetime boundaries required for a physical system to be conscious, and all models are based on the conventional assumptions of spacetime. However, since IIT is only about information and systems, one may be able to test IIT for non-local consciousness. Spacetime properties could be included in the postulates (i.e., requirements for a physical substrate to be conscious) for the calculation of ϕ (e.g., quantum links between past and future) to see how this addition affects ϕ 's value. These non-local applications of IIT would allow for the non-local effects observed in quantum mechanics and the literature reviewed in Section “Phenomena suggested by a model of nonlocal consciousness.”

Just like physicalist theories need rigorous testing to validate them, non-local consciousness theories also need testing and validating. The key to fully validating a theory of consciousness (physicalist or non-local) is to make a prediction that can be experimentally tested and quantified, thus, validating or invalidating the prediction. Theories that cannot meet the prediction can be rejected or adjusted. Unfortunately, many theoretical predictions are challenging to test experimentally, and sometimes prediction confirmation might depend on future technological innovations. Often, the theory is built with abstract terms that need further precision and elaboration. The more precise the theory and the prediction, the more it lends itself to testing. Also, the theory may be demonstrated with mathematics and yet not currently be experimentally validated.

One very small step to explore the applicability of the concept of non-local consciousness models and the motivation for developing these models in the first place is driven by phenomena that are not accounted for by physicalist theories, as described in the next section. One reason that non-local consciousness models may be useful is

that they allow for the possibility of the subjective experiences that are usually considered impossible by physicalist models or simply ignored because of the basic assumptions on which they are built.

Phenomena suggested by a model of non-local consciousness

In the next section, we propose specific phenomena that we would expect to see if non-local consciousness theories are correct.

Phenomenon #1: Perceiving information about distant locations

If consciousness were non-local, then an individual ought to be able to perceive information beyond the reach of the brain, body, and senses. For example, one might be able to gain information about a person, place, or object at a distant location. Such abilities are described as part of a classified US government program that ran from 1972 to 1995, which sought to use non-local consciousness for espionage ([May and Marwaha, 2018](#)). That program conducted over 500 operational missions, some of which are said to have resulted in actionable intelligence and also several hundred controlled experimental trials. The latter was evaluated by a professor of statistics and skeptical psychology professional. Both concluded that the evidence in those studies was statistically significant and could not be attributed to methodological flaws ([Mumford et al., 1995](#); [Utts, 2016](#)). In a typical experimental session, a “viewer” would enter a relaxed state. An interviewer would give them a random number designating the desired target and then ask them to describe and/or draw any information they perceived about that target. Both viewer and interviewer were blind to the target. Multiple meta-analyses of public domain and declassified experiments of this type have been conducted, and the results showed highly positive evidence in favor of a genuine phenomenon ([Milton, 1997](#); [Dunne and Jahn, 2003](#); [Baptista et al., 2015](#); [Cardeña, 2018](#)). This apparent ability is now used for other practical applications, such as predicting stock market movements ([Harary and Targ, 1985](#); [Kolodziejzyk, 2013](#); [Smith et al., 2014](#)), locating missing persons ([Mcmoneagle and May, 2004](#)), and finding previously unknown archaeological sites ([Schwartz, 2005, 2019](#)).

Phenomenon #2: Perceiving information from another person

If consciousness were non-local, an individual might be able to receive information about another, isolated person’s mental activity from a distance. Numerous well-controlled laboratory studies have observed this apparent phenomenon using the *ganzfeld* protocol, one of the most-repeated non-local consciousness studies. Ganzfeld originates from a German word meaning “whole field,” and Gestalt psychologists initially developed the protocol. First, a person is exposed to low-level, unpatterned sensory stimuli (e.g., red light diffused to the eyes and white noise played through headphones). Meanwhile, a second, isolated person attempts to mentally “send” a target image randomly selected out of a pool of four possible images, which

was randomly selected out of a database of many such pools. The chance of the “receiving” person correctly selecting the actual image is thus 25%. Over 120 published experiments have used this protocol, comprising about 4,000 individual trials, and the overall hit rate was just over 30%. Multiple reviews and meta-analyses on this protocol have also been conducted ([Storm et al., 2010](#); [Baptista et al., 2015](#); [Cardeña, 2018](#); [Storm and Tressoldi, 2020](#)). These results have been discussed and debated in one of the principal journals in academic psychology, *Psychological Bulletin* ([Bem and Honorton, 1994](#); [Hyman, 2010](#); [Storm et al., 2010](#)).

In a conceptually similar design, rather than testing whether one person could select a correct image sent by another isolated person, the person’s unconscious physiological state was intentionally influenced by a second person who was asked to focus their attention on them. These studies have typically used measures such as electrodermal activity ([Braud and Schlitz, 1983](#); [Radin et al., 2008](#)), electroencephalography (EEG) activity ([Standish et al., 2004](#); [Richards et al., 2005](#)), and functional magnetic resonance imaging ([Standish et al., 2003](#); [Achterberg et al., 2005](#)). To date, there have been three meta-analyses for this class of studies, with each reporting statistically significant outcomes ([Schmidt et al., 2004](#); [Schmidt, 2012](#), [2015](#)). Using this experimental paradigm, researchers discovered that the prior beliefs of the investigators were an important element in the observed outcomes. That is, working with the same subject populations, protocol, equipment, and analyses, skeptical investigators obtained null results, but investigators more open to the possibility of an effect obtained significant results ([Watt et al., 2002](#); [Schlitz et al., 2006](#)). These investigator-specific effects have been documented in psychology and are called “experimenter effects” ([Palmer and Millar, 2015](#)). Thus, it is challenging to ascertain if results are solely influenced by the experimenter effect (i.e., intentions of the investigator) or if there are intrinsic effects. Multiple-experimenter studies have been posed as a solution to solving this issue in psychological studies ([Bierman and Jolij, 2020](#)).

Phenomenon #3: Perceiving the future

If consciousness were non-local, one might be able to perceive information from non-inferable future events. Experiments testing this idea have shown that people’s physiology has reacted to randomly selected future events ([Radin and Pierce, 2015](#)), including electrodermal ([Radin, 1997](#)) and electrocortical activity ([Radin and Lobach, 2007](#); [Radin and Borges, 2009](#); [Radin et al., 2011](#)), and heart rate ([McCraty et al., 2004](#); [Tressoldi et al., 2009](#)). These laboratory studies apparently demonstrate that the body can react to randomly selected stimuli approximately 1–10 s in the future. Erotic and negative images produce more robust responses than emotionally neutral pictures, and pre-responses generally manifest in the same direction as the body would typically respond after exposure to a stimulus. Meta-analyses have evaluated multiple laboratory studies with positive effect sizes ([Mossbridge et al., 2012](#), [2014](#); [Storm et al., 2012](#); [Mossbridge and Radin, 2018](#); [Honorton et al., 2018](#)). For example, [Mossbridge et al. \(2012\)](#) analyzed 26 studies where unpredictable stimuli were presented and physiological activity was collected before, during, and after the stimuli. There was a pre-stimulus effect demonstrating a physiological response prior to the unpredictable

stimuli (fixed effect: overall effect size = 0.21, 95% CI = 0.15 – 0.27, $z = 6.9$, $p < 2.71 \times 10^{-12}$; [Mossbridge et al., 2012](#)).

Implicit bias tests with a retrocausal element provide similar findings. In one paradigm, a classic perceptual priming task was reversed, such that the prime occurred after the target images. For example, in one task, the prime “happy” might typically occur prior to the target picture of a flower. In a reverse priming task, the flower image would appear *before* the prime “happy.” These reverse priming tasks found slower response times when the prime/target pairs were incongruent (sad/flower) versus congruent (happy/flower), just as the classic task would, even though the prime occurred *prior* to the target. Some 90 independent replications of these experiments have provided evidence for a highly significant effect (overall effect size = 0.09, $z = 6.4$, $p = 1.2 \times 10^{-10}$; [Bem et al., 2015](#)).

Phenomenon #4: Apparent cognitive abilities beyond the experience/learning/skill of the person exhibiting them

If consciousness were non-local, then people might be able to gain cognitive skills without previous experience or training in those skills.

An example is the phenomenon of a person speaking a language unknown to, or xenoglossy. This phenomenon has been reported since ancient times. It refers to the ability of an individual to speak or write a language that they presumably did not know and could not have acquired by ordinary means. For example, in 400 BC, Plato mentions priestesses on the Island of Delos who spoke “in tongues.” There are also descriptions in the Bible (Corinthians 14:1-40 and Acts 2:4).

Another example is Indriði Indriðason (1883–1912), who apparently spoke multiple languages he did not know ([Haraldsson, 2012](#)). Similarly, Alec Harris spoke at length to witness Sir Alexander Cannon in Hindustani and Tibetan, two languages that Harris would have had no way of knowing, but Sir Alexander did know ([Vandersande, 2008](#), p. 113). Other xenoglossy cases have also been documented by University of Virginia scientist Ian Stevenson ([Stevenson and Pasricha, 1979, 1980](#)). While anecdotal and subject to the known biases of experiential reports, these cases have been meticulously well-documented. Similar cases of “acquired” and “spontaneous savants” refer to individuals who, either through a traumatic event or with no apparent cause at all, suddenly gain exceptional musical or mathematical skills ([Treffert, 2009](#)).

Phenomenon #5: Non-local consciousness experiences are common

If consciousness were non-local, such experiences would be highly prevalent in all humans. And indeed, non-local experiences can be found throughout history, across all cultures, and at all educational levels. Formal prevalence studies have been conducted for almost 50 years, with rates ranging from 10% to 97%, depending on the population surveyed ([Bourguignon, 1976](#); [Palmer, 1979](#); [Haraldsson, 1985, 2011](#); [Greeley, 1987](#); [Haraldsson and Houtkooper, 1991](#); [Ross and Joshi, 1992](#); [McClenon, 1993](#); [Cohn, 1994](#); [Castro et al., 2014](#); [Wahbeh et al., 2018](#)). Another survey of the

general public, scientists, and engineers in the United States found that over 90% had experienced at least one of 25 of these experiences ([Wahbeh et al., 2018](#)). With prevalence rates being well-above 10% of most populations surveyed, it is evident that these phenomena, at least in their subjective reports, are more frequent than commonly supposed.

Phenomenon #6: Cognitive abilities can be retained when the brain is seriously compromised

We usually assume that the brain is the body's driver, and if the brain is not working well, the body should not work. Suppose this is wrong and consciousness is not entirely dependent on the physical function of the brain. In that case, cognition, perception, and memory may continue to operate normally even when the brain would not be considered functional. This is consistent with what we see in a phenomenon called *terminal lucidity*. Terminal lucidity is a label given to a phenomenon in which patients with terminal neurodegenerative conditions display apparently normal cognitive function and mental clarity during the period preceding death (hours to days). While such experiences would seem impossible based on known principles of neuroscience and neuroanatomy, they have been reported in the medical literature for over 250 years ([Nahm et al., 2012](#)).

Terminal lucidity, also called paradoxical lucidity, has occurred in conditions such as waking from a long-term coma, dementia due to advanced Alzheimer's disease, brain abscesses, tumors, strokes, and meningitis ([Nahm et al., 2012](#)). A recent study of terminal lucidity reviewed 124 cases in dementia patients and found that in "more than 80% of these cases, complete remission with the return of memory, orientation, and responsive verbal ability was reported by observers of the lucid episode" ([Batthyány and Greyson, 2021](#)). For example, one reported case involved a patient with cancer that had metastasized to the brain, with little functional brain tissue remaining. However, an hour before the patient died, he regained awareness and conversed with his family for about 5 min before passing away ([Nahm et al., 2012](#)). Most terminal lucidity cases are retrospective case reports ([Kelly et al., 2007](#); [Nahm and Greyson, 2009](#); [Nahm et al., 2012](#); [Mashour et al., 2019](#); [Batthyány and Greyson, 2021](#)), but a few are prospective. Macleod and colleagues prospectively observed terminal lucidity cases ([Macleod, 2009](#)), as did Fenwick and colleagues ([Fenwick et al., 2010](#)). In these cases, the patients demonstrated normal cognitive abilities just prior to death, contrary to what objective medical findings would have predicted (e.g., EEG, neuroimaging). These patients are operating in an anomalous manner that brings into question the idea that the body is a "puppet" controlled from the inside (the brain) and that perhaps it can function alternately in some instances. Perhaps there are aspects of consciousness that could be "outside" of the body controlling it. The lucid mental functioning associated with these patients' behavior is challenging to explain under the assumption that one's sense of identity, memory, and awareness solely depends on brain activity.

Summary

In sum, we presented six phenomena regarding aspects related to non-local consciousness. Reports of individuals perceiving information from distant locations, from another person, from the future, where people gain skills beyond their normal capacity, or when the brain is apparently non-functional, have been documented in anecdotal and experimental contexts. In addition, these phenomena are ubiquitous worldwide. Note that these examples are not meant to provide definitive evidence for non-local consciousness, nor provide a comprehensive list of such phenomena, but rather to highlight that certain commonly reported phenomena, and some rare effects, present clear challenges to prevailing physicalist models of consciousness.

Of course, given the significant theoretical importance of these phenomena, each example supporting these predictions has evoked critical responses. The critiques have tended to fall into two classes. First, the phenomena suggested by these examples are deemed impossible because they violate the basic limiting principles of science. Therefore, the only possible way to interpret experiments reporting positive results is that they most likely involve flaws, fraud, or both. Critical reactions to anecdotal reports have also tended to focus on their subjective nature and the many ways that such experiences can be misinterpreted as illusions, misperceptions, or distorted memories. Such critiques can be answered by pointing out that some of the anecdotal reports involved hundreds to thousands of documented case studies, and all the experiments mentioned involved controlled experimental paradigms that were repeated in multiple laboratories and dozens to over a hundred independent replications, with overall highly significant meta-analytic outcomes ([Cardeña, 2018](#)). In some of the earliest experiments, methodological flaws were discovered but later corrected with similar results, so insisting that flaws or fraud can be the only possible explanations is not supported by analysis of the data.

The second category of critique is that perhaps the results could be accounted for by one or more physicalist explanations that we do not understand yet, given the state of the science. For example, perhaps some material explanation will eventually arise for how someone with severe brain atrophy and neurofibrillary tangles, or who was in a deep coma for an extended period, could nevertheless suddenly become lucid and maintain a coherent conversation with loved ones shortly before death. Alternatively, perhaps if it is established that the brain has quantum biological properties, then that might provide a plausible substrate for perceptual non-locality. That is, a brain that is partially acting in a quantum manner could possibly account for all these anomalous phenomena. A quantum brain would have non-local properties, so our sensory system would be spread out in space and time, and it might also have observational properties. However, even if this was true, it would not tell us anything about the nature or source of our subjective awareness. That is, from the quantum brain perspective, these phenomena would be completely explained as a purely physical phenomena (albeit within the context of the not-quite-physical nature of the quantum world).

The scientific process and perspective in the face of a paradigm shift

Our call to test non-local consciousness theories is not a proposal to discard physicalist theories. There is no question that, as a set of assumptions, materialism has proven to be outstandingly successful in elucidating the nature of physical reality, and it will likely continue to be useful. However, the phenomena we have highlighted here bring some level of doubt to the ability of physicalist theories to explain *everything*, including the nature, origin, and capacities of consciousness. Here we propose that materialism be viewed as a special case of a more comprehensive metaphysic, one that includes consciousness in some fundamental way. This approach is akin to regarding classical physics as a special case, one that describes a limited domain of the physical world. Quantum mechanics, too, is probably a special case because, so far, it is not compatible with relativistic physics. These “modern” physical theories are more comprehensive than classical physics and are special cases.

Promoting the value of more comprehensive models of reality can be challenging. As Max Planck said, “A new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it” ([Planck, 1950](#), p. 33). Researchers testing if this statement was true found that, indeed, acceptable scientific models and concepts became more varied after leaders in their field died ([Azoulay et al., 2019](#)).

A classic example of shifts in worldviews in science is the case of black holes. Imagine it is 1921, and we asked, “Do black holes exist?” In 1915, Karl Schwarzschild solved Einstein’s equations of general relativity for the limited case of a single spherical non-rotating mass. In the process, he discovered the possibility that under extreme gravitational conditions, space could collapse upon itself. Einstein denied that these “black holes” could possibly form. In 1939, he published a paper arguing that a star collapsing would spin faster and faster, eventually spinning at the speed of light with infinite energy, well before the point that it would collapse into a singularity. It was not until the 1960s when Roger Penrose published more detailed models showing how black holes could form, that other physicists considered them viable. A half-century later, astronomers finally observed a black hole (The Event Horizon Telescope Collaboration). In fact, a team of Harvard scientists just released an image of Sagittarius A-star, a black hole at the center of our Milky Way galaxy ([McDermott-Murphy, 2022](#)).

We propose that today’s understanding of the evidence for non-local consciousness is similar to what was understood about black holes in 1921. A jury of leading scientists in 1921 charged with deciding if black holes existed would have weighed the pros and cons of existing theory and data, they would have consulted with Einstein, and they would have almost certainly decided that black holes could not exist and therefore did not exist. As we know today, they would have been wrong.

Conclusion

Intriguing phenomena are alluding to consciousness being associated with – but not limited by – brain activity. We are in a position similar to those who were studying the possibility of black holes a century ago. Perhaps in 50 years, we will look back on the current transitional period between materialistic and post-materialistic paradigms in science and more clearly understand why we could not have possibly grasped the whole picture.

We can learn from the black hole example to release our desire to prove non-local consciousness and instead remain in a state of curiosity, focusing on methods and improved measures. Even if it were not possible to definitively demonstrate that consciousness is non-local but in the process of determining that, it was discovered that there were non-local aspects of consciousness that we learned more about and controlled to some extent, our world would be radically transformed with the shift in the understanding of our capacities and its practical applications. The systematic scientific study of consciousness is still in its infancy, and thus, we are at the very beginning of understanding the right questions to ask.

This review also calls for humility, open-mindedness, and collaboration in science. Is it possible to remain neutral about the various theories of consciousness? Perhaps physicalist theories will be tested and shown to be relevant in particular situations. Perhaps non-materialist theories will also be tested and shown to be valid in other situations. Could it be that multiple theories of consciousness are tested and found viable? If so, what would that mean about the nature of reality? Can these theories be evaluated for similarities and differences, perhaps combining some and ultimately testing them? Templeton World Charity Foundation Program 'Accelerating Research on Consciousness' has spearheaded such an initiative for physicalist theories. The same could be implemented for non-physicalist theories ([Templeton World Charity Foundation, 2022](#)). Is there an interaction between a non-local consciousness interfacing with the physical and/or quantum brain that is persuasively describable? Remaining open and flexible about these possibilities is essential in supporting the birth of new ideas. Remaining humble allows us to review other theories without prejudice.

To further assess the vast number of theories of consciousness, physicalist and non-local, we invite theorists to attempt to make their theories increasingly precise so that abstract terms become quantifiable predictions that can be confirmed or refuted. Furthermore, theorists could attempt to use similar language/terms to improve the clarity regarding the distinctions and commonalities across theories. Criteria could be developed, allowing researchers to easily determine the nature/aspect of consciousness discussed by the theory, the proposed processes that explain how non-local consciousness may interact with physical substrates, and the precise predictions to validate it. Physicalist theories might be closer to validating or invalidating their predictions if the assumption about the nature of consciousness being generated from the brain is correct. However, these predictions may never address the possibility that consciousness is a fundamental property of reality with non-local properties (i.e., they address physical mechanisms but not the nature of consciousness itself).

In conclusion, our reported phenomena of non-local consciousness present intriguing examples that should be addressed when evaluating whether consciousness may be more than an emergent property of brain activity. Despite sophisticated physicalist theories of consciousness dependent on brain function, these examples apparently demonstrate non-local aspects of consciousness, perceiving information in a way that is not limited by our conventional understanding of time and space and that is not dependent on the brain function. Many of these data have been observed with objective measures in the laboratory in a valid and reliable way or collected in the field with impeccable methods and exclusion of fraud. While materialism explains much in our world, it does not explain everything, including these phenomena. Non-materialist theories encompassing consciousness as fundamental and/or non-local may provide a pathway to understanding these phenomena. Perhaps holding the hypothetical assumption that consciousness *is* fundamental and focusing on what we can learn about the mechanism, mediators, moderators, and practical applications of non-local consciousness will reveal novel areas to explore.

Author contributions

HW contributed to the conceptualization, funding acquisition, project administration, and writing – original draft preparation, review, and editing. DR and AD contributed to the conceptualization, funding acquisition, and writing – original draft preparation, review, and editing. CC contributed to the conceptualization and writing – review and editing. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Bigelow Institute of Consciousness Studies and the Institute of Noetic Sciences and its members. The funders did not play a role in the design or implementation of the present study.

Acknowledgments

The authors thank the IONS science team, staff, and members for their support of this project.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Achterberg, J., Cooke, K., Richards, T., Standish, L. J., Kozak, L., and Lake, J. (2005). Evidence for correlations between distant intentionality and brain function in recipients: A functional magnetic resonance imaging analysis. *J. Altern. Complement. Med.* 11, 965–971. doi: 10.1089/acm.2005.11.965

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Albantakis, L. (2020). “Integrated information theory,” in *Beyond neural correlates of consciousness*, eds A. Kirkeby-Hinrup, J. Mogensen, and M. Overgaard (Abingdon: Routledge).

[Google Scholar](#)

Albantakis, L., Hintze, A., Koch, C., Adami, C., and Tononi, G. (2014). Evolution of integrated causal structures in animats exposed to environments of increasing complexity. *PLoS Comput. Biol.* 10:e1003966. doi: 10.1371/journal.pcbi.1003966

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Aru, J., Bachmann, T., Singer, W., and Melloni, L. (2012). Distilling the neural correlates of consciousness. *Neurosci. Biobehav. Rev.* 36, 737–746. doi: 10.1016/j.neubiorev.2011.12.003

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Azoulay, P., Fons-Rosen, C., and Zivin, J. S. G. (2019). Does science advance one funeral at a time? *Ame. Econ. Rev.* 109, 2889–2920. doi: 10.1257/aer.20161574

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Baars, B. J. (2005). Global work-space theory of consciousness: Toward a cognitive neuroscience of human experience. *Prog. Brain Res.* 150, 45–53. doi: 10.1016/S0079-6123(05)50004-9

[CrossRef Full Text](#) | [Google Scholar](#)

Baars, B. J., Geld, N., and Kozma, R. (2021). Global Workspace Theory (GWT) and prefrontal cortex: Recent developments. *Front. Psychol.* 12:749868. doi: 10.3389/fpsyg.2021.749868

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Baptista, J., Derakhshani, M., and Tressoldi, P. E. (2015). "Explicit anomalous cognition: A review of the best evidence in ganzfeld, forced choice, remote viewing and dream studies," in *Parapsychology: A handbook for the 21st century*, eds E. Cardeña, J. Palmer, and D. Marcusson-Clavertz (Jefferson, NC: McFarland & Company, Inc., Publishers), 192–214.

[Google Scholar](#)

Barrett, L. F. (2016). The theory of constructed emotion: An active inference account of interoception and categorization. *Soc. Cogn. Affect. Neurosci.* nsw154. doi: 10.1093/scan/nsw154

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Batthyány, A., and Greyson, B. (2021). Spontaneous remission of dementia before death: Results from a study on paradoxical lucidity. *Psychol. Conscious.* 8, 1–8. doi: 10.1037/cns0000259

[CrossRef Full Text](#) | [Google Scholar](#)

Bem, D. J., and Honorton, C. (1994). Does psi exist? Replicable evidence for an anomalous process of information transfer. *Psychol. Bull.* 115, 4–18. doi: 10.1037/0033-2909.115.1.4

[CrossRef Full Text](#) | [Google Scholar](#)

Bem, D. J., Tressoldi, P., Rabeyron, T., and Duggan, M. (2015). Feeling the future: A meta-analysis of 90 experiments on the anomalous anticipation of random future events. *F1000Res.* 4:1188. doi: 10.12688/f1000research.7177.2

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Bierman, D., and Jolij, J. J. (2020). Dealing with the experimenter effect. *J. Sci. Explor.* 34, 703–709. doi: 10.31275/20201871

[CrossRef Full Text](#) | [Google Scholar](#)

Blumenfeld, H., and Taylor, J. (2003). Why do seizures cause loss of consciousness? *Neuroscientist* 9, 301–310. doi: 10.1177/1073858403255624

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Bourguignon, E. (1976). *Possession*. Novato, CA: Chandler & Sharp Publishers.

[Google Scholar](#)

Braud, W., and Schlitz, M. (1983). Psychokinetic influence on electrodermal activity. *J. Parapsychol.* 47, 95–119.

[Google Scholar](#)

Cardaña, E. (2018). The experimental evidence for parapsychological phenomena: A review. *Am. Psychol.* 73, 663–677. doi: 10.1037/amp0000236

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Castro, M., Burrows, R., and Wooffitt, R. (2014). The paranormal is (still) normal: The sociological implications of a survey of paranormal experiences in Great Britain. *Sociol. Res. Online* 19:16. doi: 10.5153/sro.3355

[CrossRef Full Text](#) | [Google Scholar](#)

Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. Oxford: Oxford University Press.

[Google Scholar](#)

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36, 181–204. doi: 10.1017/S0140525X12000477

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Cleeremans, A., Achoui, D., Beauny, A., Keuninckx, L., Martin, J.-R., Muñoz-Moldes, S., et al. (2020). Learning to Be Conscious. *Trends Cogn. Sci.* 24, 112–123. doi: 10.1016/j.tics.2019.11.011

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Close, E. R. (2018). *Gimmel, life and consciousness*. Available online at: <http://www.erclosetphysics.com/2018/05/gimmel-life-and-consciousness.htmlAD> (accessed May 27, 2022).

[Google Scholar](#)

Cohn, S. A. (1994). A survey on Scottish second sight. *J. Soc. Psych. Res.* 59, 385–400. doi: 10.1001/jamaoncol.2021.6987

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Currihan, J. (2017). *The cosmic hologram: In-formation at the center of creation*. New York, NY: Simon and Schuster.

[Google Scholar](#)

de Graaf, T. A., Hsieh, P.-J., and Sack, A. T. (2012). The “correlates” in neural correlates of consciousness. *Neurosci. Biobehav. Rev.* 36, 191–197. doi: 10.1016/j.neubiorev.2011.05.012

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Dehaene, S., Charles, L., King, J.-R., and Marti, S. (2014). Toward a computational theory of conscious processing. *Curr. Opin. Neurobiol.* 25, 76–84.

[Google Scholar](#)

Dehaene, S., Naccache, L., Cohen, L., Bihan, D. L., Mangin, J.-F., Poline, J.-B., et al. (2001). Cerebral mechanisms of word masking and unconscious repetition priming. *Nat. Neurosci.* 4, 752–758. doi: 10.1038/89551

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Dehaene, S., Sergent, C., and Changeux, J.-P. (2003). A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proc. Natl. Acad. Sci. U.S.A.* 100, 8520–8525.

[Google Scholar](#)

Dossey, L. (1994). Healing and the mind: Is there a dark side? *J. Sci. Explor.* 8, 73–90.

[Google Scholar](#)

Dunne, B. J., and Jahn, R. G. (2003). Information and uncertainty in remote perception research. *J. Sci. Explor.* 17, 207–241.

[Google Scholar](#)

Faggin, F. (2021a). “Consciousness Comes First,” in *Consciousness unbound: Liberating mind from the tyranny of materialism*, eds E. F. Kelly and P. Marshall (Lanham, MD: Rowman & Littlefield Publishers), 283–322.

[Google Scholar](#)

Faggin, F. (2021b). *Silicon: From the invention of the microprocessor to the new science of consciousness*. Sea, CA: Waterside Productions.

[Google Scholar](#)

Fenwick, P., Lovelace, H., and Brayne, S. (2010). Comfort for the dying: Five year retrospective and one year prospective studies of end of life experiences. *Arch. Gerontol. Geriatr.* 51, 173–179.

[Google Scholar](#)

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Gershman, S. J. (2019). The generative adversarial brain. *Front. Artif. Intell.* 2:18. doi: 10.3389/frai.2019.00018

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Goff, P. (2019). *Galileo's error: Foundations for a new science of consciousness*. New York, NY:: Pantheon Books.

[Google Scholar](#)

Greeley, A. (1987). Mysticism goes mainstream. *Am. Health* 7, 47–49.

[Google Scholar](#)

Grossberg, S. (2013a). Adaptive resonance theory. *Scholarpedia* 8:1569. doi: 10.4249/scholarpedia.1569

[CrossRef Full Text](#) | [Google Scholar](#)

Grossberg, S. (2013b). Adaptive resonance theory: How a brain learns to consciously attend, learn, and recognize a changing world. *Neural Netw.* 37, 1–47. doi: 10.1016/j.neunet.2012.09.017

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Guillemant, P. (2016). *Theory of double causality. Time matters review*. Available online at: [http://www.guillemant.net/english/Theory of Double Causality.pdf](http://www.guillemant.net/english/Theory_of_Double_Causality.pdf)

[Google Scholar](#)

Guillemant, P., and Medale, M. (2019). A unique but flexible space–time could challenge multiverse theories. *Ann. Phys.* 409:167907. doi: 10.1016/j.aop.2019.05.005

[CrossRef Full Text](#) | [Google Scholar](#)

Hameroff, S. (2021). ‘Orch OR’ is the most complete, and most easily falsifiable theory of consciousness. *Cogn. Neurosci.* 12, 74–76. doi: 10.1080/17588928.2020.1839037

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Hameroff, S., and Penrose, R. (2014). Consciousness in the universe: A review of the ‘Orch OR’ theory. *Phys. Life Rev.* 11, 39–78. doi: 10.1016/j.plrev.2013.08.002

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Haraldsson, E. (1985). Representative national surveys of psychic phenomena: Iceland, Great Britain, Sweden, USA and Gallup's multinational survey. *J. Soc. Psych. Res.* 53, 145–158.

[Google Scholar](#)

Haraldsson, E. (2011). Psychic experiences a third of a century apart: Two representative surveys in Iceland with an international comparison. *J. Soc. Psych. Res.* 75:76.

[Google Scholar](#)

Haraldsson, E. (2012). Further facets of Indridi indridason's mediumship, including "transcendental" music, direct speech, xenoglossy and light phenomena. *J. Soc. Psych. Res.* 76, 129–149.

[Google Scholar](#)

Haraldsson, E., and Houtkooper, J. M. (1991). Psychic experiences in the multinational human values study: Who reports them. *J. Am. Soc. Psych. Res.* 85, 145–165.

[Google Scholar](#)

Harary, K., and Targ, R. (1985). A new approach to forecasting commodity futures. *Psi Res.* 4, 79–88.

[Google Scholar](#)

Haun, A., and Tononi, G. (2019). Why does space feel the way it does? Towards a principled account of spatial experience. *Entropy* 21:1160. doi: 10.3390/e21121160

[CrossRef Full Text](#) | [Google Scholar](#)

Hebb, D. O., and Penfield, W. (1940). Human behavior after extensive bilateral removal from the frontal lobes. *Arch. Neurol. Psychiatry* 44, 421–438. doi: 10.1001/archneurpsyc.1940.02280080181011

[CrossRef Full Text](#) | [Google Scholar](#)

Hoffman, D. D. (2014). The origin of time in conscious agents. *Cosmology* 18, 494–520.

[Google Scholar](#)

Hoffman, D. D. (2019). *The case against reality: Why evolution hid the truth from our eyes*. New York, NY: W.W. Norton & Company.

[Google Scholar](#)

Hoffman, D. D., Singh, M., and Prakash, C. (2015). The interface theory of perception. *Psychon Bull. Rev.* 22, 1480–1506. doi: 10.3758/s13423-015-0890-8

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Hohwy, J. (2013). *The predictive mind*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199682737.001.0001

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Hohwy, J., and Seth, A. (2020). Predictive processing as a systematic basis for identifying the neural correlates of consciousness. *Psyarxiv* [Preprint]. doi: 10.33735/phimisci.2020.II.64

[CrossRef Full Text](#) | [Google Scholar](#)

Honorton, C., Ferrari, D. B., and Hansen, G. (2018). “Meta-analysis of Forced-Choice Precognition Experiments (1935–1987),” in *The star gate archives: Reports of the united states government sponsored psi program, 1972-1995. volume 2: Remote viewing, 1985-1995*, eds E. C. May and S. B. Marwaha (Jefferson, NC: McFarland & Company, Inc., Publishers), 291.

[Google Scholar](#)

Huang, Y., and Rao, R. P. N. (2011). Predictive coding. *Wiley Interdiscip. Rev. Cogn. Sci.* 2, 580–593. doi: 10.1002/wcs.142

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Hudetz, A. G. (2012). General anesthesia and human brain connectivity. *Brain Connect.* 2, 291–302. doi: 10.1089/brain.2012.0107

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Hyman, R. (2010). Meta-analysis that conceals more than it reveals: Comment on storm et al. (2010). *Psychol. Bull.* 136, 486–490. doi: 10.1037/a0019676

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Kalra, A. P., Hameroff, S., Tuszynski, J., Dogariu, A., Sachin, N., and Gross, P. J. (2020). *Anesthetic gas effects on quantum vibrations in microtubules – Testing the Orch OR theory of consciousness*. OSF. Available online at: <https://osf.io/zqnjd/> (accessed May 5, 2022).

[Google Scholar](#)

Kastrup, B. (2017). Self-transcendence correlates with brain function impairment. *J. Cogn. Neuroethics* 4, 33–42.

[Google Scholar](#)

Kastrup, B. (2021). "Analytic idealism and psi: How a more teneable metaphysics neutralizes a physicalist taboo," in *Consciousness unbound: Liberating mind from the tyranny of materialism*, eds E. F. Kelly and P. Marshall (Lanham: Rowman & Littlefield Publishers), 257–283.

[Google Scholar](#)

Kelly, E. W., Greyson, B., and Kelly, E. F. (2007). "Unusual experiences near death and related phenomena," in *Irreducible mind: Toward a psychology for the 21st century*, eds E. F. Kelly, E. W. Kelly, A. Crabtree, A. Gauld, M. Grosso, and B. Greyson (Lanham: Rowman and Littlefield), 367–421.

[Google Scholar](#)

Keppler, J. (2018). The role of the brain in conscious processes: A new way of looking at the neural correlates of consciousness. *Front. Psychol.* 9:1346. doi: 10.3389/fpsyg.2018.01346

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Koch, C. (2018). What is consciousness? *Nature* 557, S8–S12. doi: 10.1038/d41586-018-05097-x

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Koch, C., Massimini, M., Boly, M., and Tononi, G. (2016). Neural correlates of consciousness: Progress and problems. *Nat. Rev. Neurosci.* 17, 307–321. doi: 10.1038/nrn.2016.22

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Kolodziejzyk, G. (2013). Greg Kolodziejzyk's 13-year associative remote viewing experiment results. *J. Parapsychol.* 76, 349–368.

[Google Scholar](#)

Lamme, V. A. F. (2006). Towards a true neural stance on consciousness. *Trends Cogn. Sci.* 10, 494–501. doi: 10.1016/j.tics.2006.09.001

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Lamme, V. A. F. (2010). How neuroscience will change our view on consciousness. *Cogn. Neurosci.* 1, 204–220. doi: 10.1080/17588921003731586

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Lamme, V. A., and Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci.* 23, 571–579. doi: 10.1016/s0166-2236(00)01657-x

[CrossRef Full Text](#) | [Google Scholar](#)

Lau, H. (2019). Consciousness, metacognition, & perceptual reality monitoring. *PsyArXiv* [Preprint]. doi: 10.31234/osf.io/ckbyf

[CrossRef Full Text](#) | [Google Scholar](#)

Lau, H., and Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends Cogn. Sci.* 15, 365–373. doi: 10.1016/j.tics.2011.05.009

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Leopold, D. A., and Logothetis, N. K. (1999). Multistable phenomena: Changing views in perception. *Trends Cogn. Sci.* 3, 254–264. doi: 10.1016/S1364-6613(99)01332-7

[CrossRef Full Text](#) | [Google Scholar](#)

Macleod, A. D. S. (2009). Lightening up before death. *Palliat. Support. Care* 7, 513–516. doi: 10.1017/S1478951509990526

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Mashour, G. A., Frank, L., Batthyany, A., Kolanowski, A. M., Nahm, M., Schulman-Green, D., et al. (2019). Paradoxical lucidity: A potential paradigm shift for the neurobiology and treatment of severe dementias. *Alzheimers Dement. J. Alzheimers Assoc.* 15, 1107–1114. doi: 10.1016/j.jalz.2019.04.002

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Mashour, G. A., Roelfsema, P., Changeux, J.-P., and Dehaene, S. (2020). Conscious processing and the global neuronal workspace hypothesis. *Neuron* 105, 776–798. doi: 10.1016/j.neuron.2020.01.026

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

May, E. C., and Marwaha, S. B. (2018). *The star gate archives: reports of the united states government sponsored psi program, 1972-1995*, Vol. 1. Jefferson, NC: McFarland & Company, Inc., Publishers.

[Google Scholar](#)

McClenon, J. (1993). Surveys of anomalous experience in Chinese, Japanese, and American samples. *Sociol. Relig.* 54, 295–302.

[Google Scholar](#)

McCraty, R., Atkinson, M., and Bradley, R. T. (2004). Electrophysiological evidence of intuition: Part 1. The surprising role of the heart. *J. Altern. Complement. Med.* 10, 133–143. doi: 10.1089/107555304322849057

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

McDermott-Murphy, C. (2022). *The dawn of a new era in astronomy*. *Harvard Gazette*. Available online at: <https://news.harvard.edu/gazette/story/2022/05/the-dawn-of-a-new-era-in-astronomy/>

[Google Scholar](#)

Mcmoneagle, J. W., and May, E. C. (2004). “The possible role of intention, attention and expectation in remote viewing,” in *Proceedings of Presented Papers. The Parapsychological Association 47th Annual Convention*, (Vienna: Vienna University).

[Google Scholar](#)

Milton, J. (1997). Meta-analysis of free-response ESP studies without altered states of consciousness. *J. Parapsychol.* 61, 279–319.

[Google Scholar](#)

Mossbridge, J., and Radin, D. (2018). Precognition as a form of prospection: A review of the evidence. *Psychol. Conscious.* 5:78.

[Google Scholar](#)

Mossbridge, J., Tressoldi, P., and Utts, J. (2012). Predictive physiological anticipation preceding seemingly unpredictable stimuli: A meta-analysis. *Front. Psychol.* 3:390. doi: 10.3389/fpsyg.2012.00390

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Mossbridge, J., Tressoldi, P., Utts, J., Ives, J. A., Radin, D., and Jonas, W. B. (2014). Predicting the unpredictable: Critical analysis and practical implications of predictive anticipatory activity. *Front. Hum. Neurosci.* 8:146. doi: 10.3389/fnhum.2014.00146

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Mumford, M., Rose, A., and Goslin, D. (1995). *An evaluation of remote viewing: Research and applications*. Arlington: American Institutes for Research.

[Google Scholar](#)

Nahm, M., and Greyson, B. (2009). Terminal lucidity in patients with chronic schizophrenia and dementia: A survey of the literature. *J. Nerv. Ment. Dis.* 197, 942–944. doi: 10.1097/NMD.0b013e3181c22583

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Nahm, M., Greyson, B., Kelly, E. W., and Haraldsson, E. (2012). Terminal lucidity: A review and a case collection. *Arch. Gerontol. Geriatr.* 55, 138–142. doi: 10.1016/j.archger.2011.06.031

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Neppe, V. M., and Close, E. R. (2015). Refuting atomic materialism. *IQNexus* 11, 1074–1083.

[Google Scholar](#)

Neppe, V. M., and Close, E. R. (2020). The Neppe-Close triadic dimensional vortical paradigm: An invited summary. *Int. J. Phys. Res. Appl.* 3, 001–004. doi: 10.29328/journal.ijpra.1001018

[CrossRef Full Text](#) | [Google Scholar](#)

Oizumi, M., Albantakis, L., and Tononi, G. (2014). From the phenomenology to the mechanisms of consciousness: Integrated information theory 3.0. *PLoS Comput. Biol.* 10:e1003588. doi: 10.1371/journal.pcbi.1003588

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Palmer, J. (1979). A community mail survey of psychic experiences. *J. Am. Soc. Psych. Res.* 73, 221–251.

[Google Scholar](#)

Palmer, J., and Millar, B. (2015). “Experimenter effects in parapsychological research,” in *Parapsychology: A handbook for the 21st century*, eds E. Cardeña, J. Palmer, and D. Marcusson-Clavertz (Jefferson, NC: McFarland & Company, Inc., Publishers), 293–300.

[Google Scholar](#)

Planck, M. (1950). in *Scientific autobiography and other papers*, ed. F. Gaynor (London: Williams & Northgate).

[Google Scholar](#)

Radin, D. I. (1997). Unconscious perception of future emotions: An experiment in presentiment. *J. Sci. Explor.* 11, 163–180.

[Google Scholar](#)

Radin, D. I., Stone, J., Levine, E., Eskandarnejad, S., Schlitz, M., Kozak, L., et al. (2008). Compassionate intention as a therapeutic intervention by partners of cancer

patients: Effects of distant intention on the patients' autonomic nervous system. *Explore J. Sci. Heal.* 4, 235–243. doi: 10.1016/j.explore.2008.04.002

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Radin, D. I., Vieten, C., Michel, L., and Delorme, A. (2011). Electrocortical activity prior to unpredictable stimuli in meditators and non-meditators. *Explore* 7, 286–299. doi: 10.1016/j.explore.2011.06.004

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Radin, D., and Borges, A. (2009). Intuition through time: What does the seer see? *Explore* 5, 200–211. doi: 10.1016/j.explore.2009.04.002

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Radin, D., and Lobach, E. (2007). Toward understanding the placebo effect: Investigating a possible retrocausal factor. *J. Altern. Complement. Med.* 13, 733–739. doi: 10.1089/acm.2006.6243

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Radin, D., and Pierce, A. (2015). “Psi and Psychophysiology,” in *Parapsychology: A handbook for the 21st century*, ed. E. Cardena (Jefferson, NC: McFarland & Company, Inc., Publishers).

[Google Scholar](#)

Rao, R. P. N., and Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2, 79–87. doi: 10.1038/4580

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Reardon, S. (2019). ‘Outlandish’ competition seeks the brain’s source of consciousness. Available online at: <https://www.science.org/content/article/outlandish-competition-seeks-brain-s-source-consciousness> (accessed May 16, 2022).

[Google Scholar](#)

Richards, T. L., Kozak, L., Johnson, L. C., and Standish, L. J. (2005). Replicable functional magnetic resonance imaging. Evidence of correlated brain signals between physically and sensory isolated subjects. *J. Altern. Complement. Med.* 11, 955–963. doi: 10.1089/acm.2005.11.955

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Ross, C. A., and Joshi, S. (1992). Paranormal experiences in the general population. *J. Nerv. Ment. Dis.* 180, 357–361.

[Google Scholar](#)

Saxe, J. G. (2016). *The blind men and the elephant*. Hong Kong: Enrich Spot Limited.

[Google Scholar](#)

Schlitz, M., Wiseman, R., Watt, C., and Radin, D. (2006). Of two minds: Sceptic-proponent collaboration within parapsychology. *Br. J. Psychol.* 97(Pt 3), 313–322. doi: 10.1348/000712605X80704

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Schmidt, S. (2012). Can we help just by good intentions? A meta-analysis of experiments on distant intention effects. *J. Altern. Complement. Med.* 18, 529–533. doi: 10.1089/acm.2011.0321

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Schmidt, S. (2015). “Experimental research on distant intention phenomena,” in *Parapsychology: A handbook for the 21st century*, eds E. Cardeña, J. Palmer, and D. Marcusson-Clavertz (Jefferson, NC: McFarland & Company, Inc., Publishers), 244–257. doi: 10.1016/j.explore.2015.10.001

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Schmidt, S., Schneider, R., Utts, J., and Walach, H. (2004). Distant intentionality and the feeling of being stared at: Two meta-analyses. *Br. J. Psychol.* 95(Pt 2), 235–247. doi: 10.1348/000712604773952449

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Schooler, J. (2014). “Bridging the objective/subjective divide: Towards a meta-perspective of science and experience,” in *Open MIND*, (Frankfurt: Frankfurt am Main: MIND Group).

[Google Scholar](#)

Schwartz, S. A. (2005). *The secret vaults of time: Psychic archaeology and the quest for man's beginnings*, Vol. 12. Newburyport, MA: Hampton Roads Publishing.

[Google Scholar](#)

Schwartz, S. A. (2019). The location and reconstruction of a byzantine structure in marea, egypt, including a comparison of electronic remote sensing and remote viewing. *J. Sci. Explor.* 33, 451–480.

[Google Scholar](#)

Selimbeyoglu, A., and Parvizi, J. (2010). Electrical stimulation of the human brain: Perceptual and behavioral phenomena reported in the old and new literature. *Front. Hum. Neurosci.* 4:46. doi: 10.3389/fnhum.2010.00046

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends Cogn. Sci.* 17, 565–573. doi: 10.1016/j.tics.2013.09.007

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Seth, A. K., and Bayne, T. (2022). Theories of consciousness. *Nat. Rev. Neurosci.* 23, 439–452. doi: 10.1038/s41583-022-00587-4

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Shapiro, K. L., Raymond, J. E., and Arnell, K. M. (1997). The attentional blink. *Trends Cogn. Sci.* 1, 291–296. doi: 10.1016/S1364-6613(97)01094-2

[CrossRef Full Text](#) | [Google Scholar](#)

Simons, D. J., and Chabris, C. F. (1999). Gorillas in our midst: Sustained inattention blindness for dynamic events. *Perception* 28, 1059–1074. doi: 10.1068/p281059

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Smith, C. C., Laham, D., and Moddel, J. (2014). Stock market prediction using associative remote viewing by inexperienced remote viewers. *J. Sci. Explor.* 28, 7–16.

[Google Scholar](#)

Standish, L. J., Johnson, L. C., Kozak, L., and Richards, T. (2003). Evidence of correlated functional magnetic resonance imaging signals between distant human brains. *Altern. Ther.* 9, 122–125.

[Google Scholar](#)

Standish, L. J., Kozak, L., Johnson, L. C., and Richards, T. (2004). Electroencephalographic evidence of correlated event-related signals between the brains of spatially and sensory isolated human subjects. *J. Altern. Complement. Med.* 10, 307–314. doi: 10.1089/107555304323062293

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Steriade, M., Timofeev, I., and Grenier, F. (2001). Natural waking and sleep states: A view from inside neocortical neurons. *J. Neurophysiol.* 85, 1969–1985. doi: 10.1152/jn.2001.85.5.1969

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Stevenson, I., and Pasricha, S. (1979). A case of secondary personality with xenoglossy. *Am. J. Psychiatry* 136, 1591–1592. doi: 10.1176/ajp.136.12.1591

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Stevenson, I., and Pasricha, S. (1980). A preliminary report on an unusual case of the reincarnation type with xenoglossy. *J. Am. Soc. Psych. Res.* 74, 331–348.

[Google Scholar](#)

Storm, L., and Tressoldi, P. (2020). Meta-analysis of free-response studies 2009-2018: Assessing the noise-reduction model ten years on. *J. Soc. Psych. Res.* 84, 193–219.

[Google Scholar](#)

Storm, L., Tressoldi, P. E., and Di Risio, L. (2010). Meta-analysis of free-response studies, 1992-2008: Assessing the noise reduction model in parapsychology. *Psychol. Bull.* 136, 471–485. doi: 10.1037/a0019457

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Storm, L., Tressoldi, P. E., and Di Risio, L. (2012). Meta-analysis of ESP studies, 1987-2010: Assessing the success of the forced-choice design in parapsychology. *J. Parapsychol.* 76, 243–273.

[Google Scholar](#)

Templeton World Charity Foundation. (2022). *Accelerating research on consciousness*. Templeton world charity foundation. Available online at: <https://www.templetonworldcharity.org/our-priorities/accelerating-research-consciousness> (accessed May 19, 2022).

[Google Scholar](#)

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

[CrossRef Full Text](#) | [Google Scholar](#)

Tononi, G., Boly, M., Massimini, M., and Koch, C. (2016). Integrated information theory: From consciousness to its physical substrate. *Nat. Rev. Neurosci.* 17, 450–461. doi: 10.1038/nrn.2016.44

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Treffert, D. A. (2009). The savant syndrome: An extraordinary condition. A synopsis: Past, present, future. *Philos. Trans R. Soc. Lond. B. Biol. Sci.* 364, 1351–1357. doi: 10.1098/rstb.2008.0326

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Tressoldi, P., Martinelli, M., Zaccaria, E., and Massaccesi, S. (2009). Implicit intuition: How heart rate can contribute to predict future events. *J. Soc. Psych. Res.* 73, 1–16.

[Google Scholar](#)

Utts, J. (2016). Appreciating Statistics. *J. Am. Stat. Assoc.* 111, 1373–1380. doi: 10.1080/01621459.2016.1250592

[CrossRef Full Text](#) | [Google Scholar](#)

Vandersande, J. W. (2008). *Life after death: Some of the best evidence*. Denver: Outskirts Press, Inc.

[Google Scholar](#)

VanRullen, R., and Kanai, R. (2021). Deep learning and the Global Workspace Theory. *Trends Neurosci.* 44, 692–704. doi: 10.1016/j.tins.2021.04.005

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Wahbeh, H., Radin, D., Mossbridge, J., Vieten, C., and Delorme, A. (2018). Exceptional experiences reported by scientists and engineers. *Explore* 14, 329–341. doi: 10.1016/j.explore.2018.05.002

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Watt, C., Wiseman, R., and Schlitz, M. (2002). Tacit information in remote staring research: The Wiseman-Schlitz interviews. *Paranormal Rev.* 24, 18–25.

[Google Scholar](#)

Keywords: non-local consciousness, orch-OR, integrated information theory, global work-space theories, higher-order theories, re-entry and predictive processing, analytic idealism, interface theory of perception

Citation: Wahbeh H, Radin D, Cannard C and Delorme A (2022) What if consciousness is not an emergent property of the brain? Observational and empirical challenges to materialistic models. *Front. Psychol.* 13:955594. doi: 10.3389/fpsyg.2022.955594

Received: 28 May 2022; **Accepted:** 19 August 2022;
Published: 07 September 2022.

Edited by:

[Antonino Raffone](#), Sapienza University of Rome, Italy

The nonlocal universe

[Andrew Lohrey](#) ^{a,✉}, [Bruce Boreham](#) ^b

- Author information
- Article notes
- Copyright and License information

PMCID: PMC7588183 PMID: [33149801](#)

ABSTRACT

We propose that the universe is nonlocal and that the appropriate worldview or paradigm for this understanding is nonlocal realism. Currently the worldview of local realism guides and frames the understanding and interpretations of science. Local realism was the worldview employed by Einstein in his relativity theories, but the principles of this paradigm have operated as the guiding framework for the rest of classic science for more than a century. This paper points to incoherencies in local realism and to the violation of its principles by recent experiments; it suggests that these negative effects have undermined the credibility and legitimacy of this worldview. We offer a more inclusive worldview for the future of science called nonlocal realism. Unlike local realism, the worldview of nonlocal realism encompasses meaning, mind and universal consciousness.

KEYWORDS: Local realism, nonlocal realism, locality, nonlocality, meaning, consciousness, paradigm, future of science

Introduction

A paradigm or worldview is not a theory, or a series of theories, or a hypothesis. Thomas Kuhn, with whom the now customary use of the term originated, wrote that ‘On the one hand, it stands for the entire constellation of beliefs, values, techniques, and so on shared by the members of a given community. On the other hand, it denotes one sort of element in that constellation, the concrete puzzle-solutions.’ [\[1\]](#) Kuhn also suggested that a paradigm was a ‘disciplinary matrix’ but never satisfactorily resolved the distinction between a constellation and one element of that constellation.

We are applying the notion of a disciplinary matrix to the term ‘paradigm’. Kuhn’s disciplinary matrix represents a network of norms, values and rules common to the practitioners of a particular discipline. A similar view is held by Iain McGilchrist who, in the Foreword to the *Galileo Commission Report*, compares a paradigm with a ‘lens through which we apprehend reality, the problem being that, while such paradigms are indispensable, we tend to be oblivious to the inevitably distorting effect of the lens’.[1](#)

The idea of a ‘lens’ implies a worldview or paradigm that is not innocent but contains a set of principles, habits of mind and predispositions. Like all other disciplines, scientific

practices, whether theoretical, mathematical or experimental, are always subject to a background framework involving hidden assumptions, predispositions and at times a set of principles. A paradigm or worldview operates like a pilot that guides the scientist's actions and dictates what is observed and measured, how results are interpreted, what research is carried out, what are the appropriate controls and equipment for experiments and finally what and how theories are developed.

These framing conditions are addressed directly in this paper. They are not concerned with scientific methods, experiments or their results, but have a focus on the psychological realities of paradigms. How a paradigm determines certain kinds of interpretations is instructive in how scientific results are understood. How science is interpreted is a subject area outside the conventions of physics, biology, mathematics or medicine, yet it is one that deeply influences every aspect of science.

The psychological reality of hidden assumptions and predispositions that create a paradigm will determine how measurements are conducted, or even what is meant by measurement. The practices of measurement are of central significance for all of science, and so these determining factors are critical to every aspect of science. The hidden assumptions and predispositions of the paradigm of local realism (discussed below) mandates that scientific measurements are essentially physical and 'good' measurements should produce factual information that is reliable. What is missing from this paradigm is any consideration of the psychological reality concerning the hidden assumptions and predispositions that have produced this paradigm. These are not inconsiderable factors.

The paradigm that takes account of the psychological reality of hidden assumptions and predispositions we call nonlocal realism. From the viewpoint of this paradigm scientific measurements are products of the mind, and hence all paradigms are essentially psychological. This paper is written from the worldview of nonlocal realism and as a consequence proposes that a paradigm operates as a psychological navigation aid that guides the scientist's actions and dictates how things are measured and interpreted. From the point of view of nonlocal realism, scientific measurements always have at least two levels: the first is the psychology of the paradigm itself; the second will involve a series of instruments and controls, and that will include the symbols of mathematics, and all of these will be arranged and used according to the sightlines and principles of the paradigm.

Local realism

The paradigm called local realism represents the dominant lens or worldview through which mainstream mechanical science (and that includes Einstein's theories of relativity) view the world. This framework of mind has grown out of the culture of scientific activity and philosophical debate over the last three centuries. We argue in this paper that one of the major drawbacks of the paradigm of local realism is it presents a

divided and partial picture of a universe that is actually whole and nonlocal. To a large extent this criticism embraces those criticisms within biology leveled at reductionist methods and related to dissecting biological systems into their constituent parts and then expecting this reductive approach to explain the complexity of living systems that have emergent properties.

This expectation ignores the observer's role in detecting complexity, where the system's collective behavior gives rise to structural modifications and new hierarchical arrangements [2]. Similarly, in physics, emphasizing the constituent parts at the expense of the whole physical system can lead to oversimplified models and experiments that cannot be extrapolated to accurately describe complex physical systems [3,4]. Emerging systems in physics are examples of complex systems where new and complicated properties emerge in response to collective properties and not to properties of the individual parts [5–7].

However, rather than focusing on the limitations of reductionism in biology or physics we are attempting a broader view that has a focus on the psychology of paradigms. This approach enables us to highlight the overall limitations and partiality of the paradigm of local realism when it is employed by anyone in any discipline, as well as how to identify how these constraints can be overcome. In order to achieve that aim we need to begin by identifying the two basic principles of local realism: the 'local' and the 'real'.

The principle of the 'real' in local realism is created by denying the role of the mind while over-valuing the physical world, which is taken to be a 'reality' separate from the subjective minds of observers. In terms of the 'local', physical objects are assumed to be influenced only by their immediate (local) physical surroundings. In physics, for example, a particle is assumed to have properties that are localized in that particle and which cannot influence what happens elsewhere. In biology it is assumed that because living systems are made of atoms this should allow us to explain individual component parts solely in terms of their physicochemical properties. Hence, the paradigm of local realism assumes that individual inanimate physical parts are the primary building blocks of living systems as well as the components of a dead physical universe.

The benefits of the paradigm of local realism are its pragmatic techniques that focus behavior and thinking in ways that provide solutions to practical and technical problems. These are the useful skills and techniques that belong to the technician. The major deficiency of local realism comes from the elevation of these entirely local methods into the general field of philosophy so as to advocate a binary and hence divided worldview, which then becomes applicable to all scientific investigation, physical and social, as well as applying throughout the entire universe. Thus, inherent in these principles of the 'local' and the 'real' is an absolute dualism that divides and separates the physical world from the minds of humans and importantly, the meanings they make [8]. The assumption that mind and body are distinct: [9].

The elevation of local, technical and separating practices into a worldview carries with it the denial of any larger picture involving mind, meaning and the all-embracing context of universal consciousness. In addition, the implications of this worldview reinforce each other in ways that are often incoherent. For example, the absolute dualism of local realism divides the universe into mind and matter. This amounts to the denial that scientists have minds that are embedded within the theories, practices and interpretations they use. Here is the incoherence of the objective approach where it is assumed that the practical scientists have a 'no-mind mind'.

Along with the absolute dualism of local realism comes a series of binary values and preferences, such as quantity in preference to quality; the physical in preference to the mental; objectivity in preference to subjectivity; explicit details in preference to hidden implicit contexts; controlled variables in preference to hidden and uncontrolled variables; the local in preference to the nonlocal. Such binary thinking tends to become habitual so that interconnected and holistic systems will be treated as if they have gaps, divisions and separations. In addition, local realism has a series of important contextual omissions and deletions which are commonly reflected in scientific practices and investigation. These are the deletions or erasures of meaning, mind and consciousness. When these important contexts are deleted from practices and theory the general result again is the separation of parts within wholes along with the fragmentation of knowledge.

When David Bohm wrote in *Wholeness and the Implicate Order* [10] about fragmentation and wholeness he was taking a critical stance against the worldview of mechanical science that focuses on fragments but treats these details as fundamentals. As the term implies, a fragment is part of a whole, but when it is seen as a fundamental feature of the world we change it from being a fragment into a separate and over-valued detail or difference that exists without reference to the overarching contexts of life or consciousness. When these contexts of life and consciousness are ignored or considered non-existent, any part, fragment, difference or feature within a scientific investigation will automatically become over-valued. As Bohm suggests, this kind of thinking is incoherent, while the general denial of context produces the fragmentation of meaning and, as a consequence, of knowledge.

Bohm went on to write that fragmentation is a confusion around the question of difference and sameness and '*To be confused about what is different and what is not, is to be confused about everything*' [10, p. 16]. This is a major category of confusion which we suggest is the hallmark of the paradigm of local realism. This raises the question of the credibility, legitimacy and continuing usefulness of the worldview of local realism as the central psychological guide for mainstream science. These are serious questions about the future effectiveness of this paradigm. Yet such questions do not seem to have had much influence on many scientists. For example, Einstein's commitment to this scientific paradigm was manifest in his insistence on the entirely local nature of particle connections and his rejection of their nonlocal exchanges, or what he called 'spooky action at a distance' [11]. Einstein rejected the idea of 'instantaneous communication'

between particles because in local reality his theory of relativity forbids any signal to travel faster than the speed of light [\[12\]](#).

Nonlocal realism

We consider that the future of science rests on changing its psychological reference frame from local realism to that of nonlocal realism. Nonlocal realism is the term we use to represent a holistic and integrated worldview that displays the reality (hence realism) of a singular integrated, interconnected universal field of relationships. The reality of this paradigm is established by the evidence that the universe we perceive and measure is nonlocal and by an analysis of the integrated nature of the meanings we make.

Nonlocal realism is closely related to the view espoused by Erwin Schrödinger that the overall number of minds within the universe is just one.² He went further by suggesting that mind has erected the physical outside world out of its own mental stuff [\[13\]](#). The interconnected universal consciousness implied by the concept of one mind constitutes the nonlocal, singular implicit reality of a universal consciousness that has embedded within itself the local and explicit conscious mind of each individual.

The paradigm of nonlocal realism denies the absolute dualism inherent within local realism along with the specific separation of mind and the more general psychological tendency to divide and separate all aspects of the world. As a consequence of this emphasis, the primary status that has traditionally been given to an independent physical world disappears. Yet the disappearance of this primary status does not make the physical world disappear. The withdrawal of this status comes from its reconstitution as the secondary effects produced by local minds through the processes of perception and conception. As Lohrey argues, this reallocation of status comes about because the context of universal consciousness is taken into account as the fundamental context; a context in which local minds with perceptions and conceptions of a physical world have a secondary status [\[14\]](#).

The reallocation of the physical world's status comes about because of the asymmetrical structure of universal consciousness, which stands as the permanent background context on which all of science is writ. Universal consciousness has a structural order in which hidden implicit meaning always has primacy over the secondary status of the explicit meanings of distinctions and differences, and these always involve objects and forms. This order tells us that fragments, parts, objects and forms always arise from wholes (not the reverse), and also that explicit meaning always arises from implicit meaning, and that humans always arise from Nature. This asymmetrical order informs us that consciousness has a local mind and a physical body, rather than the reverse that says the body has a local mind that is consciousness.

However, while individual human minds have a secondary status within the paradigm of nonlocal realism, they are always the ground on which all interpretations and

observations are developed. Unlike local realism which ignores or sidelines the human mind or treats it as an entity, nonlocal realism has at its core the human mind as the context for all meaning-making and psychological activity. This means that the observer is never simply a local subjectivity but represents the local entrance into a broader implicit context of nonlocal, implicit universal consciousness.

The status of the observer as context becomes more obvious when we begin to take seriously the evidence that *the only way that we are able to know the physical world (and that includes the brains that neuroscientists study) is by observation and conceptualization, and both of these psychological processes operate within the mind of the observer.* This crucial evidence tends to be dismissed or devalued by local realists.

Hence, any scientific interpretation that is framed by the paradigm of nonlocal realism will take on a different character from those seen through the lens of local realism. First, that difference will be in replacing gaps, divisions and separations with an integrated sense of inclusion and interiority. For example, all relations, both local and nonlocal, are internal to the singularity of universal consciousness and operate entirely within this domain, with no relations existing or functioning outside it. This universal domain of consciousness is not the same as the domain of subjectivity, a term applicable to the dualism of local realism. Thus the subject does not stand as an entity independent of the physical world but, rather, functions as the primary always open local doorway into the implicit knowing of universal consciousness.

We can say then that the paradigm of nonlocal realism automatically has at its center the mind of the observer. To put it somewhat differently, in terms of nonlocal realism the observer becomes the center of the universe and that holds for every observer. How is it possible for the center of the universe to be in different locations? It is possible because each mind has a common foundation of implicit meaning embedded within universal consciousness. This implicit field functions prior to any sense of difference (non-symmetry), and therefore the observer, while always a secondary feature to universal consciousness, will at the same time be the central locale for all potential and actual social exchanges. *One of the important social consequences that flow from this is that all observers are equal in being at the center of the universe, with none left out.*

What evidence is there that the universe is nonlocal and how does this evidence relate to the creation of a paradigm of nonlocal realism?

To answer that question, we turn first to the mind of the observer, which is where we find the primary evidence for the paradigm of nonlocal realism. The mind of the observer has a similar structure to Bohm's two orders of the universe: the implicate and explicate orders. We relate these two orders to the structure and function of meaning. Why meaning? Bohm considered that meaning was the essential nature of

consciousness [15]. We fully agree, but go further by proposing that meaning is the content of language, mind and consciousness. In this paper we have interpreted Bohm's two great orders not as physical orders but as orders of meaning and hence as the metaphysical orders of universal consciousness.

As a consequence, Bohm's implicate order represents a universe-wide, permanently hidden context that is actually full of the implicit relations of meaning, that is, a plenum of implicitness or implicit knowing [14]. Likewise, the explicate order represents that multiplicity of secondary and derivative explicit relations that are located within the conscious and explicit minds of individuals. For mainstream science this explicit order is elevated to 'the physical world' and then its differential forms are over-valued and made to appear separate and divided through the deletion of the natural integrating forces of the implicate order. The distinctions between the implicate order (implicit meaning) and the explicate order (explicit meaning) are concerned not so much with the differences between the visible and the invisible, but more generally with the differences associated with two kinds of meaning, knowing and intelligibility.

The two kinds of knowing are discovered within the relationship of Bohm's two orders and, more directly, with the identical relationship between implicit and explicit meaning. As with Bohm's two orders, meaning has two poles of attraction that are entirely integrated. These are, i) the pole of implicit meaning, and ii) the pole of explicit meaning. The knowing of explicit meaning comes in the creation of distinctions, differences, contrasts and marks. In contrast, implicit knowing produces links, connections, unities and wholes and represents the hidden content of every context in every area of endeavor. Implicit knowing has the structure and function of symmetry. Explicit knowing is structured by the relations of non-symmetry and asymmetry, which gives rise to distinctions and differences within systems [14]. An example of the symmetry of implicit knowing is the intelligence shown by twin particles who know instantly the orientation of the other.

The poles of implicit and explicit meaning have the further attributes of nonlocality and locality. While both these terms have come from scientific developments, they are not exclusive to science as they represent general terms applicable to all human endeavors, while adding significantly to the scope of our understanding of the surface differences and the depth dimensions of universal consciousness and its relationship to local minds. The pole of explicit knowing always has locality in that it creates a focus on locally situated space and time objects, forms, movements and actions (Bohm's explicate order). This means that explicit meaning is evident in every conscious action across a wide range of behavior, from mathematics to rock climbing, from child-minding to scientific experimentation. On the other hand, the pole of implicit meaning has a nonlocal, infinite and boundless character of knowing and hence functions as a dynamic movement of hidden and entangled relations across and beyond the distinctions of time, space, and the material body and the physical world.

Like Bohm's two universal orders, these two kinds of knowing are not binary and do not form the kind of absolute dualism common to local realism; rather they are entirely integrated yet allow for two kinds of emphasis: a local and a nonlocal focus. For example, a nonlocal emphasis produces the view of the universe as an infinite singularity, a single system of nonlocal consciousness that has two major features: a background of permanently hidden, implicit relations (that cannot be measured) and a foreground of revealed explicit differences and forms (that produce measurements). This is the holistic gestalt-like structure of universal consciousness. Within this singular system of universal consciousness every relationship is related to every other relationship. This universal singularity of meaning is reflected in physics by the state that has been called superdeterminism [\[16\]](#).

These two kinds of knowing present us with the choice of where we place the emphasis, whether locally or nonlocally. For local realists that emphasis is always on the local, while for nonlocal realists the emphasis is on the nonlocal. However, quite apart from our choices the structure of this infinitely interconnected singular system of universal consciousness is gestalt-like and remains unaffected by whatever emphasis we employ. What is 'a gestalt-like structure'? The relationship between Bohm's implicate and explicate orders and the relationship between implicit and explicit meaning are identical in that in both cases these two features cannot be separated from each other and cannot be treated in isolation. This indicates that both orders and meanings are entirely integrated into a single gestalt-like holism in which every local form will always have hidden implicit features or what in quantum physics has sometimes been called hidden variables.

The word 'hidden' can be used in two very different ways. One use comes with the assumption that this is a temporary situation and that what is hidden can be, or may be, revealed later. The other use implies that hidden is a permanent condition and this means there is no possibility under any circumstance of measuring this state. It is this second use of hidden that refers to the character of nonlocal connections, for these are relationships that will remain forever hidden, or, more accurately, remain forever non-explicit. Roughly speaking this is the situation scientists find in relation to the question about a particle's position or momentum, which cannot be simultaneously measured.

Hence, every explicit particle and every life form will have some permanently hidden implicitness (variables) that represent the background context of universal consciousness, a context in which every particle and every life form is immersed. From a nonlocal worldview these two kinds of knowing mean that fragments (in whatever form they take) are always treated as parts of a larger whole and that every local particle, form or object always operates within a nonlocal system of relationships. A fragment may be a virus, a particle, a mathematical equation, a word, a narrative, money, a house, a city, or the measurements of time or the three dimensions of space. In contrast, the local emphasis of local realism is to treat such fragments, not as fragments but as distinctions without having a qualifying context.

What is the scientific evidence that supports the paradigm of nonlocal realism?

In the last almost seventy years the scientific evidence for nonlocal connections between entangled particles has directly challenged the credibility of Einstein's common-sense local realist worldview that prohibits instantaneous nonlocal connections. In 1951 Bohm proposed an experiment using correlated pairs of particles that could test instantaneous communication between them [\[17,18\]](#). He proposed the creation of correlated states between properties of quantum particles that have a discrete set of possible values (e.g. vertically and horizontally polarized light or light spinning up and down). However, this measurement is not enough to confirm action at a distance. It is still possible that correlation observed between the particles could be due to hidden variables (that is, some property of the particle system that predetermines the outcome of the measurement that is currently hidden from view but may be revealed at a later time).

In the early 1960s John Stewart Bell proposed his now famous theorem that local hidden variables cannot account for all the predictions of quantum mechanics (QM) and that local determinism is incapable of describing what is happening in quantum experiments [\[19–21\]](#). He developed what has come to be known as 'Bell's Inequalities'. These inequalities can be used to determine the difference between how strongly entangled particles would be correlated in the different cases of quantum mechanics alone and entanglement involving hidden variables. Such experiments have been carried out many times over several decades, confirming the QM predictions and consistently demonstrating experimentally that the principles of local realism are violated by nonlocal connections that link entangled particles across vast distances [\[22–28\]](#). Recent research conducted in 2016 by a worldwide research team led by the Institute of Photonic Sciences (ICFO) in Barcelona has again confirmed these findings.³ These results imply that quantum mechanics violates either locality or realism.

In addition to these experiments, the Free Will Theorem of Conway and Kochen [\[29–31\]](#) claims to show that 'in principle' correlations between particles cannot be due to hidden variables. Based on a number of axioms, they provide a proof that if experimenters have free will then so must elementary particles. Therefore, the hidden variable hypothesis should be abandoned. This claim has attracted both support and criticism, with some critics suggesting that the theorem applies but in a more restricted sense than its authors propose [\[32,33\]](#). Further, although the starting axioms have been modified by Conway and Kochen in response to criticism, they still draw on concepts from local reality in limiting the speed of propagation of communication and cannot be assumed therefore to support the concepts of nonlocal realism.

A nonlocal universe would be an interconnected, holistic universe and if this is the basic structure of a nonlocal reality then its implications for science and society are profound

and revolutionary. For example, it will mean that contexts are to be considered when doing science, and principally that means the contexts of meaning, language, mind and consciousness. A science in which consciousness was not alien to QM experiments is one in which scientists will not be surprised that one of a pair of photons originating from the same source knows instantly the state of its twin even though they are many kilometers apart. If photons 'know' something it indicates they have consciousness and that means the universe as a whole has consciousness and is alive. This is not to say that photons have free will for such an idea fails to take account of the gestalt-like structure of universal consciousness. Free will is an issue about differential and explicit decisions made within the local human mind and it is not an issue related to universal consciousness, which is an infinite field that is entirely implicit [14]. The significance of individual scientists and the particles they experiment with not functioning independently is considered further below in the discussion of Wheeler's delayed choice experiment.

The possibility of violation of realism has led some scientists to revisit the notion of the world as a simulation or local illusion. For instance, Quantum Bayesianism suggests that there is no objective reality but only a subjective estimation of it [34]. Muller takes this further by suggesting that physical reality is fundamentally observer-relative.⁴ The possibility that the world is a simulation has been pursued by Bostrom, suggesting that we are more likely than not to be living in a computer simulation [35]. More recently Erwin et al. propose that the universe is a self-simulation that might exist as a broad class of possible theoretical models of reality obeying the *principle of efficient language* axiom [36]. However, a broad consensus has emerged that it is locality that is violated, not realism, and that QM involves some type of nonlocality, although what is the exact nature of the nonlocality remains controversial [37].

Even though experiments in QM have consistently violated the principles of local realism and upheld the principle of nonlocality and, as we have already argued, the significant omissions of local realism produce only a partial view of the universe, in the face of all this overwhelming evidence there is still a tenacious adherence to these principles by most scientists. We can only assume that the very idea that science is host to the guiding principles of a partial paradigm is so alien to many that they treat any criticism of local realism as a mere detail. As a consequence, many seem unable to tell the difference between a fatal blow to their worldview and what maybe thought of as an incidental embarrassing mistake by Einstein and others.

In contrast to this mainstream view there has been over the last century a wide range of scientists who have given voice to, and provided evidence of, a more inclusive, integrated and holistic universe. An appreciation of the point can be gained from David Bohm's model of the two great orders of the universe, which provide an integrated structure of a holistic universe.

More recently, in her paper 'Non-locality as a Fundamental Principle of Reality', Elizabeth Rauscher sets out to confirm the proposition presented in the title of her paper

and for the most part succeeds [38]. Yet we are left to wonder what is the ‘reality’ to which she refers. Rauscher does not address this question directly but seems to imply that the reality she refers to is an interconnected, holistic universe, for she writes, ‘It is clear that this principle of nonlocality has profound implications about the nature of a nonlocal universe’. *She goes on to suggest that locality maybe considered as a ‘special and limiting case’ of the more basic nonlocality.*

Some early pioneers of quantum mechanics [39–41], including the mathematician John von Neumann, advanced theories and postulates concerning the possible role of consciousness in the collapse of the wave function and the measurement problem while arguing that the human mind has a direct influence on the collapse (the von Neumann–Wigner interpretation). Today these ideas remain contentious and the subject of debate [42,43]. Wigner originally argued that experimenter consciousness was essential for wave function collapse but moved away from this interpretation in later life [44].⁵ While the term ‘know’ implies consciousness, the well-established nonlocal connection between twin particles suggests a holistic universe that is interconnected through consciousness. Yet a universe that has a background of instant interconnecting consciousness is not the type of nonlocal universe that Rauscher argues for, as she believes that ‘superluminal signals must be invoked’ even though she refers to the universe as ‘nonlocal’.

Rauscher’s proposed nonlocal universe appears to be interconnected by superluminal (faster than light) signals rather than implicit knowing. However, it is commonly agreed that in quantum phenomena, superluminal signaling is impossible in practice. Moreover, many scientists believe that the so-called ‘no-communication theorem’ or ‘no-signalling principle’ excludes superluminal signaling in principle as well as in practice. There are a number of proofs of this theorem to be found in the literature [37, p. 25; 45],⁶ for example, proposing superluminal signaling invokes locality (signals are local and potentially measurable] while on the other hand contradicting locality with the concept of superluminal exchanges. In addition, if nonlocality is a fundamental principle of the universe what is it about faster-than-light signals that is nonlocal? Surely, a faster-than-light movement can be measured to make it so, which in turn makes it entirely local. These are interesting issues because they again raise the question of the psychological framework we are using to understand the nature of nonlocality.

The idea of quantum nonlocality was introduced by Niels Bohr in 1935. To allow *the possibility* that when two particles interacted it might be possible, by measuring one particle, to work out some of the properties of the other without needing to measure it directly (that is, through their entangled history), he suggested that the state of both particles simply became ‘real’ at the same time, that is to say, measuring one particle would instantaneously influence the other, regardless of the distance separating them [46]. This distance can be extremely large, even thousands of light years.

This proposition implied more than just a set of correlations between distant events for it includes the notion that it becomes impossible to measure the features of one particle (position, momentum, polarity) without instantly steering the other into a corresponding state even across vast distances. This means that correlated particles are no longer separate particles, one influencing the other, but two parts of a single system separated but interdependent in space [47]. (or in other words entangled particles are an example of an emergent system).

We see from Bohr's description that entanglement of a state shared by two particles is a necessary but not sufficient condition of quantum nonlocality. There exist some states that are genuinely entangled but admit a local model [48]. Entanglement is more commonly viewed as an algebraic concept, noted for being a *prerequisite* to nonlocality, whereas nonlocality is defined according to experimental statistics. As we have seen, nonlocality allows two distant observers to obtain experimental statistics that cannot be described by any classical common cause, as if they *instantly* swapped details on this experimental intrusion across a distance⁷ rather than propagating continuously in space-time. This means that quantum nonlocality is defined in terms of the processes it allows, not by its fundamental nature.

Berkovitz has examined the proposition of nonlocality in detail as well as an alternative common view that these influences are due to some type of holism and/or non-separability of states of composite systems, which are characteristic of systems in entangled states (like the spin singlet state), and which exclude the very possibility of action at a distance [45,49, p. 8]. Holism is essentially the idea that at least some properties of wholes are not determined by the physical properties of their parts. There are various forms of holism in the literature. For example, Berkovitz defines holism as a violation of the condition of Particularism, that is, **the proposition that** the world is composed of individuals. All individuals have non-relational properties and all relations supervene upon the non-relational properties of the relata and the spatiotemporal relations between them.

Berkovitz considered in detail the nature of holism and non-separability as manifested by various interpretations of quantum mechanics and whether these interpretations predicate the existence of action at a distance. He concludes that quantum action at a distance would be explained by the holistic nature of the quantum realm and/or non-separability of the states of the systems involved where action at a distance is defined as 'a phenomenon in which a change in intrinsic properties induces a change in the intrinsic properties of a distinct system without their being a process that carries this influence contiguously in space and time'. Conditions are identified under which all three of holism, non-separability and action at a distance apply and under which action at a distance only is involved. According to Berkovitz all the collapse models postulate the same kinds of holism and non-separability as orthodox quantum mechanics.

In summary, the interpretations of QM considered by Berkovitz involve characterizations of holism, non-separability and/or some type of action at a distance. While holism, non-separability and action at a distance could explain the process of *nonlocality obtained empirically*, they do not address its basic nature or cause, that is, the nature of its connections.

While the physics of nonlocality have been repeatedly verified over the last century, its oblique definition has remained a problem to most scientists. Perhaps this is because one of the defining characteristics of nonlocality appears to be its unmeasurable nature, that is, nonlocality is impossible to measure in any way (see for example Berkovitz and Hemmo⁸). If something is unmeasurable then mainstream science, which rests on the worldview of local realism, is unable to deal with this situation because the measurement of quantities is the *raison d'être* for this brand of science. Hence, if two particles know instantly the orientation of the other, even though they are many kilometers apart, the instantaneous knowing represents the area that is unmeasurable.

An instantaneous knowing (even by particles) is unmeasurable because in this domain of knowing there are no distinctions or differences that can be measured. In other words, we suggest that any measurement of any kind can only be made when there is a field, space or system containing distinctions and differences that may be compared and correlated. Such comparisons and correlations constitute the processes of measurement. Thus, as a general rule we can say that any measurement, scientific or otherwise, cannot occur in the absence of differences or distinction. Instantaneous connections are a case in point. Another more human case is insight, intuition and extrasensory perception, processes that do not involve differences that can be measured, perhaps the reason why these human abilities are comprehensively dismissed by local realist scientists.

If the first characteristic of a nonlocal connection is the impossibility of measurement, then a second characteristic represents the hidden nature of these connections. We have previously noted that the character of nonlocal connections is such that they will remain forever hidden. There are further distinctions to be made regarding the hidden character of nonlocal connections. These involve the conditions of the visible and the invisible. This distinction relates directly to the two major functions of the human mind: perception and conception. Concept formations, that is, thinking, represents the processes of interpretation, yet these processes are invisible to us for they do not register with the naked eye and hence are hidden from direct sight. Images of visual perceptions, however, are by their very nature visible. As a consequence, one of the critical differences between perceptions and conceptions is the distinction between the visible world and our invisible interpretations of that world. This difference is significant as it allows for the wide variety of interpretations of the same images of the world.

In relation to the processes of knowing these are entirely hidden in the sense that they are not visible. Yet while invisible, the processes of knowing (and we can include the

term ‘intelligibility’) are the very basis by which we humans know and make meaning of the world. This is also the case for relationships. Relations are known to us because we use them all the time to make decisions based upon comparisons, assessments, measurements and interpretations. These kinds of relations are explicit. Yet even explicit relations tend to be invisible because they do not register visually for us, though they are known to us. We can realize that two particles can know instantly the orientation of the other and yet such knowing is permanently invisible to us and hence these connections will be unmeasurable. These kinds of relationships have the character of nonlocal connections and again will be permanently invisible.

Rauscher refers to ‘a nonlocal universe’ as one in which nonlocality is ‘a’ fundamental principle. Yet for the paradigm of nonlocal realism it is not so much ‘a’ but ‘the’ fundamental principle. As a consequence, we would expect that a nonlocal universe will be the same everywhere and exhibit the character of nonlocality in every place and every aspect of the universe. What further evidence is there that can substantiate this claim?

We find evidence of the observer as the central context in John Wheeler’s delayed choice experiment.⁹ In Young’s classic two slit experiment when electrons are aimed at a barrier containing two slits, the electrons behave like waves, that is, after traversing the two slits, they at once produce an interference pattern on a detector on the far side of the barrier. If, however, the slits are closed off one at a time the interference pattern disappears, and the electrons pass through the barrier like individual particles. In Wheeler’s delayed choice experiment, the experimenter decides to close one slit or leave both open not before but *after* the electron passes through the barrier, with the same results as when the decision is made before the electron passes through the barrier. That is, the electrons seem to know in advance how the experimenter will choose to observe it. Wheeler’s delayed choice experiment has been experimentally realized by Jacques et al.¹⁰

In terms of local realism, the results of the delayed choice experiment make no sense at all. However, from the perspective of nonlocal realism these results synchronize with the principles and sightlines of this paradigm. This is because universal consciousness is omnipresent, ‘beginning at the level of the most basic quantum particles ... and proceeding to produce a deeper understanding of the entire cosmos’ [50, quoted in 14, p 42]. This means that local individual scientists, along with the electrons they are experimenting with, do not function in isolation, rather they both exist within the nonlocal, implicit context of a knowing universal consciousness. If particles know in advance how the experiment chooses to observe them, they know because universal consciousness is immanent within the scientist’s mind as well as in the particles of the experiment.

One of the oldest interpretations of QM is the Copenhagen interpretation devised in the 1920s by Niels Bohr and Werner Heisenberg. Some of the effects of nonlocal realism

can be appreciated when we apply it to the Copenhagen interpretation. According to this interpretation, physical systems do not have definite differential and explicit properties until measured by local minds. This is an accurate description of the functions of the human mind. However, local minds do not exist on their own as assumed by local realism, but only exist within the context of nonlocal consciousness, which has a foundation role to play in all local measurements.

In the vocabulary of QM, the act of measurement affects the system, causing a set of probabilities to reduce to one *after* measurement. Measurement is defined in local realist terms as the testing or manipulation of a physical system in order to yield a numerical result. Bohr believed that the results of measurement could only be described in the language of classical physics and that it made no sense to ask what was going on in the invisible quantum realm. From the worldview of nonlocal realism, he is entirely correct; it makes no sense to try and describe distinctions within the invisible unmeasurable realm of universal consciousness because there are no conscious distinctions within this domain. In this domain there are only hidden relations, and the human uncertainty that always attends nonlocal implicit exchanges.

Another window into the dilemma of local reality is obtained from the more minimalist interpretation of quantum determinism given in the statistical interpretation or ensemble interpretation of QM.^{[11](#)} This is based on Max Born, The Statistical Interpretation of Quantum Mechanics, Nobel Lecture, December 11, 1927. In this interpretation the probability distribution function is taken at face value. This function is used only to predict probabilities. The behavior of an ensemble of particles can be predicted but not the behavior of individual systems or particles.

In order to know the outcome for an individual particle the experimenter must observe it, that is, the role of the observer is passive. There is no physical object or process that collapses because the wave function has no physical meaning. The problem that there is no prescription in QM for the collapse of the wave function is bypassed. While this has resolved one dilemma it introduces another. In this view quantum physics is unable to provide a deterministic outcome and we arrive back at Bohr's view where we have nothing more to say about quantum reality, or we say that the Copenhagen interpretation is simply a metaphor for the mathematics. In either case it is not possible to know in advance the outcome of a measurement.

We have effectively arrived at a limit of understanding where according to Bohm 'the statistical features of quantum theory are thus regarded as representing a kind of irreducible lawlessness of *individual* phenomena in the quantum domain. All individual laws (classical physics) are then regarded as limiting cases of the probability laws of quantum theory, approximately valid for systems involving large numbers of molecules.' [[10](#), p. 69] From the perspective of nonlocal realism, Bohr was correct insofar as the results of measurement could only be described in terms of classical physics.

However, the vocabulary of nonlocal realism answers this difficulty in that the unmeasurable relations of meaning operate prior to the visible macro world. The gestalt-like structure of this metaphysical domain of consciousness allows for both a description of the 'invisible quantum realm' as well as an integrated description of the measurement results. In regard to the invisible quantum realm this is the hidden universal realm of implicit knowing called consciousness. In relation to the language of classical physics this is directed at the multiple and differential realm of Bohm's explicate order involving explicit distinctions, in which the physical behaviors of the experimenting scientist take place.

There are not two independent orders involved in scientific investigations or in life. There is only one. The so-called 'potential realities' of the invisible quantum realm represent the implicit meanings and knowing of the implicate order, which is the fundamental nonlocality of the universe. The local realist view concerning the measurements of the physical world do not have the status of an independent order, for such measurements and their discourses represent only a set of derivative distinctions and differences that arise from the prime order of universal consciousness. Measurements are explicit maps that have their semantic foundations within this prime order of knowing.

In addition, there is no physical object or process that collapses with the exchanges of these two orders of meaning, so the phrase 'the wave function collapse' appears to be a local realist construction and overly physical and incongruous. What continually occurs between these two orders of meaning are the transformations of implicit meaning into explicit meaning: transformations that contain the structure of *implicit-to-explicit* exchanges and represent the key feature that gives rise to the local conscious mind of the individual. This significant change from nonlocal to local, from implicate to explicate, from implicit to explicit meaning has under the guidance of local realism undergone a rough categorization as 'the wave function collapse'. These transformations do not occur externally, *out there* in the physical world, but operate entirely within the singularity of universal consciousness while involving the local level of the human scientist's mind.

For the scientists who employ nonlocal realism to organize their work it will mean that all experiments will take place within a nonlocal as well as a local world. In other words, every experiment in whatever discipline will have classic, macro, local time and space objects, movements and measurements operating concurrently with nonlocal, unmeasurable interconnections of universal consciousness of which the scientist's mind will be a part. This integrated duality tells us that within every scientific experiment there are not two contradictory conceptual schemes at work (micro and macro) as many local realists have assumed. Rather, nonlocal realism takes account of the encompassing singularity of universal consciousness and that unifying reality involves relations that are implicit and nonlocal as well as those that are explicit, practical and local. In other words, nonlocal realism embraces invisible, unmeasurable relations as well as those visible objects and forms associated with explicit and differential relations.

For the bench scientist this means overcoming the constraints of local realism by placing greater emphasis on the nonlocal, that is, addressing the entire context of an experiment via a holistic and inclusive approach that recognizes connectedness, complexity, the role of the collective properties of emerging systems, and the role of consciousness. It also requires that experimental controls be developed and exercised that are appropriate to the broader more holistic approach of nonlocal realism [51,52].¹² Nonlocal realism does not provide the scientist with a set of levers that can be pulled to control an experiment better and nor does it give a sharper definition to binary logic. Rather, by employing the psychological navigational aid of nonlocal realism we become more aware of the unities and affinities that connect everything to everything else. For example, [53], argues that space and time are not physical, rather the differentials of space and time are always simply aspects of the local human mind, while the symmetry that underlies both is a central and unmeasurable feature of universal consciousness. Both these local and nonlocal features form the integrated systems we call space and time or the spacetime continuum.

Applied to biology, the nonlocal realist would not refer to 'life' when looking through the electron microscope at the structure of a bacterium. The integrative reference here should be a local 'life form'. The general designation 'life' properly applied refers to that animate field of universal consciousness because life is one of its characteristics. Plant neurobiology is a recently developed field of plant biology research 'aimed at understanding how plants perceive their circumstances and respond to environmental inputs in an integrated fashion' [54]. This field of research brings up the controversial question of 'plant intelligence' and the less controversial newly developed field of bioinformatics which seeks to explain plant communication, now well established as plant signaling.

We would speculate that by applying a nonlocal approach to plant neurobiology research and specifically to the problems of plant communication this change would be more productive than using the orthodox paradigm of local realism. One example of that change would be the deletion the term 'information' from research and replacing it with 'meaning' [55]. 'Information' is a mechanical, local realist term, the separating theory of which tends to prevent deeper understandings of the implications inherent within any and all forms of communication. Hence, when communication is understood to indicate an exchange of meaning the latent unities and integrated connections within communication exchanges between organisms of any kind will tend to become more obvious. These integrative aspects in plant communication should also assist with the controversial question of plant intelligence because the sightlines within such nonlocal observations inherently take account of consciousness and hence would not be inhibited by local realist predispositions for divisions and separations. In other words, the 'controversy' about plant intelligence is generated by the worldview of local realism and so it does not come from any scientific evidence.

Thus, scientific investigations and interpretations framed by nonlocal realism will have a very different character to classic local realist ones. Under the influence of nonlocal

realism, the universe is no longer a dead, insentient world but rather an animate participatory universe where every single physical object, from particles to galaxies, from microbes to monkeys will contain more hidden connecting meaning than what is revealed by the appearance of their physical shapes, exchanges and movements.

Finally, an example of nonlocal realism comes from the inventor and philosopher Arthur Young (1905–1995), who stated that photons have no mass, charge or time, which suggests their behavior transcends physics in the manner that consciousness does [56]. Young also wrote that ‘Light is not like other things ... ’ and ‘Light is not seen; it is a seeing [56, p. 11].’ When light becomes a ‘seeing’ there is no more dualism of subjective seer and objective seen, rather there is simply a process within the singularity of an interconnected universe that jointly involves a local mind along with nonlocal consciousness. What is given in this participatory process is the sight within seeing, that is, the seeing essence of visual perception. Yet the ownership of that sight does not come from a separate local mind or the intention of an individual who has almost no control over seeing or not seeing. Rather, the ownership and agency of the sight within our seeing entirely rests with nonlocal universal consciousness.

Conclusion

Our review of the paradigms of local realism and nonlocal realism suggests that the usefulness of local realism as a worldview is severely limited and that science in the future will most likely turn to a more inclusive, integrated paradigm that has the character of nonlocal realism.

Notes

[1]

<https://explore.scimednet.org/>

[2]

<https://www.organism.earth/library/quotes/178>.

[3]

[https://thebigbelltest.org/science/\(The Big Bell test of 2016\); Handsteiner J](https://thebigbelltest.org/science/(The%20Big%20Bell%20test%20of%202016);Handsteiner%20J) [57]).

[4]

Muller 2019: arXiv:1712.01826

[5]

John Wheeler also believed that consciousness plays a role in bringing the universe into existence – the Participatory Anthropic Principle (PAP) ('The anthropic universe' Abc.net.au. 2006-02-18. Retrieved 2011-01-24). In contrast, Bohr rejected the idea that the experimental outcome is due to the observer (Faye J [58] 'It is certainly not possible for the observer to influence the events which appear under the conditions he has arranged' (Bohr N [59])). David Bohm himself developed a theory of QM (the pilot wave theory or the de-Broglie-Bohm theory). This theory is nonlocal (thereby satisfying Bell's inequality), taking place in a single space-time and is deterministic. The measurement problem is resolved because particles have definite positions at all times [60].
[6]

https://en.wikipedia.org/wiki/No-communication_theorem
[7]

https://en.wikipedia.org/wiki/Quantum_entanglement
[8]

See Sections 8.1 and 9 in [61].
[9]

This is discussed in John Horgan's (2018) article on Wheeler: 'Do our questions create the world? (<https://blogs.scientificamerican.com/cross-check/do-our-questions-create-the-world/>); Wheeler JA [62].

[10]The experimental realisation of Wheeler's delayed-choice Gedanken Experiment. In V [51], *Science*, 315(5814), pp 966-8. Also, in D [52]. Consciousness and the double-slit interference pattern: Six experiments. *Physics Essays* 25(2). They describe using a double-slit optical system to test the possible role of consciousness in the experiment. In this experiment the ratio of the interference pattern's double-slit spectral power to its single-slit spectral was predicted to decrease when attention was focused toward the double slit as compared to away from it. Data contributed by 137 people in 6 experiments indicate that the spectral ratio decreased as predicted and it was concluded these results were consistent with a consciousness-related interpretation of the quantum measurement problem.

[11]This is based on Max Born, *The Statistical Interpretation of Quantum Mechanics*, Nobel Lecture, December 11, 1954. Bohr accepted Born's statistical interpretation because he believed that the wave function has only a symbolic meaning and does not represent anything real. Therefore, it makes no sense to

talk about the collapse of the wave function (Faye J [58], p. 7). Bohr stressed the importance of experiments to understand the quantum world. Each experiment in effect asks a question of the quantum world. The questions are highly colored by everyday experience and the experiments are rooted in classical physics. (Gribbin J [63]).

[12]<https://www.embopress.org/doi/10.15252/embr.201949110>

Disclosure of Potential Conflicts of Interests

The authors report no conflict of interest.

References

- [1].Kuhn T. The structure of scientific revolutions. Chicago: University of Chicago Press, p. 175; 1970. [[Google Scholar](#)]
- [2].Licata I, Sakaji A. Physics of emergence and organization. US and UK: World Scientific; 2008. [[Google Scholar](#)]
- [3].Boreham BW, Hora H. Energy spectra of electrons emitted from laser irradiated low-density gas and the correspondence principle of electromagnetic interaction. *Laser Particle Beams*. 1995;13(1):71–81. [[Google Scholar](#)]
- [4].Frauchiger D, Renner R. Quantum theory cannot consistently describe the use of itself. *Nat Commun*. 2018;9(3711). DOI: 10.1038/s41467-018-05739-8 [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- [5].Anderson P. More is different. *Science*. 1972;177(4047):393–396. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- [6].Carroll R. On the emergent theme of physics. Singapore: World Scientific; 2012. [[Google Scholar](#)]
- [7].Wen X-G. Topological order: from long-range entangled quantum matter to a unified origin of light and electrons. *Int Sch Res Notices*. 2013. DOI: 10.1155/2013/198710 [[DOI](#)] [[Google Scholar](#)]
- [8].Crane T, Patterson S. History of mind-body problem; 2001. pp 1–2.
- [9].Hart WD. Dualism. In: Guttenplan S, editor. *A companion to the philosophy of mind*. Oxford: Blackwell; 1996. p. 265–267. [[Google Scholar](#)]
- [10].Bohm D. Wholeness and the implicate order. London: Arc; 1983. [[Google Scholar](#)]

- [11].Einstein A, Born M. Letter from Einstein to Max Born, 3 March 1947. In: Walker BM, editor. The Born-Einstein Letters: correspondence between Albert Einstein and Max and Hedwig Born from 1916 to 1955. New York, NY: Macmillan; 1971. p. 157–158. [[Google Scholar](#)]
- [12].Einstein A, Podolsky B, Rosen N. Can quantum-mechanical description of physical reality be considered complete? Phys. Rev. 1935;47(10):777. [[Google Scholar](#)]
- [13].Schrödinger E. *What is life*. Cambridge: University Press; 1993. p. 121. [[Google Scholar](#)]
- [14].Lohrey A. The evolution of consciousness: a new science. Princeton New Jersey: ICRL Press; 2018a. [[Google Scholar](#)]
- [15].Hiley B, Peat F, eds. Quantum implications: essays to honour David Bohm. London: Routledge; 1991. p. p 436. [[Google Scholar](#)]
- [16].Hossenfelder and Palmer. 2020. [accessed date: 2020 March 12]. Available from: <http://nautil.us/issue/83/intelligence/how-to-make-sense-of-quantum-physics>
- [17].Bohm D. Quantum theory. NY: Prentice-Hall; 1951. [[Google Scholar](#)]
- [18].Bohm D, Aharonov Y. Discussion of experimental proof for the paradox of Einstein, Rosen, and Podolsky. Phys Rev. 1957;108(4):1070–1076. [[Google Scholar](#)]
- [19].Bell JS. Speakable and unspeakable in quantum mechanics. Cambridge, England: Cambridge University Press; 1988. p. 14. [[Google Scholar](#)]
- [20].Bell J. On the Einstein-Podolsky-Rosen paradox. Physics. 1964;1:195–200. [[Google Scholar](#)]
- [21].Bell J. On the problem of hidden variables in quantum mechanics. Rev Mod Phys. 1966;38:447–452. [[Google Scholar](#)]
- [22].Freedman S, Clauser J. Experimental test of local hidden variables theory. Phys Rev Lett. 1972;28:934–941. [[Google Scholar](#)]
- [23].Aspect A, Grangier P, Roger G. Experimental realization of Einstein-Podolsky-Rosen-Bohm gedankenexperiment: A new violation of Bell's inequalities. Phys Rev Lett. 1982;49:91–94. [[Google Scholar](#)]

- [24].Tittel W, Brendel J, Zbinden H, et al. Violation of Bell inequalities by photons more than 10 km apart. Phys Rev Lett. 1998;81(17):3563–3566. [[Google Scholar](#)]
- [25].Ma X-S, Zotter S, Kofler J, et al. Experimental delayed-choice entanglement swapping. Nat Phys. 2012;8(6):479–484. [[Google Scholar](#)]
- [26].Megidish E, Halevy A, Shacham T, et al. Entanglement swapping between photons that have never coexisted. Phys. Rev. Lett. 2013;110(21):210403. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- [27].Lemos GB, Borish V, Cole GD, et al. Quantum imaging with undetected photons. Nature. 2014;512(7515):409–412. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- [28].Hensen B, Bernien H, Dréau AE, et al. Loophole-free Bell Inequality violation using electron spins separated by 1.3 kilometres. Nature. 2015;526(7575):682–686. . [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- [29].Conway J, Kochen S. The free will theorem. Foundations Phys. 2006;36(10):1441–1473. [[Google Scholar](#)]
- [30].Conway JH, Kochen S. The strong free will theorem. Notices AMS. 2009;56(2):226–232. [[Google Scholar](#)]
- [31].Wheeler N. The ‘Free Will Theorem’ of Conway & Kochen. Reed College Physics Department; 2009. Available from: <https://www.reed.edu/physics/faculty/wheeler/documents/Quantum%20Mechanics/Miscellaneous%20Essays/Free%20Will%20Theorem/Free%20Will%20Theorem.pdf>
- [32].Hodgson D. Chapter 7: science and determinism. Published to Oxford Scholarship Online. 2012. [[Google Scholar](#)]
- [33].Goldstein S, Tausk DV, Tumulka R, et al. What does the free will theorem actually prove? Notices AMS. 2010;57(1):1451–1453. [[Google Scholar](#)]
- [34].Fuchs CA. On participatory realism. In: Durham IT, Rickles D, editors. Information and interaction: Eddington, wheeler, and the limits of knowledge. US: Springer; 2017; arXiv:1601.04360. [[Google Scholar](#)]
- [35].Bostrom N. Are we living in a computer simulation? Philosophical Quarter. 2003;53(211):243–255. [[Google Scholar](#)]
- [36].Irwin K, Amaral M, Chester D. The self-simulation hypothesis interpretation of quantum mechanics. Entropy. 2020;22(2):247. [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]

- [37].Berkovitz J. Action at a distance in quantum mechanics, section 3. The Stanford Encyclopedia of Philosophy (Spring 2016 Edition). 2016.
- [38].Rauscher E. Non-locality as a fundamental principle of reality: bell's theorem and space-like interconnectedness, cosmos and history. J Nat Social Philosophy. 2017. 13(2):204. [[Google Scholar](#)]
- [39].von Neumann J. The mathematical foundations of quantum mechanics. Princeton NJ: Princeton University Press; 1932. (English translation 1955, *Mathematical Foundations of Quantum Mechanics*). [[Google Scholar](#)]
- [40].Wigner E, Margenau H. Symmetries and reflections, scientific essays. Am J Phys. 1967;35(12):1169–1170. [[Google Scholar](#)]
- [41].London F, Bauer E. La théorie de l'observation en mécanique quantique, English translation. In: Wheeler JA, Zurek WH, editors. Quantum theory and measurement. Vol. 1983. Princeton NJ: Princeton University Press; 1939. p. 217–259. [[Google Scholar](#)]
- [42].Ball P. BBC Earth; 2017. February 16. [accessed date: 2017 February 16]. Available from: <http://www.bbc.com/earth/story/20170215-the-strange-link-between-the-human-mind-and-quantum-physics>.
- [43].Ball P. Beyond weird. Penguin Random House; 2018. p. 116. [[Google Scholar](#)]
- [44].Esfeld M. Essay review: Wigner's view of physical reality. In: Studies in history and philosophy of modern physics. Vol. 30B. Elsevier Science Ltd; 1990. p. 145–154. [[Google Scholar](#)]
- [45].Zalta EN, ed. n.d.. <https://plato.stanford.edu/archives/spr2016/entries/qm-action-distance/> [[Google Scholar](#)]
- [46].McEvoy P. Neils Bohr, reflections on subject and object, the theory of interacting systems. Vol. 1. San Francisco: Microanalytix; 1997. [[Google Scholar](#)]
- [47].Ball P. Beyond weird. London: The Bodley Head; 2018. p. 184. [[Google Scholar](#)]
- [48].Augusiak R, Demianowicz M, Tura J, et al. Entanglement and nonlocality are inequivalent for any number of parties. Phys Rev Lett. 2015;115(3):030404. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- [49].Berkovitz J. Action at a distance in quantum mechanics. Stanford Encyclopedia of Philosophy (Spring 2016 Edition). 2016/2007.

- [50].Morgart E. The theory of everything has nine dimensions. USA Today Magazine. 2014. January, pp 66–68
- [51].Jacques V, Wu E, Grosshans F, et al. The experimental realisation of Wheeler’s delayed-choice Gedanken experiment. Science. 2007;315(5814):966–968. [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
- [52].Radin D, Michel L, Galdamez K, et al. Consciousness and the double-slit interference pattern: six experiments. Phys Essays. 2012;25(2):157–171. [\[Google Scholar\]](#)
- [53].Lohrey A. Five realities, one truth: the human condition. Tasmania: Rishi Publishing; 2020. [\[Google Scholar\]](#)
- [54].Brenner ED, Stahlberg R, Mancuso S, et al. Plant neurobiology: an integrated view of plant signaling. Trends Plant Sci. 2006;11(8):413–419. www.sciencedirect.com [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
- [55].Lohrey A. The language virus of information theory. Edgescience. 2018b;35(5):5. [\[Google Scholar\]](#)
- [56].Young A. The reflective universe. Cambria, CA: Anodos Foundation, p. 10; 1999. [\[Google Scholar\]](#)
- [57].Handsteiner J. Cosmic bell test. Phys Rev Letts. 2017;118(7). DOI: 10.1103/PhysRevLett.118.060401. [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
- [58].Faye J. Copenhagen interpretation of quantum mechanics. Stanford Encyclopaedia of Philosophy. 2014. 25 June, p 4.
- [59].Bohr N. Essays 1932–1957 on atomic physics and human knowledge. Vol. II. Woodbridge: Ox Bow Press; 1958/1987. reprinted as *The Philosophical Writings of Niels Bohr* [\[Google Scholar\]](#)
- [60].Maudlin T. Why Bohm’s theory solves the measurement problem. Philosophy Sci. 1995;62(3):479–483. [\[Google Scholar\]](#)
- [61].Berkovitz J, Hemmo M. A modal interpretation in terms of relational properties. In: Demopoulous W, Pitowsky I, editors. Physical theory and its interpretation: essays in honour of Jeffery Bub. New York: Springer, 2006. Western Ontario Series in Philosophy of Science,1–28. [\[Google Scholar\]](#)
- [62].Wheeler JA. The ‘past’ and the ‘delayed choice’ double slit experiment. ed. ARM Marlo, Mathematical Foundation of Quantum Theory. NY: Academic Press; 1978. [\[Google Scholar\]](#)

- [63].Gribbin J. Search of Schrodinger's cat. Reading, Berks. UK: Corgi; 1994. [[Google Scholar](#)]
-

Explaining the Unexplainable with Non-local Consciousness

October 3, 2022
Science Team

How are we aware that we are aware? Consciousness remains a perplexing mystery for science. Today's neuroscience assumes that consciousness emerges from neurons interacting with each other in the brain. Yet, we still do not have an explanation for exactly how these neural interactions cause subjective experience or the experience of being conscious. Or, to be more specific, how interacting neurons could cause us to be aware of the profound and moving experiences of falling in love, enjoying a delightful piece of art, or feeling a wave of dazzling inspiration.

In a recently published [paper](#) in the mainstream science journal Frontiers of Psychology, we look at how science is currently tackling consciousness and possible alternative ways of viewing the problem that could help us solve this long-held human mystery.

The Brain Generates Consciousness

The current mainstream scientific worldview assumes that matter is the primary building block of the universe. This philosophy is called *materialism* and proposes that we can reduce every phenomenon into smaller physical parts to explain it. In this model, neurons work together to somehow *generate* consciousness. The field, however, cannot agree on just how this happens. In the paper, we review the main neuroscience theories of consciousness and propose that perhaps there are gaps in the approaches because some assumptions— such as that the brain generates consciousness — are wrong or incomplete.

Unexplained, Exceptional Human Experiences

Why? The materialist theories of consciousness do not easily explain certain human experiences and cause us to question whether consciousness is exclusively dependent on brain activity. For example, there are well-documented experiences of people perceiving information from distant locations, the future, and mental impressions from other people, which are difficult to explain with the current materialist models. Have you ever had a dream that came true exactly as you saw it? Impossible, says materialism.

Additionally, while there are many examples of correlations between neural integrity and human behavior, such as when damage to a brain region (e.g., from a stroke) causes changes in human behavior (e.g., losing the ability to speak), there are also many cases where the correlation is not so clear cut. For instance, there are documented cases where people, through a traumatic event or with no apparent cause, gain cognitive skills without previous training or experience in the skills, such as acquired and spontaneous savants, who suddenly gain exceptional musical or mathematical skills. There are also cases of people who can speak languages that they never previously learned. In yet another example, there are medical cases where the brain has degenerated to the point that barely any functional brain tissue remains, but where the individual inexplicably temporarily regains memory, orientation, and the ability to speak and respond with awareness. If brain function is perfectly correlated with human behavior and experiences, we would not expect to see these exceptional cases.

Can Non-local Consciousness Explain the Unexplainable?

To account for these exceptional cases where consciousness appears to extend beyond the physical brain and body in both space and time, alternative models of consciousness – or *non-local consciousness* theories – have been suggested. We review some of these theories in the new paper. These models do not assume that consciousness needs to be confined to and generated by the brain. This is hard to imagine using a model of reality where matter is the fundamental building block of the universe, but not if the model is flipped and consciousness becomes the fundamental building block instead, becoming more fundamental than matter and spacetime.

Yes, including non-local theories of consciousness may require a change in worldview, but this would not be the first time a shift in a worldview in science has happened. One notable example is the case of black holes. In 1915, the possibility of the existence of black holes was proposed by Karl Schwarzschild. But in 1939, Albert Einstein denied that black holes could exist. In the 1960s, Roger Penrose published models explaining how black

holes could form, causing other physicists to consider the possibility. Finally, a half-century later, astronomers observed a black hole. If the possibility of black hole existence had been permanently dropped, as Einstein proposed, we never would have discovered them.

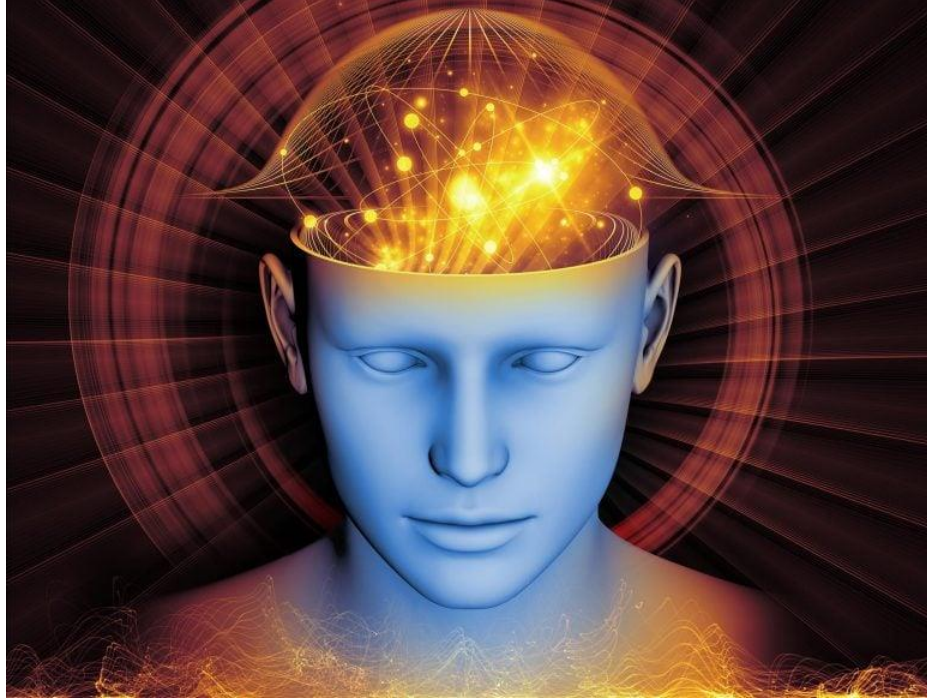
Including all Theories

This tension between resistance and advancement is common in science and helps us cautiously explore theories without abandoning past evidence. Yet, stories such as the one about black holes are a good reminder to approach all phenomena and theories with curiosity, humility, and the acknowledgment that at any given point in time, we do not have everything figured out. There is always more to discover about our universe. Including other models that do not fit the current scientific worldview can help advance our understanding of the world we live in and could even speed up scientific progress by explaining the unexplainable.

It is important for mainstream science to incorporate theories of non-local consciousness, such as [publishing these theories in journals such as Frontiers in Psychology as we did](#), because while materialism explains much in our world, it does not explain everything, including these fascinating exceptional cases. Non-materialist theories that suggest consciousness is fundamental and/or non-local may provide a road to understanding these phenomena. They may even help us better understand ourselves and our human experiences. To get there, we need testable theories and open minds to truly investigate this stunning, elusive thing we call consciousness.

Groundbreaking Study Affirms Quantum Basis for Consciousness: A Paradigm Shift in Understanding Human Nature

BY STACEY SCHMEIDEL, WELLESLEY COLLEGE SEPTEMBER 14, 2024 [40 COMMENTS](#) 3 MINS READ



New research indicates that anesthesia's effectiveness via microtubule interaction supports a quantum theory of consciousness, potentially transforming neuroscientific approaches and treatments.

A groundbreaking study has provided experimental evidence suggesting a quantum basis for consciousness.

By demonstrating that drugs affecting microtubules within neurons delay the onset of unconsciousness caused by anesthetic gases, the study supports the quantum model over traditional classical physics theories. This quantum perspective could revolutionize our understanding of consciousness and its broader implications, potentially impacting the treatment of mental illnesses and our understanding of human connection to the universe.

Exploring the Quantum Basis of Consciousness

For decades, one of the most fundamental and vexing questions in neuroscience has been: what is the physical basis of consciousness in the brain? Most researchers favor classical models, based on classical physics, while a minority have argued that consciousness must be quantum in nature, and that its brain basis is a collective quantum vibration of "microtubule" proteins inside neurons.

New research by Wellesley College professor Mike Wiest and a group of Wellesley College undergraduate students has yielded important experimental results relevant to this debate, by examining how anesthesia affects the brain. Wiest and his research team found that when they gave rats a drug that binds to microtubules, it took the rats significantly longer to fall unconscious under an anesthetic gas. The research team's microtubule-binding drug interfered with the anesthetic action, thus supporting the idea that the anesthetic acts on microtubules to cause unconsciousness.

"Since we don't know of another (i.e., classical) way that anesthetic binding to microtubules would generally reduce brain activity and cause unconsciousness," Wiest says, "this finding supports the quantum model of consciousness."

Implications for the Future of Neuroscience

It's hard to overstate the significance of the classical/quantum debate about consciousness, says Wiest, an associate professor of neuroscience at Wellesley. "When it becomes accepted that the mind is a quantum phenomenon, we will have entered a new era in our understanding of what we are," he says. The new approach "would lead to improved understanding of how anesthesia works, and it would shape our thinking about a wide variety of related questions, such as whether coma patients or non-human animals are conscious, how mysterious drugs like lithium modulate conscious experience to stabilize mood, how diseases like Alzheimer's or schizophrenia affect perception and memory, and so on."

More broadly, a quantum understanding of consciousness "gives us a world picture in which we can be connected to the universe in a more natural and holistic way," Wiest says. Wiest plans to pursue future research in this field, and hopes to explain and explore the quantum consciousness theory in a book for a general audience.

Broader Research Impact

Wellesley students who co-authored the paper with Wiest are Sana Khan '25, Yixiang Huang '25, Derin Timucin '27, Shantelle Bailey '24, Sophia Lee '23, Jessica Lopes '26, Emeline Gaunce '26, Jasmine Mosberger '25, Michelle Zhan '24, Bothina Abdelrahman '26 and Xiran Zeng '27. Published on September 1 in the journal *eNeuro*, the Wellesley study demonstrates that anesthesia works by binding to microtubules inside neurons, thus providing important evidence for a quantum theory of consciousness while reviving a focus on microtubules in anesthesia.

eNeuro is the Society for Neuroscience's open-access journal.

Reference: "Microtubule-Stabilizer Etoposide B Delays Anesthetic-Induced Unconsciousness in Rats" by Sana Khan, Yixiang Huang, Derin Timuçin, Shantelle Bailey, Sophia Lee, Jessica Lopes, Emeline Gaunce, Jasmine Mosberger, Michelle Zhan, Bothina Abdelrahman, Xiran Zeng and Michael C. Wiest, 15 August 2024, *eNeuro*.

[DOI: 10.1523/ENEURO.0291-24.2024](https://doi.org/10.1523/ENEURO.0291-24.2024)

Physicists explain how the brain might connect to the quantum realm

Down the rabbit hole ...

BY CRISTIANE DE MORAIS SMITH

UPDATED: FEB. 20, 2024

ORIGINALLY PUBLISHED: JULY 25, 2021



One of the most important open questions in science is how our consciousness is established. In the 1990s, [long before winning](#) the 2020 Nobel Prize in Physics for his [prediction of black holes](#), physicist Roger Penrose teamed up with anesthesiologist Stuart Hameroff to propose an ambitious answer.

[They claimed](#) that the brain's [neuronal](#) system forms an intricate network and that the consciousness this produces should obey the rules of [quantum mechanics](#) — the theory that determines how tiny particles like electrons move around. This, they argue, could explain the mysterious complexity of human consciousness.

Penrose and Hameroff were met with incredulity. Quantum mechanical laws are usually only found to apply at [very low temperatures](#). Quantum computers, for example, currently operate at around [-272°C](#). At higher temperatures, classical mechanics takes over.

Since our body works at room temperature, you would expect it to be governed by the classical laws of physics. For this reason, the quantum consciousness theory has been [dismissed outright](#) by many scientists — though others are [persuaded supporters](#).

Instead of entering into this debate, I decided to join forces with colleagues from China, led by Professor Xian-Min Jin at Shanghai Jiaotong University, to test some of the principles underpinning the quantum theory of consciousness.

In [our new paper](#), we've investigated how quantum particles could move in a complex structure like the brain – but in a lab setting. If our findings can one day be compared with activity measured in the brain, we may come one step closer to validating or dismissing Penrose and Hameroff's [controversial theory](#).

Brains and fractals

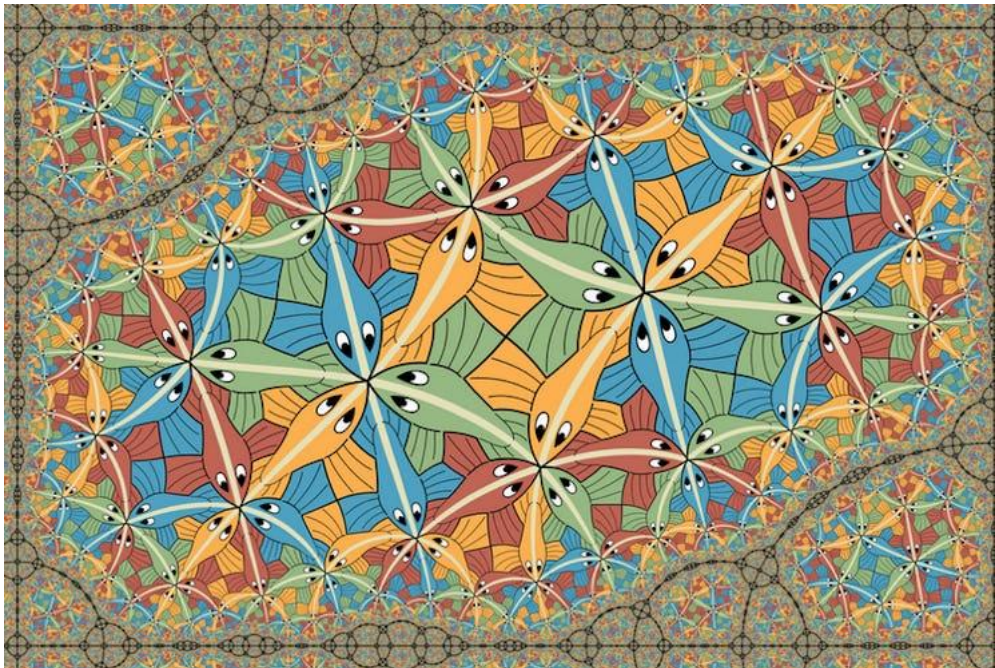
Our brains are composed of cells called neurons, and their combined activity is believed to generate consciousness. Each neuron contains [microtubules](#), which transport substances to different parts of the cell. The Penrose-Hameroff theory of quantum consciousness argues that microtubules are structured in a [fractal pattern](#) which would enable quantum processes to occur.

Fractals are structures that are neither two-dimensional nor three-dimensional but are instead some fractional value in between. In mathematics, fractals emerge as [beautiful patterns](#) that repeat themselves infinitely, generating what is seemingly impossible: a structure that has a finite area, but an infinite perimeter.

This might sound impossible to visualize, but fractals actually occur frequently [in nature](#). If you look closely at the florets [of cauliflower](#) or the branches [of ferns](#), you'll see that they're both made up of the same basic shape repeating itself over and over again, but at smaller and smaller scales. That's a key characteristic of fractals.

The same happens if you look inside your own body: the structure of [your lungs](#), for instance, is fractal, as are the [blood vessels](#) in your circulatory system. Fractals also feature in the enchanting repeating artworks of [MC Escher](#) and [Jackson Pollock](#), and they've been used for decades in technology, such as in the [design of antennas](#).

These are all examples of classical fractals — fractals that abide by the laws of classical physics rather than quantum physics.



This extension of Escher's Circle Limit III shows its fractal, repeating nature.

[VLADIMIR-BULATOV/DEVIANTART](#)

It's easy to see why fractals have been used to explain the complexity of human consciousness. Because they're infinitely intricate, allowing complexity to emerge from simple repeated patterns, they could be the structures that support the mysterious depths of our minds.

But if this is the case, it could only be happening on the quantum level, with tiny particles moving in fractal patterns within the brain's neurons. That's why Penrose and Hameroff's proposal is called a theory of "quantum consciousness".

Quantum consciousness

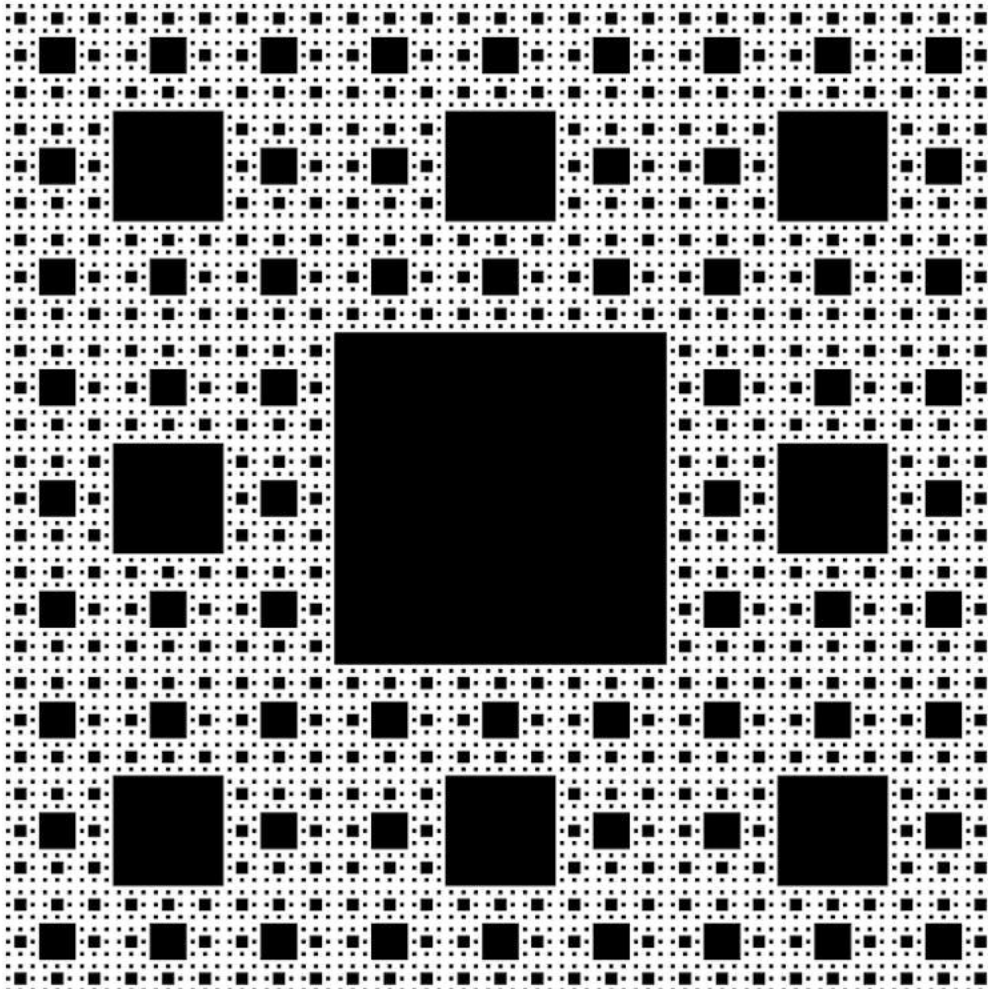
We're not yet able to measure the behavior of quantum fractals in the brain — if they exist at all. But advanced technology means we can now measure quantum fractals in the lab. In [recent research](#) involving a [scanning tunneling microscope](#) (STM), my colleagues at Utrecht and I carefully arranged electrons in a fractal pattern, creating a quantum fractal.

When we then measured the wave function of the electrons, which describes their quantum state, we found that they too lived at the fractal dimension dictated by the physical pattern we'd made. In this case, the pattern we used on the quantum scale was the [Sierpiński triangle](#), which is a shape that's somewhere between one-dimensional and two-dimensional.

This was an exciting finding, but STM techniques cannot probe how quantum particles move – which would tell us more about how quantum processes might occur in the brain. So in [our latest research](#), my colleagues at Shanghai Jiaotong University and I

went one step further. Using state-of-the-art photonics experiments, we were able to reveal the quantum motion that takes place within fractals in unprecedented detail.

We achieved this by [injecting photons](#) (particles of light) into an artificial chip that was painstakingly engineered into a tiny Sierpiński triangle. We injected photons at the tip of the triangle and watched how they spread throughout its fractal structure in a process called [quantum transport](#). We then repeated this experiment on two different fractal structures, both shaped as squares rather than triangles. And in each of these structures, we conducted hundreds of experiments.



We also conducted experiments on a square-shaped fractal called the Sierpiński carpet. [JOHANNES RÖSSEL/WIKIMEDIA](#)

Our observations from these experiments reveal that quantum fractals actually behave in a different way to classical ones. Specifically, we found that the spread of light across a fractal is governed by different laws in the quantum case compared to the classical case.

This new knowledge of quantum fractals could provide the foundations for scientists to experimentally test the theory of quantum consciousness. If quantum measurements

are one day taken from the human brain, they could be compared against our results to definitely decide whether consciousness is a classical or a quantum phenomenon.

Our work could also have profound implications across scientific fields. By investigating quantum transport in our artificially designed fractal structures, we may have taken the first tiny steps towards the unification of physics, mathematics, and biology, which could greatly enrich our understanding of the world around us as well as the world that exists in our heads.

Consciousness: here, there and everywhere?

[Giulio Tononi](#)¹, [Christof Koch](#)^{2,✉}

- Author information
- Article notes
- Copyright and License information

PMCID: PMC4387509 PMID: [25823865](#)

Abstract

The science of consciousness has made great strides by focusing on the behavioural and neuronal correlates of experience. However, while such correlates are important for progress to occur, they are not enough if we are to understand even basic facts, for example, why the cerebral cortex gives rise to consciousness but the cerebellum does not, though it has even more neurons and appears to be just as complicated. Moreover, correlates are of little help in many instances where we would like to know if consciousness is present: patients with a few remaining islands of functioning cortex, preterm infants, non-mammalian species and machines that are rapidly outperforming people at driving, recognizing faces and objects, and answering difficult questions. To address these issues, we need not only more data but also a theory of consciousness—one that says what experience is and what type of physical systems can have it. Integrated information theory (IIT) does so by starting from experience itself via five phenomenological axioms: *intrinsic existence*, *composition*, *information*, *integration* and *exclusion*. From these it derives five postulates about the properties required of physical mechanisms to support consciousness. The theory provides a principled account of both the quantity and the quality of an individual experience (a quale), and a calculus to evaluate whether or not a particular physical system is conscious and of what. Moreover, IIT can explain a range of clinical and laboratory findings, makes a number of testable predictions and extrapolates to a number of problematic conditions. The theory holds that consciousness is a fundamental property possessed by physical systems having specific causal properties. It predicts that consciousness is graded, is common among biological organisms and can occur in some very simple systems. Conversely, it predicts that feed-forward networks, even complex ones, are not conscious, nor are aggregates such as groups of individuals or heaps of sand. Also, in sharp contrast to widespread functionalist beliefs, IIT implies that digital computers, even if their behaviour were to be functionally equivalent to ours, and even if they were to run faithful simulations of the human brain, would experience next to nothing.

Keywords: mind body problem, causation, existence, neuronal correlates of consciousness, awareness, cerebral cortex

1. Consciousness: here, there and everywhere?

I know I am conscious: I am seeing, hearing, feeling something *here*, inside my own head. But is consciousness—subjective experience—also *there*, not only in other

people's heads, but also in the head of animals? And perhaps *everywhere*, pervading the cosmos, as in old panpsychist traditions and in the Beatles' song? While these kinds of questions may seem scientifically inappropriate, we argue below that they can be approached in a principled and testable manner. Moreover, obtaining an answer is urgent, not only because of difficult clinical cases and in our interactions with other species but also because of the advent of machines that are getting closer to passing the Turing test—computers programmed to perform many tasks as well as us, and often far better than some brain-damaged patients.

2. Here

That I am conscious, here and now, is the one fact I am absolutely certain of—all the rest is conjecture. This is, of course, the gist of the most famous deduction in Western thought, Descartes' *je pense, donc je suis*. Everything else—what I think I know about my body, about other people, dogs, trees, mountains and stars, is inferential. It is a reasonable inference, corroborated first by the beliefs of my fellow humans and then by the intersubjective methods of science. Yet consciousness itself—the central fact of existence—still demands a rational explanation.

The past two centuries of clinical and laboratory studies have revealed an intimate relationship between the conscious mind and the brain, but the exact nature of this relationship remains elusive. Why is the brain associated with consciousness but not the liver or the heart, as previous cultures believed? Why certain parts of the brain and not others? Why is consciousness lost in some stages of sleep? Why does red feel like red and not like the sound of a violin? Is consciousness just an epiphenomenon, or does it have a function? Can computers be conscious? Could a system behave like us and yet be devoid of consciousness—a zombie? Such questions seem to resist the empirical, reductionist approach that has been so successful for other aspects of the natural world. Nevertheless, thanks to experimental and theoretical progress in the past decades [1–5], we are in a better position to understand which systems under which conditions can be conscious. That is, the study of consciousness is becoming a science. In doing so, it is leaving behind the defeatist dictum of the physiologist Emil du Bois-Reymond, *ignoramus et ignorabimus* (we don't know and never will), espousing instead the upbeat maxim of the mathematician David Hilbert, *Wir müssen wissen—wir werden wissen* (we must know and we will know).

3. There

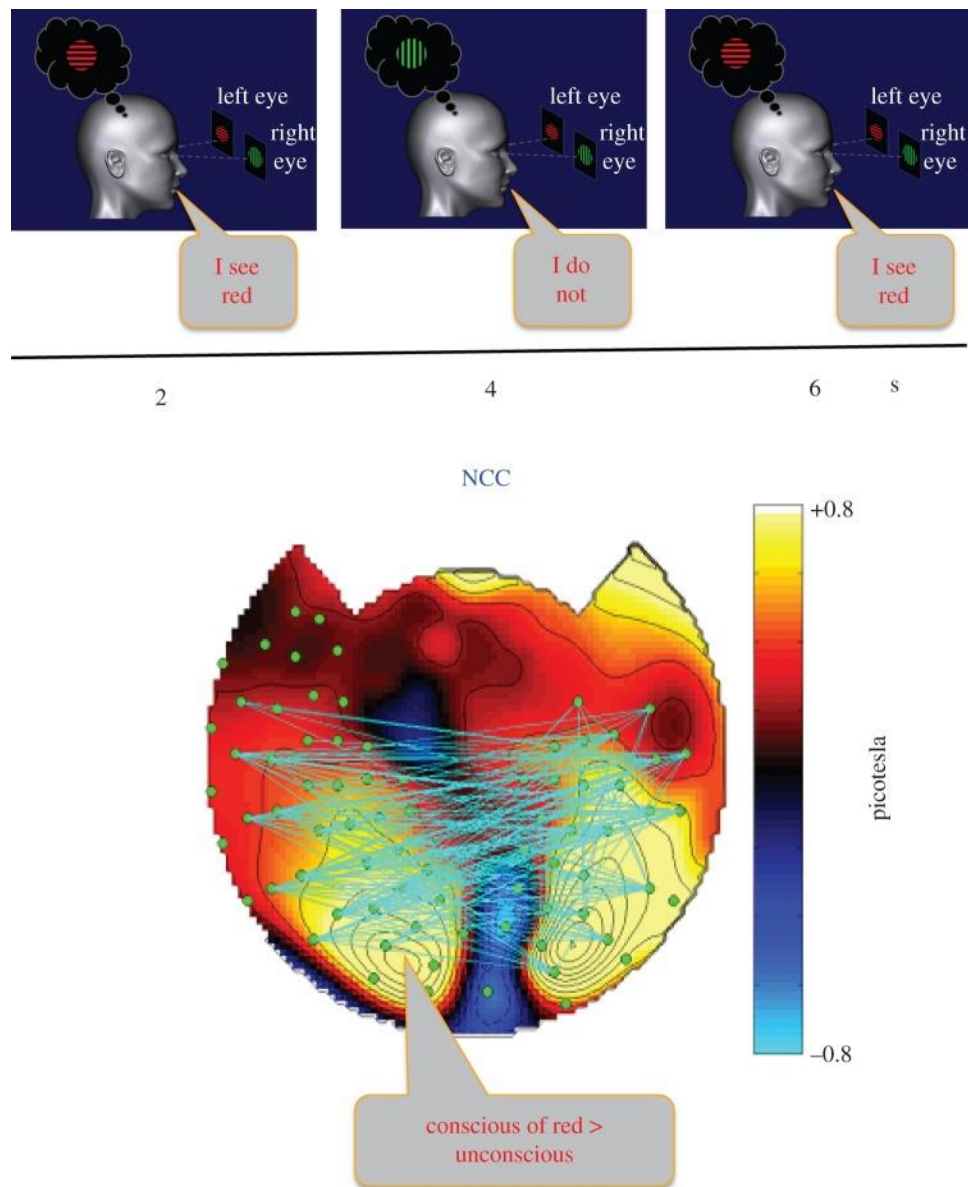
We usually grant consciousness to others—of the same kind we experience in the privacy of our own mind—if they can tell us what they feel, or if they look and behave more or less like us. However, we become less and less confident in attributing consciousness to those who cannot talk about their experiences, such as infants or severely brain injured patients. Many assume that animals closely related to *homo sapiens*—apes and other primates—are conscious, though presumably less than we are, based on the similarity of their behaviour and their brain. But should we attribute

experience to all mammals,¹ to all vertebrates, to invertebrates such as cephalopods and bees or even to all multi-cellular animals? What about cultured organoids that mimic the cellular organization of the developing human brain [8]? And finally, what about the sophisticated machines that run software designed to substitute for conscious humans in many complicated tasks?

(a). Behavioural correlates of consciousness and reportability

Traditionally, we assess consciousness by observing behaviour ([figure 1a](#)). If someone is awake and acts meaningfully, we have little doubt he is conscious. If he speaks, and especially if he can answer questions about what he is conscious of, we are fully confident. In the laboratory, the ability to report one's experiences has become the gold standard for judging the presence of consciousness. Reportability is often reduced to a binary forced choice, in which the subject pushes one of two buttons for 'seen' versus 'not seen', or 'angry face' versus 'happy face'. One can also ask subjects how confident they are in their judgements (*confidence rating* [10]), ask them to further describe their experiences (*perceptual awareness scale* [11,12]) or get them to make an economic judgement following each response (*post-decision wagering* [13]). These kinds of meta-cognitive and confidence reports can also be obtained from trained monkeys and other animals, with so many similarities to our own reports that there is little doubt as to the presence of consciousness [14,15] (but see [16]).

Figure 1.



[Open in a new tab](#)

Behavioural (BCC) and neuronal correlates of consciousness (NCC). The top row shows a schematic diagram of a binocular rivalry experiment. A horizontal red grating is shown to the left eye and a vertical green grating to the right eye throughout the experiment (courtesy of Naotsugu Tsuchiya and Olivia Carter). The subject does not see a juxtaposition of both stimuli but experiences either the red grating or the green one, switching back and forth every few seconds. Even if the stimuli do not change, what one sees consciously does, as is inferred by the subject's report. The bottom row shows the results of an experiment using magnetoencephalography (MEG), in which the red grating was flashed at one frequency and the green one at another. Yellow indicates areas of the cortex (seen from the top) that had more power at the frequency of the red grating when it was experienced than when it was not. The cyan lines indicate increased coherence (synchronization) between distant brain regions associated with experiencing the grating (from [9]).

But behaviour can be misleading: a person may walk and speak in her sleep, yet it is quite dubious whether she is experiencing anything. Or a person can be asleep, immobile, silent and unresponsive, yet she may be dreaming—vividly conscious of an imaginary environment. In such cases, reportability can be used as retrospective evidence of consciousness, by waking up the sleeper to obtain a ‘dream report’. However, reportability, too, can be problematic. Since we obviously experience things in dreams whether or not we are woken up to report them, we should accept the possibility that in certain situations consciousness can be present even if it is not reported [17,18]. Moreover, insisting on reportability elevates language to a king-maker role, which makes inferring consciousness in non-verbal infants, preterm babies, fetuses or animals problematic.² Clearly, if we want to understand what is really going on, we must also investigate the brain mechanisms that underlie consciousness.

(b). Neural correlates of consciousness

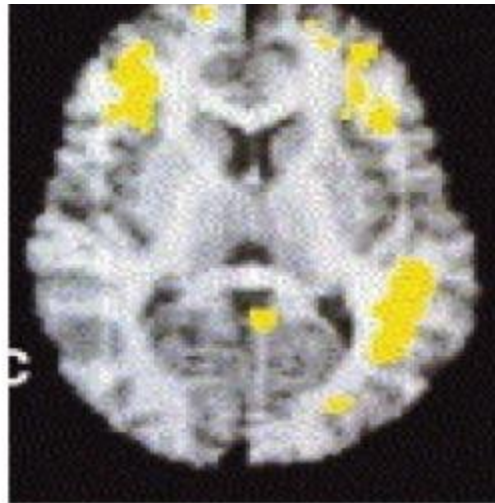
The neural correlates of consciousness (NCC) have been defined as the minimal neural mechanisms that are jointly sufficient for any one conscious percept, thought or memory, under constant background conditions (figure 1b) [1,23,24]. The latter are the distal or proximal *enabling factors* that must be present for any conscious experience to occur—the heart must beat and supply the brain with oxygenated blood, various nuclei in the midbrain reticular formation and brainstem must be active [25–27], cholinergic release needs to occur within the cortico-thalamic complex [28] and so on.

Every experience will have an associated NCC: one for seeing a red patch, another one for hearing a high C. Inducing the NCC by manipulating the relevant neuronal populations via magnetic stimulation, optogenetics or other means will give rise to the associated conscious percept. Interfering with the NCC by disabling the underlying neural circuits will eliminate the percept.

The NCC are typically assessed by determining which aspects of neural function change depending on whether a subject is conscious or not, as established using behavioural reports. This can be done by considering a global change in the level of consciousness, as when awareness is lost during deep sleep or general anaesthesia [29,30]. Or it can be done by considering changes in a particular content of consciousness, as when a subject's awareness of a particular stimulus is experimentally manipulated (‘seen’ versus ‘not seen’ [31,32]). In optimally controlled experiments, the stimulus and the behavioural report (such as a button press) are kept constant while the subject sometimes sees the percept and sometimes does not [3,33,34]. Once a particular NCC has been sufficiently validated, it can be used to extrapolate to situations in which reports are not available. Both functional brain imaging in magnetic scanners and as high-density electroencephalography (EEG) recordings from outside the skull have been put to use to track down the footprints of consciousness in the brain of healthy adult observers. Popular candidates include strong activation of high level fronto-parietal cortices (figure 1b), high-frequency electrical activity in the gamma range (35–80 Hz), and the occurrence of an EEG event known as the P300 wave [1,3,29].

However, there is still no consensus on whether any of these signs can be treated as reliable ‘signatures’ of consciousness. In particular, there can be consciousness without frontal cortex involvement [35–37], gamma activity without consciousness [38], such as during anaesthesia [39,40], and consciousness without a frontal P300, for example, during dreaming sleep [41,42]. Moreover, it is likely that many of the signatures proposed as possible NCC may actually be correlates of neural activity that is needed leading up to a conscious percept [43,44], or for giving a report following a conscious percept [36,37,44], rather than for having an experience. A major challenge is to keep constant cognitive functions such as selective attention, memory, decision making and task monitoring, in order to isolate the ‘naked’ substrate of consciousness at the neuronal level [45,46]. Finally, NCC obtained in healthy adults may or may not apply to brain-damaged patients, to infants, to animals very different from us, not to mention machines (figure 2).

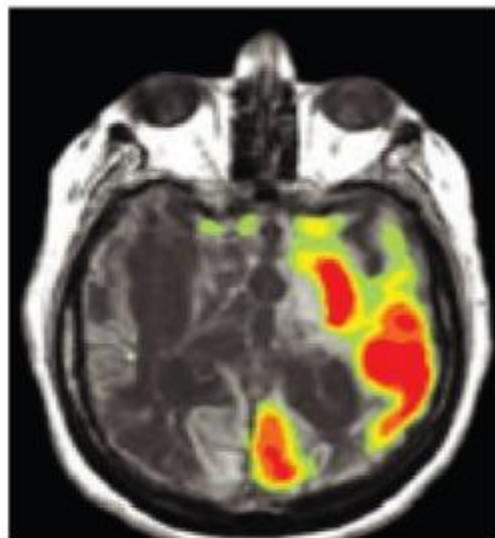
Figure 2.



sleepwalking



ketamine anaesthesia



brain 'islands', vegetative patient

Six instances in which it becomes progressively more difficult to infer the existence of consciousness, since the behavioural repertoire and the underlying mechanisms (brains) differ substantially from that of typical persons able to speak about their experiences ([figure 1](#)).

(c). Patients and infants

Patients with widespread cortical or thalamic damage pose a poignant challenge. Emergency room personnel quickly evaluate the severity of a head injury behaviourally by assigning a number to a patient's auditory, visual, verbal and motor functions as well as communication and arousal level. Various NCC, such as the presence of a P300 wave in response to a non-standard stimulus, are increasingly being used to complement the behavioural assessment and occasionally modify the diagnosis. In some cases, NCC can be decisive. Thus, if a patient who lies mute and immobile can nevertheless respond to commands by appropriately activating certain brain areas, it is fair to conclude that she is conscious [\[47\]](#). Yet most of the proposed 'signatures' of consciousness are inadequate. For example, the P300 wave is absent in many minimally conscious patients and even in some brain-damaged patients who can communicate [\[48\]](#). And what should one make of patients in whom, amidst widespread destruction and inactivity, one or a few isolated cortical areas may show signs of metabolic activation and electrophysiological 'markers' of consciousness [\[49\]](#)? Is an island of functioning brain tissue sufficient for generating a limited kind of awareness, maybe just awareness of sound or of pain? In other words, 'what is it like' to be a brain island, if it feels like anything at all? And how big must the island be to qualify?

By the same token, what is it like to be a newborn baby with an immature brain and restricted connectivity among cortical structures [\[50\]](#)? Again, considering NCC can be helpful: for example, a wave resembling the P300 wave has been reported in six to 16 months old infants, although weaker, more variable and delayed compared with adults [\[51\]](#). But does this mean that newborn and preterm babies or even fetuses experience nothing because they do not show a P300?

(d). Animals

The problem becomes even more acute when turning to other species. The study of consciousness in nature has been hindered for centuries by a strong belief in human exceptionalism. Yet the range and complexity of animal behaviour has laid rest to this belief, at least among biologists. This is particularly true for mammals. In psychophysical tasks involving simple button presses, trained macaque monkeys act very similarly to human volunteers, including signalling when they do not see anything [\[14\]](#). Visual recognition of self, meta-cognition (knowing one's mind), theory of mind, empathy and long-range planning have all been demonstrated in primates, rodents and other orders [\[52\]](#).

It is also difficult to find anything exceptional about the human brain [53]. Its constitutive genes, synapses, neurons and other cells are similar to those found in many other species. Even its size is not so special, as elephants, dolphins and whales have even bigger brains [54]. Only an expert neuroanatomist, armed with a microscope, can tell a grain-sized piece of neocortex of a mouse from that of a monkey or a human. Biologists emphasize this structural and behavioural continuity by distinguishing between *non-human* and *human* animals [55]. Given this continuity, it seems unjustified to claim that only one species has consciousness while everybody else is devoid of experience, is a zombie. It is far more likely that all mammals have at least some conscious experiences, can hear the sounds and see the sights of life.

As we consider species that are progressively further removed from *Homo sapiens* in evolutionary and neuronal terms, the case for consciousness becomes more difficult to make. Two observations, one relating to complexity of behaviour and another one to complexity of the underlying nervous system, are critical. First, ravens, crows, magpies, parrots and other birds, tuna, coelacanths and other fish, octopuses and other cephalopods, bees and other members of the vast class of insects are all capable of sophisticated, learnt, non-stereotyped behaviours that we associate with consciousness if carried out by people [56–58]. Darwin himself set out ‘to learn how far the worms acted consciously’ and concluded that there was no absolute threshold between ‘lower’ and ‘higher’ animals, including humans, which would assign higher mental powers to one but not to the other [59]. Second, the nervous systems of these species display a vast and ill-understood complexity. The bee contains about 800 000 nerve cells whose morphological and electrical heterogeneity rivals that of any neocortical neuron. These cells are assembled in highly nonlinear feedback circuits whose density is up to ten times higher than that of neocortex [60]. Thus, neural signatures of consciousness that have some validity in humans and other mammals may not apply at all to invertebrates.

On the other hand, the lessons learnt from studying the behavioural (BCC) and neuronal correlates of consciousness in people must make us cautious about inferring its presence in creatures very different from us, no matter how sophisticated their behaviour and how complicated their brain. Humans can perform complex behaviours—recognizing whether a scene is congruous or incongruous, controlling the size, orientation and strength of how one's finger should grip an object, doing simple arithmetic, detecting the meaning of words or rapid keyboard typing—in a seemingly non-conscious manner [61–66]. When a bee navigates a maze, does it do so like when we consciously deliberate whether to turn right or left, or rather like when we type on a keyboard? Similarly, consider that an extraordinarily complicated neuronal structure in our brain, the cerebellum, home to 69 of the 86 billion nerve cells that make up the human brain [54], apparently has little to do with consciousness. Patients who lose part or nearly all of their cerebellum owing to stroke or other trauma show ataxia, slurred speech and unsteady gait [67] but do not complain of a loss or diminution of consciousness. Is the bee's brain central complex more like the cerebellum or more like the cerebral cortex with respect to experience? Thus, the extent to which non-

mammalian species share with us the gift of subjective experience remains hard to fathom.³

(e). Machines

Difficulties in attributing sentience become even more apparent when considering digital computers. These have a radically different architecture and provenance from biological organisms shaped by natural selection. Owing to the relentless decrease in transistor size over the past 50 years and the concomitant exponential increase in computational power and memory capacity, present-day computers executing appropriate algorithms outperform us in many tasks that were thought to be the sole prerogative of the human mind. Prominent examples include IBM's Deep Blue that beat the reigning chess world master in 1997; another IBM computer, Watson, that can answer questions posed in spoken English and won the quiz show *Jeopardy* in 2011; smart phones that answer questions by speech; Google's driverless cars that have logged more than half a million miles on open roads; and machine vision algorithms for face detection in security and commercial applications [68]. People playing chess, supplying meaningful answers to questions, driving a car or picking out a face are assumed to be conscious. But should we say the same for these digital creatures?

4. Integrated information theory

Clearly, as we move away from people, BCC and NCC become progressively less helpful to establish the presence of consciousness. Even in the normal human brain, we need to understand *why* and *how* certain structures are associated with experience (the cerebral cortex or, possibly, the claustrum [69,70]) while others are not (the cerebellum), and why they do so under certain conditions (wake, dreams) and not others (deep sleep, seizures). Some philosophers have claimed that the problem of explaining how matter can give rise to consciousness may forever elude us, dubbing it the *Hard* problem [71–73]. Indeed, as long as one starts from the brain and asks how it could possibly give rise to experience—in effect trying to ‘distill’ mind out of matter [74], the problem may be not only hard, but almost impossible to solve. But things may be less hard if one takes the opposite approach: start from consciousness itself, by identifying its essential properties, and then ask what kinds of physical mechanisms could possibly account for them. This is the approach taken by integrated information theory (IIT) [75–79], an evolving formal and quantitative framework that provides a principled account for what it takes for consciousness to arise, offers a parsimonious explanation for the empirical evidence, makes testable predictions and permits inferences and extrapolations (table 1).⁴

Table 1.

Some terms used in integrated information theory (IIT).

Axioms. Properties of consciousness that are taken as self-evident. The only truths that, with Descartes, cannot be doubted and do not need proof. They are intrinsic existence, composition, information, integration and exclusion (figure 3, left).

Postulates. Assumptions, derived from axioms, about the physical substrates of consciousness (mechanisms must have cause–effect power, be irreducible, etc.), which can be formalized and form the basis of the mathematical framework of IIT. It is as yet unproven whether the mapping from axioms to postulates is unique. There are five postulates, matching the five axioms ([figure 3](#), right).

Element. A minimal component of a system, for example, a neuron in the brain or a logic gate in a computer, having at least two states, inputs that can affect those states and outputs that depend on them. Strictly speaking, such elements are *macro*-elements constituted of *micro*-elements such as molecules, which are constituted in turn of atoms and so on. IIT predicts that, if neurons are the relevant elements for consciousness, intrinsic cause–effect power within the system must be highest at the level of such macro-elements rather than at the level of the constituting micro-elements [[79](#)].

Mechanism. Any subset of elements within a system, first- and higher order, including the system itself, which has cause–effect power within the system.

Cause–effect repertoire. The probability distribution of potential past and future states of a system as informed by a mechanism in its current state.

Integrated information (ϕ). Information that is specified by a mechanism above and beyond the information specified by its (minimal) parts. ϕ measures the integration or irreducibility of the cause–effect repertoire specified by a mechanism.

MIP (minimum information partition). The partition that makes the least difference—in other words, the minimum ‘difference’ partition.

Complex. A set of elements within a system that specifies a local maximum of integrated conceptual information $\Phi^{\max}$. Only a complex exists as an entity from its own intrinsic perspective.

Concept. A mechanism and the maximally irreducible cause–effect repertoire it specifies, with its associated value of integrated information $\phi^{\max}$. The concept expresses the cause–effect power of a mechanism within a complex.

Conceptual structure. The set of all concepts specified by a system set with their respective $\phi^{\max}$ values, which can be plotted as a constellation of concepts in cause–effect space.

Cause–effect space (or qualia space). A high-dimensional space with one axis for each possible past and future state of the system in which a conceptual structure can be represented.

Integrated conceptual information (Φ). Conceptual information that is specified by a system above and beyond the conceptual information specified by its (minimal) parts. Φ measures the intrinsic integration or irreducibility of a constellation of concepts (integration at the system level), a non-negative number.

Quale. A conceptual structure specified by a complex in a state that is maximally irreducible intrinsically (synonymous with constellation in qualia space).

[Open in a new tab](#)

(a). Axioms: essential phenomenological properties of consciousness

Taking consciousness as primary, IIT first identifies *axioms* of experience ([figure 3](#), left), then derives a set of corresponding *postulates* ([figure 3](#), right) about its physical substrate [[77,80](#)]. The axioms of IIT are assumptions about our own experience that are the starting point for the theory. Ideally, axioms are essential (apply to all experiences), complete (include all the essential properties shared by every experience), consistent (lack contradictions) and independent (not derivable from each other). Whether the current set of five axioms are truly valid, complete and independent remains open.⁵ The five axioms are intrinsic existence, composition, information, integration and exclusion.

Figure 3.

axioms
essential properties of every experience

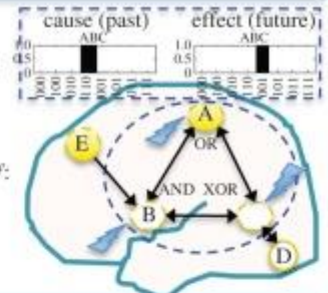
postulates
properties that physical systems (elements in a state)
must have to account for experience



intrinsic existence

consciousness exists intrinsically: each experience is real, and it exists from its own *intrinsic perspective*, independent of external observers (it is intrinsically real)

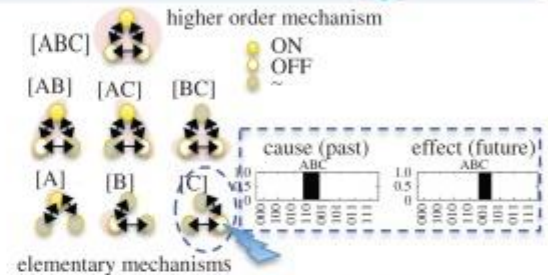
to account for experience, a system of mechanisms in a state must exist intrinsically. To exist, it must have *cause-effect power*; to exist from its own *intrinsic perspective*, independent of extrinsic factors, it must have cause-effect power *upon itself*: its present mechanisms and state must 'make a difference' to the probability of some past and future state of the system (its *cause-effect space*)



composition

consciousness is structured: each experience is composed of phenomenological distinctions, elementary or higher-order, which *exist* within it

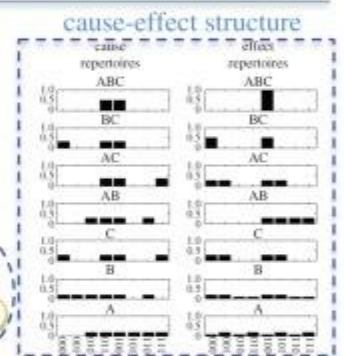
the system must be structured: subsets of system elements (composed in various combinations) must have cause-effect power upon the system



information

consciousness is specific: each experience is *the particular way it is* (it is composed of a specific set of specific phenomenological distinctions), thereby differing from other possible experiences (*differentiation*)

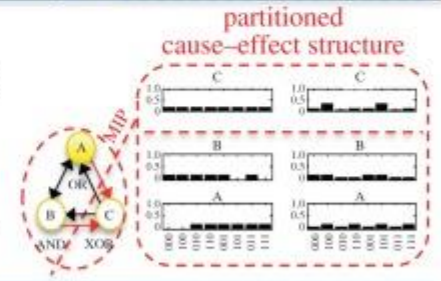
the system must specify a cause-effect structure that is *the particular way it is*: a specific set of specific cause-effect repertoires—thereby differing in its specific way from other possible structures (*differentiation*). A *cause-effect repertoire* specifies the probability of all possible causes and effects of a mechanism in a state. A *cause-effect structure* is the set of cause-effect repertoires specified by all subsets of system elements and expresses how the system gives an *actual form* to the space of possibilities



integration

consciousness is unified: each experience is *irreducible* to non-interdependent subsets of phenomenal distinctions

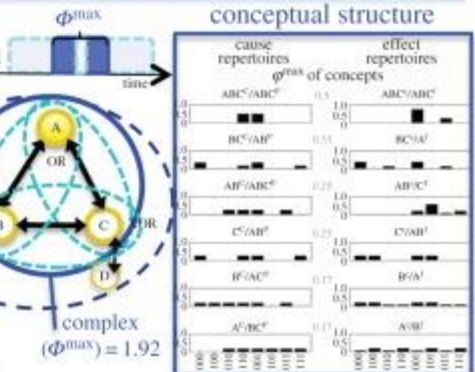
the cause-effect structure specified by the system must be unified: it must be intrinsically *irreducible* to that specified by non-interdependent sub-systems ($\Phi > 0$) across its weakest (unidirectional) link: MIP = minimum information partition



exclusion

consciousness is definite, in content and spatio-temporal grain: each experience has the set of phenomenal distinctions it has, not less or more, and flows at the speed it does, not faster or slower

the cause-effect structure specified by the system must be definite: specified over a *single set* of elements—not less or more—and spatio-temporal grains—not faster or slower; this is a cause-effect structure that is maximally irreducible intrinsically ($\Phi^{\max}$), called *conceptual structure*, made of maximally irreducible cause-effect repertoires (*concepts*)



Axioms and postulates of integrated information theory (IIT). The illustration is a colourized version of Ernst Mach's 'View from the left eye' [84]. See also the mechanisms in [figure 4](#).

(i). Intrinsic existence

Consciousness *exists*: my experience just *is*. Indeed, that my experience here and now exists—it is real or actual—is the only fact I am immediately and absolutely sure of, as Descartes realized four centuries ago. Moreover, my experience exists from its own *intrinsic* perspective, independent of external observers.

(ii). Composition

Consciousness is *structured*: each experience is composed of many *phenomenological distinctions*, elementary or higher order, which also exist. Within the same experience, for example, I may distinguish a book, a blue colour, a blue book and so on.

(iii). Information

Consciousness is *specific*: each experience is *the particular way it is*—it is composed of a specific set of specific phenomenal distinctions—thereby differing from other possible experiences (*differentiation*). Thus, an experience of pure darkness and silence is what it is because, among other things, it is not filled with light and sound, colours and shapes, there are no books, no blue books and so on. And being that way, it necessarily differs from a large number of alternative experiences I could have. Just consider all the frames of all possible movies: the associated visual percepts are but a small subset of all possible experiences.

(iv). Integration

Consciousness is *unified*: each experience is *irreducible* to non-interdependent subsets of phenomenal distinctions. Thus, I experience a whole visual scene, not the left side of the visual field independent of the right side (and vice versa). For example, the experience of seeing written in the middle of a blank page the word 'HONEYMOON' is irreducible to an experience of seeing 'HONEY' on the left plus the experience of seeing 'MOON' on the right. Similarly, seeing a blue book is irreducible to seeing a grey book plus the disembodied colour blue.

(v). Exclusion

Consciousness is *definite*, in content and spatio-temporal grain: each experience has the set of phenomenal distinctions it has, neither less (a subset) nor more (a superset), and it flows at the speed it flows, neither faster nor slower. Thus, the experience I am having is of seeing a body on a bed in a bedroom, a bookcase with books, one of which

is a blue book, but I am not having an experience with less content—say, one lacking the phenomenal distinction blue/not blue, or coloured/not coloured; nor am I having an experience with more content—say, one endowed with the additional phenomenal distinction high/low blood pressure. Similarly, my experience flows at a particular speed—each experience encompassing a hundred milliseconds or so—but I am not having experience that encompasses just a few milliseconds or instead minutes or hours.

(b). Postulates: properties that physical mechanisms must have to support consciousness

To parallel these axioms that capture the essential properties of every experience, IIT proposes a set of postulates concerning the requirements that must be satisfied by any physical system to account for experience ([figure 3](#), right). For simplicity, physical systems are considered as *elements in a state*, such as neurons or logic gates that are either ON or OFF. All that is required is that such elements have two or more internal states, inputs that can influence these states in a certain way and outputs that in turn depend on these states.

(i). Intrinsic existence

A system of mechanisms in a state must *exist intrinsically*. Specifically, in order to exist, it must have *cause–effect power*, as there is no point in assuming that something exists if nothing can make a difference to it, or if it cannot make a difference to anything [\[88\]](#).⁶ Moreover, to exist from its own intrinsic perspective, independent of external observers, it must have cause–effect power *upon* itself, independent of extrinsic factors ([figure 3](#), intrinsic existence). Cause–effect power can be established by considering a *cause–effect space* with an axis for every possible state of the system in the past (causes) and in the future (effects). Within this space, it is enough to show that an ‘intervention’ that sets the system in some initial state, keeping the state of the elements outside the system fixed (background conditions), can lead with probability different from chance to its present state (cause); conversely, setting the system to its present state leads with probability different from chance to some other state (effect).

(ii). Composition

The system must be *structured*: subsets of the elementary mechanisms of the system, *composed* in various combinations, also have cause–effect power within the system. Thus, if a system ABC comprises elements A, B and C ([figure 3](#), composition), any subset of elements, including A, B, C; AB, AC, BC; as well as the entire system, ABC, can compose a mechanism having cause–effect power. Composition allows for elementary (first-order) mechanisms to form distinct higher order mechanisms, and for multiple mechanisms to form a structure.

(iii). Information

The system must *specify* a cause–effect structure that is *the particular way it is*: a specific set of specific cause–effect repertoires—thereby differing from other possible ones (*differentiation*). A *cause–effect repertoire* characterizes in full the cause–effect power of a mechanism within a system by making explicit all its cause–effect properties. It can be determined by perturbing the system in all possible ways to assess how a mechanism in its present state makes a difference to the probability of the past and future states of the system. Together, the cause–effect repertoires specified by each composition of elements within a system specify a *cause–effect structure*. Consider for example, within the system ABC ([figure 3](#), information), the mechanism implemented by element C, an XOR gate with two inputs (A and B) and two outputs (the OR gate A and the AND gate B). If C is OFF, its cause repertoire specifies that, at the previous time step, A and B must have been either in the state OFF,OFF or in the state ON,ON, rather than in the other two possible states (OFF,ON; ON,OFF); and its effect repertoire specifies that the next time step B will have to be OFF, rather than ON. Its cause–effect repertoire is specific: it would be different if the state of C were different (ON), or if C were a different mechanism (say, an AND gate). Similar considerations apply to every other mechanism of the system, implemented by different compositions of elements. Thus, the cause–effect repertoire specifies the full cause–effect power of a mechanism in a particular state, and the cause–effect structure specifies the full cause–effect power of a system of mechanisms. Note that the notion of information in IIT differs substantially from that in communication theory or in common language, but it is faithful to its etymology: information refers to how a system of mechanisms in a state, through its cause–effect power, specifies a form (‘informs’ a conceptual structure) in the space of possibilities.

(iv). Integration

The cause–effect structure specified by the system must be *unified*: it must be intrinsically *irreducible* to that specified by non-interdependent sub-systems obtained by unidirectional partitions. Partitions are taken unidirectionally to ensure that cause–effect power is intrinsically irreducible—from the system’s intrinsic perspective—which implies that every part of the system must be able to both affect and be affected by the rest of the system. Intrinsic irreducibility can be measured as integrated information (‘big phi’ or Φ , a non-negative number), which quantifies to what extent the cause–effect structure specified by a system’s mechanisms changes if the system is partitioned (cut or reduced) along its minimum partition (the one that makes the least difference). For example, the system in [figure 3](#) is integrated, because partitioning it through its weakest link destroys several cause–effect repertoires and changes others (compare the cause–effect structure under ‘information’ and under ‘integration’ in [figure 3](#)). By contrast, if a system of mechanisms can be divided into two sub-systems and the partition makes no difference to the associated cause–effect structure, then the whole is reducible to those parts. Being intrinsically irreducible is another precondition for existence having to do with causation: there is no point in assuming that the whole exists in and of itself, if it has no cause–effect power above and beyond its parts. This postulate also applies to individual mechanisms: a subset of elements can contribute a specific aspect of

experience only if its cause–effect repertoire within the system is irreducible by the minimum partition of the mechanism ('small phi' or ϕ).

(v). Exclusion

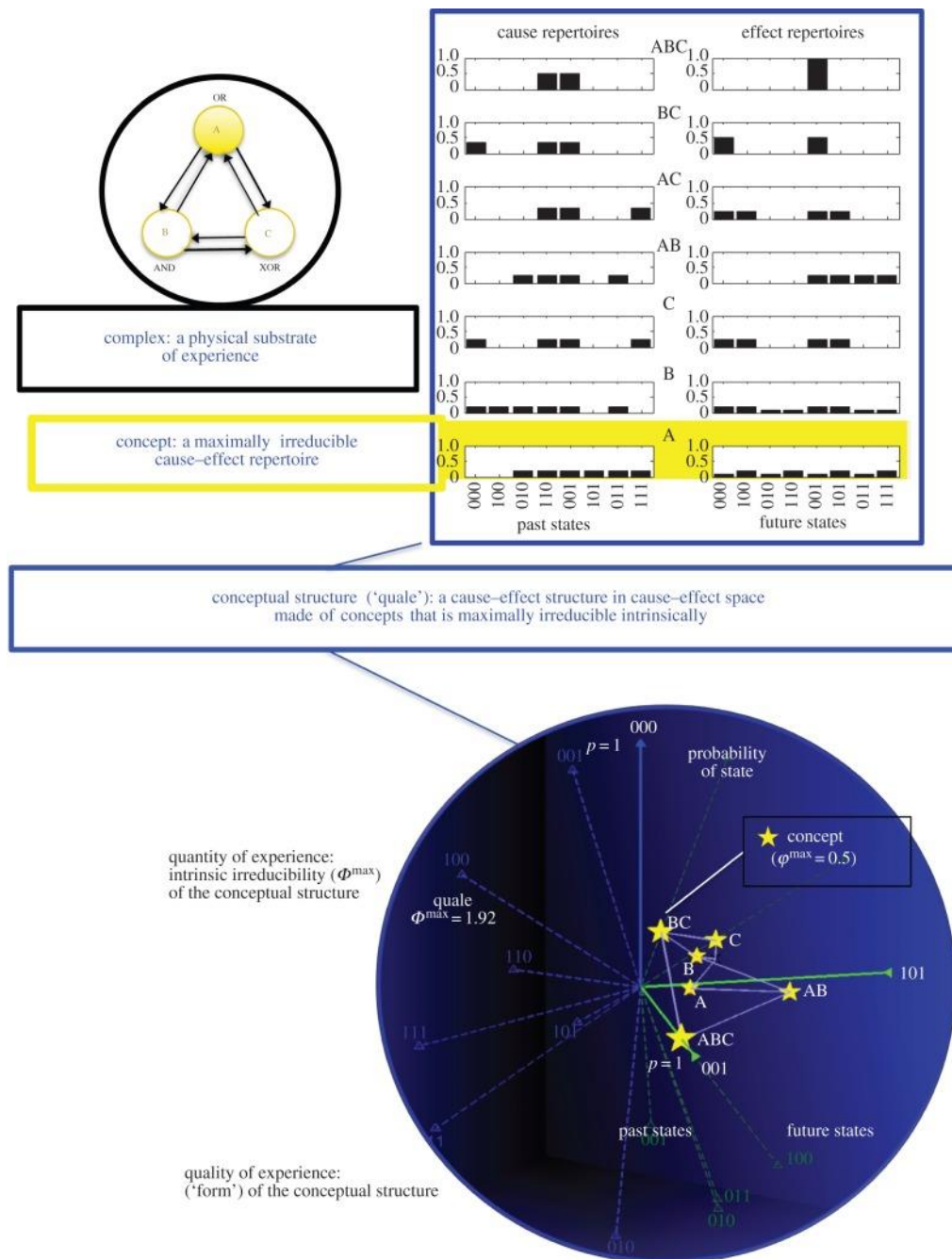
The cause–effect structure specified by the system must be *definite*: it is specified over a *single* set of elements—neither less nor more—the one over which it is *maximally irreducible* ($\phi^{\max}$) from its intrinsic perspective, thus laying maximal claim to existence. For example ([figure 3](#), exclusion), within ABCDE, many candidate systems could specify cause–effect structures, including AB, AC, BC, ABC, ABCD, ABCDE and so on. Among these, the system that specifies the cause–effect structure that is maximally irreducible intrinsically is the set of elements ABC, rather than any of its subsets or supersets. The exclusion postulate provides a sufficient reason why the contents of the experience should be what they are—neither less nor more. With respect to causation, this has the consequence that the 'winning' cause–effect structure excludes alternative cause–effect structures specified over overlapping elements: if a mechanism in a state (say A OFF) specifies a particular cause–effect repertoire within one system (ABC), it should not *additionally* specify an overlapping cause–effect repertoire as part of other, overlapping systems (say AB or ABCD), otherwise one would be counting multiple times the difference that mechanism makes. The exclusion postulate can be said to enforce Occam's razor (entities should not be multiplied beyond necessity): it is more parsimonious to postulate the existence of a single cause–effect structure over a system of elements—the one that is maximally irreducible—than a multitude of overlapping cause–effect structures whose existence would make no further difference. The exclusion postulate also applies to individual mechanisms: a subset of elements in a state specifies the cause–effect repertoire within the system that is maximally irreducible ($\phi^{\max}$), called a *core concept*, or *concept* for short. Again, it cannot *additionally* specify a cause–effect repertoire overlapping over the same elements, because otherwise the difference a mechanism makes would be counted multiple times. A maximally irreducible cause–effect structure composed of concepts is called a *conceptual structure*. The system of mechanisms that specifies a conceptual structure is called a *complex*.⁷ It is useful to think of a conceptual structure as existing as a form in cause–effect space, whose axes are given by all possible past and future states of the complex. In this space, every concept is a point (star), whose size is given by its irreducibility $\phi^{\max}$, and a conceptual structure is a 'constellation' of points, that is, a *form*. Finally, this postulate also applies to spatio-temporal grain. For example, a mechanism cannot have effects at a fine temporal grain, and additional effects at a coarser grain, otherwise causal exclusion would be violated. On the other hand, if the effects at a coarser grain are more irreducible than those at a finer grain, then the coarser grain of causation excludes the finer one [[79](#)].⁸

(c). The central identity: experience as a conceptual structure

Altogether, the elements of a complex in a state, composed into higher order mechanisms that specify concepts, form a *conceptual structure* that is *maximally*

irreducible intrinsically, also known as a *quale*. The constellation of all concepts specifies the overall *form* or shape of the quale ([figure 4](#)).

Figure 4.



[Open in a new tab](#)

A didactic example of how to calculate the quality and quantity of consciousness given a system of elements in a state. On the upper left are three gates with binary states (either ON or OFF: ABC = 100; see also [figure 3](#)) that are wired together as shown. An analysis based on the postulates of IIT [80] reveals that the system forms a complex. The complex in its present state specifies a quale—a conceptual structure that is

maximally irreducible intrinsically. The quale is presented both as the set of maximally irreducible cause–effect repertoires (concepts) specified by each mechanism (top) and as a two-dimensional projection in which each concept is a ‘star’ in cause–effect space (bottom). Cause–effect space or qualia space is a high-dimensional (here, 2×8 dimensions) space in which each axis is a possible past (in blue) and future (in green) state of the complex, and the position along the axis is the probability of that state. Each concept is a star whose position indicates how a mechanism composed of a subset of elements affects the probability of past and future states of the system (its cause–effect repertoire, which specifies what the concept contributes to experience) and whose size ($\phi^{\max}$) measures how irreducible the concept is (how much it contributes to experience). In IIT, $\phi^{\max}$ —a non-negative number—measures the intrinsic irreducibility of the entire quale, how much consciousness there is—the quantity of experience. The ‘form’ or shape of the quale (constellation of stars) is identical to the quality of the experience. Different shapes correspond to different experiences: they feel the way they do—red feeling different from blue or from a headache—because of the distinct shapes of their qualia.

On this basis, the central identity of IIT can be formulated quite simply: *an experience is identical to a conceptual structure that is maximally irreducible intrinsically*. More precisely, a conceptual structure completely specifies both the quantity and the quality of experience: *how much* the system exists—the quantity or level of consciousness—is measured by its $\phi^{\max}$ value—the intrinsic irreducibility of the conceptual structure; *which way* it exists—the quality or content of consciousness—is specified by the shape of the conceptual structure. If a system has $\phi^{\max} = 0$, meaning that its cause–effect power is completely reducible to that of its parts, it cannot lay claim to existing. If $\phi^{\max} > 0$, the system cannot be reduced to its parts, so it exists in and of itself. More generally, the larger $\phi^{\max}$, the more a system can lay claim to existing in a fuller sense than systems with lower $\phi^{\max}$. According to IIT, the quantity and quality of an experience are an intrinsic, fundamental property of a complex of mechanisms in a state—the property of informing or shaping the space of possibilities (past and future states) in a particular way, just as it is considered to be intrinsic to a mass to bend space–time around it.⁹

At any given time, then, consciousness is supported by a set of neuronal elements forming a complex of high $\phi^{\max}$ that specifies a conceptual structure that is maximally irreducible intrinsically. The particular set of neurons that form the major complex, the one of highest $\phi^{\max}$ in the brain, may change to some extent from moment to moment, as well as their state—which neurons are firing and which are not. For example, let us assume that while I watch a scene of a movie containing the actress Jennifer Aniston (JA), the major complex in my brain is made up of neurons within certain parts of the cerebral cortex.¹⁰ Every neuron within the complex necessarily shapes the probability of possible past states (causes) and future states (effects) of the complex, depending on how it is connected to the other neurons and on its state (say firing strongly for 100 ms). Thus, a neuron firing strongly in a certain visual area may specify as more likely those past states of the complex that are compatible with the invariant concept ‘J.A.’s face’, as well as certain appropriate future states. Another neuron firing strongly in another visual

area may specify that there likely was a horizontal edge in a certain position of the visual field, and so on. Yet other neurons that are part of the complex but are silent may specify that certain past (and future) states are unlikely to have occurred (or to occur), such as those having to do with the invariant concepts ‘book’, ‘square’ and so on. Moreover, combinations of neurons may specify higher order concepts, such as ‘J.A. with a red hat sitting on the couch on the left’. Note that all the concepts are specified by elements of the complex, specify cause–effect repertoires over elements of the complex, and acquire meaning intrinsically, in relation to the other concepts in the quale, and not by referring to external inputs (J.A. is just as meaningful when daydreaming about her, or in a dream) [80].

In principle, then, the postulates of IIT offer a way to analyse any system of mechanisms in a particular state and determine whether it constitutes a complex, over which spatial and temporal grain,¹¹ and which quale it specifies. Furthermore, while in practice it is not possible to determine the quale and $\Phi^{\max}$ precisely for a realistic system, it is already possible to employ IIT for prediction, explanation and extrapolation.

(d). Predictions

A straightforward experimental prediction of IIT is that the loss and recovery of consciousness should be associated with the breakdown and recovery of the brain's capacity for information integration. This prediction has been confirmed using transcranial magnetic stimulation (TMS) in combination with high-density EEG in conditions characterized by loss of consciousness [95,96]. These include deep sleep, general anaesthesia obtained with several different agents and brain-damaged patients (vegetative, minimally conscious, emerging from minimal consciousness, locked-in). If a subject is conscious when the cerebral cortex is probed with a pulse of current induced by the TMS coil from outside the skull, the cortex responds with a complex pattern of reverberating activations and deactivations that is both widespread (integrated) and differentiated in time and space (information rich) [95]. By contrast, when consciousness fades, the response of the cortex becomes local (loss of integration) or global but stereotypical (loss of information). The *perturbational complexity index* (PCI), a scalar measure of the compressibility of the EEG response to TMS inspired by IIT, decreases distinctly in all the different conditions of loss of consciousness and, critical for a clinically useful device, is high instead in each conscious healthy subject or neurological patient tested so far [96].

A theory is the more powerful the more it makes correct predictions that violate prior expectations. One counterintuitive prediction of IIT is that a system such as the cerebral cortex may generate experience even if the majority of its pyramidal neurons are nearly silent, a state that is perhaps approximated through certain meditative practices that aim at reaching ‘naked’ awareness without content [97,98]. This corollary of IIT contrasts with the common assumption that neurons only contribute to consciousness if they are active in such a way that they ‘signal’ or ‘broadcast’ the information they represent and ‘ignite’ fronto-parietal networks [3]. That silent neurons can contribute to consciousness

is because, in IIT, information is not in the message that is broadcast by an element, but in the form of the conceptual structure that is specified by a complex. Inactive elements of a complex specify a cause–effect repertoire (the probability of possible past and future states) just as much as active ones (think of the dog that did not bark in the famous Sherlock Holmes story). Conversely, if the same neurons were not merely inactive, but pharmacologically or optogenetically inactivated, they would cease to contribute to consciousness: even though their actual state is the same, they would not specify a cause–effect repertoire, since they do not affect the probability of possible past and future states of the complex.¹²

Another counterintuitive prediction of IIT is that if the efficacy of the 200 million callosal fibres through which the two cerebral hemispheres communicate with each other were reduced progressively, there would be a moment at which, for a minimal change in the traffic of neural impulses across the callosum, there would be an all-or-none change in consciousness: experience would go from being a single one to suddenly splitting into two separate experiencing minds (one linguistically dominant), as we know to be the case with split-brain patients [101,102]. This would be the point at which $\Phi^{\max}$ for the whole brain would fall below the value of $\Phi^{\max}$ for the left and for the right hemisphere taken by themselves.

More generally, IIT predicts that, whatever the neural correlate of consciousness (NCC) turns out to be—whether it is global or local within the cortex, anterior or posterior, medial or lateral, whether it includes primary areas or not, the thalamus or not, whether it encompasses neurons in supragranular, infragranular layers of cortex or not—it should be a local maximum of Φ , and thus of a maximum of intrinsic, irreducible cause–effect power. IIT also predicts that the NCC is not necessarily fixed, but may expand, shrink and even move within a given brain depending on various conditions. In fact, there may even be multiple NCCs in a single brain, as shown by split-brain patients, in which case there should be multiple local maxima of integrated information. Finally, IIT makes precise predictions about the physical elements that constitute the NCC and the time intervals and levels of activity at which they operate [77,79]: they should have a spatial scale that achieves the highest value of Φ , as opposed to finer or coarser grains (say, either individual neurons or local groups of neurons rather than neuronal compartments or brain areas); they should operate most effectively (highest value of Φ) at the time scale of consciousness, as opposed to finer or coarser scales (say, hundred milliseconds rather than a millisecond or ten seconds); and the activity states that make the most difference to the NCC should be the ones that support phenomenological distinctions (say, bursting, high mean firing, low mean firing). In short, the general rule is that the NCC must always correspond to a maximum of intrinsic, irreducible cause–effect power.

(e). Explanations

IIT offers a coherent, principled account of the NCC—which it identifies with the major complex in a particular state—and of many disparate empirical observations. For

example, why is consciousness generated by the cerebral cortex (or at least some parts of it), but the cerebellum does not contribute to it, despite the latter having even more neurons; [103]? Why does consciousness fade early in sleep, although the brain remains active? Why is it lost during generalized seizures, when neural activity is intense and synchronous? Why is there no direct contribution to consciousness from neural activity within sensory pathways (the retina) and motor pathways (the motoneurons in the spinal cord), or within neural circuits looping out of the cortex into subcortical structures and back, despite their manifest ability to influence the content of experience?

These and other well-known facts find a parsimonious explanation based on the postulates of IIT. Thus, a prominent feature of the cerebral cortex, which is responsible for the content of consciousness, is that it is composed of elements that are functionally specialized and at the same time can interact rapidly and effectively. This is the kind of organization that yields a comparatively high value of $\Phi^{\max}$. Instead, the cerebellum is composed of small modules that process inputs and produce outputs largely independent of each other [104,105]. Simulations also show that input and output pathways, while capable of affecting the major complex and being affected by it, can remain excluded from it, because they are not part of a local maximum of integrated information. The same applies to loops that may exit the major complex and reenter it. Other simulations show that $\Phi^{\max}$ is low when the effective connectivity among a set of elements is weak or is organized in homogeneous manner. Indeed, as was mentioned above, when consciousness fades during deep slow wave sleep or in certain states of general anaesthesia, the interactions among different cortical regions become weaker or highly stereotypical, as they do during generalized epileptic seizures.

(f). Extrapolations

Finally, the more the postulates of IIT are validated in situations in which we are reasonably confident about whether and how consciousness changes, the more we can use the theory to extrapolate and make inferences about situations where we are less confident—brain-damaged patients, newborn babies, alien animals, complicated machines and other far-fetched scenarios, as we shall consider next.

5. Everywhere?

In the ‘Canticle of the Creatures’, Saint Francis addressed animals, flowers and even stones as if endowed with soul, and praised them as mother earth, brother sun, sister moon, the stars, the air, water and fire. And he was not alone. Some of the brightest minds in the West embraced some form of the ancient philosophical doctrine of panpsychism, starting with the Presocratics and Plato. The Renaissance philosophers Patrizi, Bruno, Telesio and Campanella took the position that matter and soul are one substance. Later, Spinoza, Leibniz, Schopenhauer and, closer to modern times, James, Whitehead, Russell, and Teilhard de Chardin espoused panpsychist notions [106,107]. Strawson [108,109] is a well-known contemporary defender of panpsychism. Eastern

traditions, such as Buddhism, have always emphasized the continuity of consciousness across life.

Materialism, or its modern offspring, physicalism, has profited immensely from Galileo's pragmatic stance of removing subjectivity (mind) from nature in order to describe and understand it objectively—from the extrinsic perspective of a manipulator/observer. But it has done so at the cost of ignoring the central aspect of reality from the intrinsic perspective—experience itself. Unlike idealism, which does away with the physical world, or dualism, which accepts both in an uneasy marriage, panpsychism is elegantly unitary: there is only one substance, all the way up from the smallest entities to human consciousness and maybe to the World Soul (*anima mundi*). But panpsychism's beauty has been singularly barren. Besides claiming that matter and mind are one thing, it has little constructive to say and offers no positive laws explaining how the mind is organized and works.

IIT was not developed with panpsychism in mind (*sic*). However, in line with the central intuitions of panpsychism, IIT treats consciousness as an intrinsic, fundamental property of reality. IIT also implies that consciousness is graded, that it is likely widespread among animals, and that it can be found in small amounts even in certain simple systems. Unlike panpsychism, however, IIT clearly implies that not everything is conscious. Moreover, IIT offers a solution to several of the conceptual obstacles that panpsychists never properly resolved, like the problem of aggregates (or combination problem [[107](#),[110](#)]) and can account for its quality. It also explains why consciousness can be adaptive, suggesting a reason for its evolution.

(a). Consciousness is a fundamental property

The axioms and postulates of IIT say that consciousness is a fundamental, observer-independent property that can be accounted for by the intrinsic cause–effect power of certain mechanisms in a state—how they give form to the space of possibilities in their past and their future. An analogy is mass, which can be defined by how it curves space–time around it—except that in the case of experience the entities having the property are not elementary particles but complexes of elements, and experience comes not in two but in a trillion varieties. In this general sense, at least, IIT is not at odds with panpsychism.

(b). Consciousness comes in various qualities

Unfortunately, panpsychism is mute when it comes to explaining the way any one conscious experience feels—why the perception of red feels different from that of blue and why colours are experienced as different from tones. Instead, at least in principle, IIT says exactly what determines the quality of an experience—what makes it the particular way it is: an experience is a maximally irreducible conceptual structure or quale—a shape in a fantastically high-dimensional cause–effect space specified by a complex of neurons in a particular state. This is the constellation of concepts through

which the neurons of the major complex, in various combinations, give form to the space of its possible past and future states ([figure 4](#)). Different experiences—every different scene in a movie or in a dream—correspond to different shapes, with some shapes being measurably closer (red and blue) and some more distant within the space (a black screen and a city scene). Indeed, there is much scope for future research to begin mapping psychophysics, for example, the circular nature of colour space, onto the geometry of shapes in cause–effect space—except that a shape in cause–effect space, unlike the shape of an object in 3D space, is the shape within, the shape of experience itself. It is the voice in the head, the light inside the skull.

(c). Consciousness is adaptive

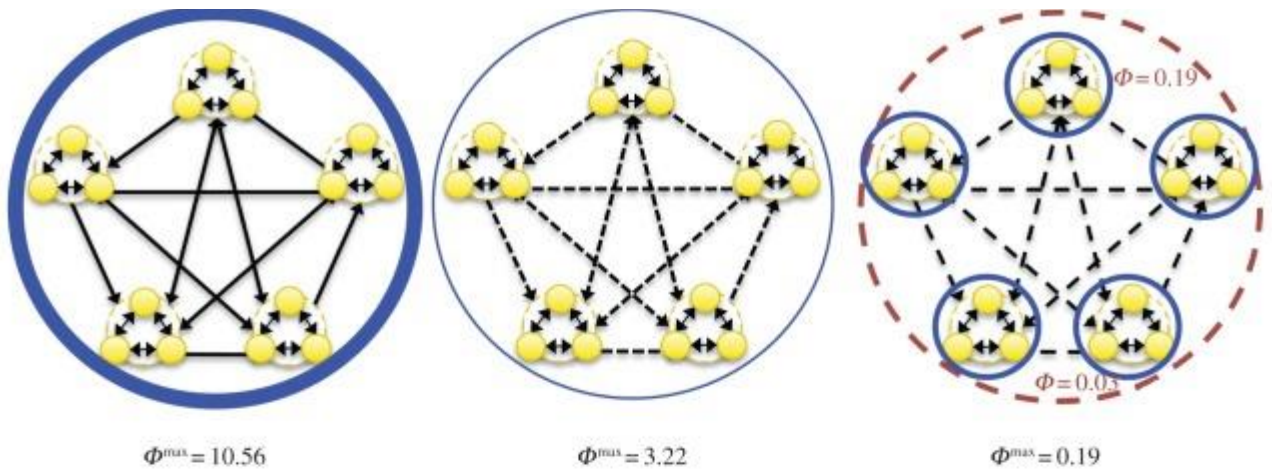
IIT takes no position on the function of experience as such—similar to physics not having anything to say about the function of mass or charge. However, by identifying consciousness with integrated information, IIT can account for why it evolved, another aspect about which panpsychism has nothing to say. In general, a brain having a high capacity for information integration will better match an environment with a complex causal structure varying across multiple time scales, than a network made of many modules that are informationally encapsulated. Indeed, artificial life simulations ('animats') of simple Braitenberg-like vehicles that have to traverse mazes and whose brains evolve by natural selection over 60 000 generations show a monotonic relationship between (simulated) integrated information and adaptation [[111,112](#)]. That is, the more adapted individual animats are to their environment, the higher the integrated information of the major complex in their brain. Similar animats, evolved to catch falling blocks in a *Tetris*-like scenario, demonstrate that increased adaptation leads to increased number of concepts in the major complex and an associated increase in integrated information that depends on the complexity of the animats' environment [[113](#)]. Thus, evolution by natural selection gives rise to organisms with high $\Phi^{\max}$ because, given constraints on the number of elements and connections, they can pack more functions per element than their less integrated competitors and thus are more adept at exploiting regularities in a rich environment.

(d). Consciousness is graded

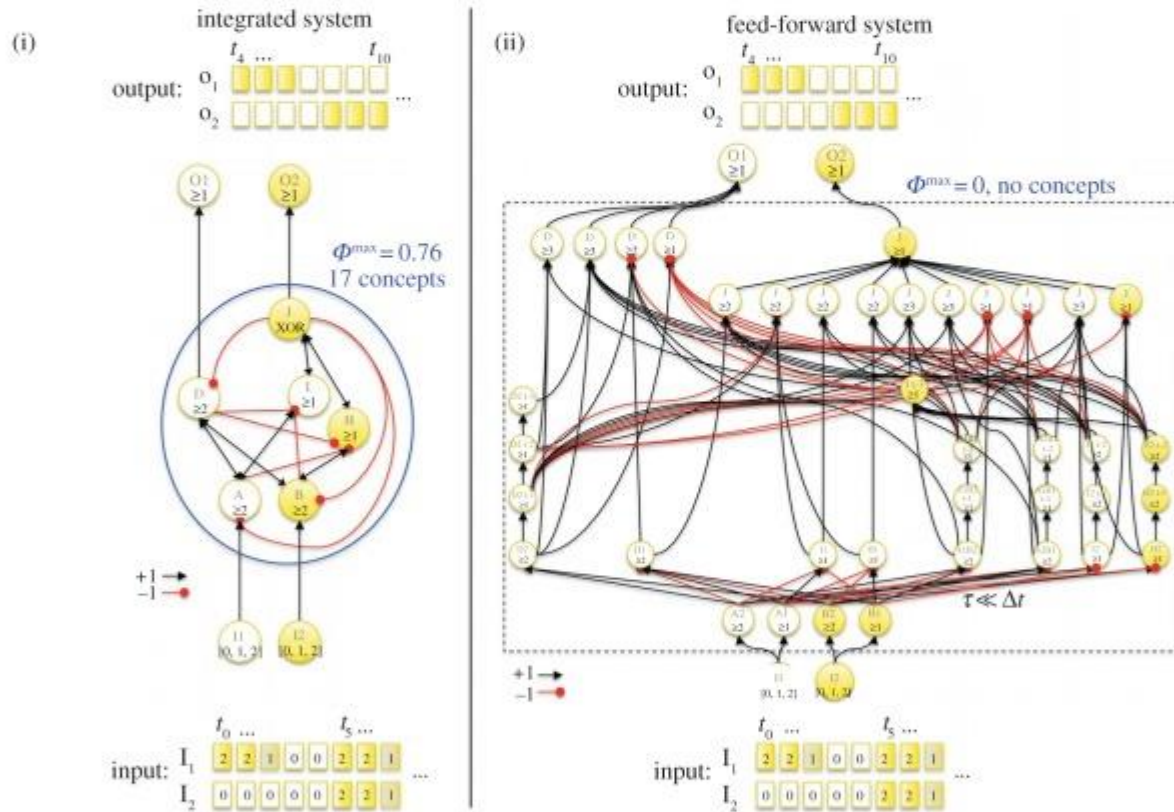
IIT does side with the panpsychist intuition that consciousness may be present across the animal kingdom, and even beyond, but in varying degrees. Everything else being equal, integrated information, and with it the richness of experience, is likely to increase as the number of neurons and the abundance of their interconnections grow, although sheer number of neurons is not a guarantee, as shown by the cerebellum. It is also likely that consciousness is graded across the lifetime of any one organism. In us it becomes richer as we grow from a baby to an adult whose brain has fully matured and becomes more functionally specialized. It can also wax and wane when we are highly alert or drowsy, intoxicated by drugs or alcohol, or become demented in old age. This is illustrated schematically in [figure 5a](#), where a set of 'cortical' areas is integrated into a major complex of 'high' $\Phi^{\max}$ when the inter-areal connections are strong, undergoes a reduction in $\Phi^{\max}$ when connection strength is reduced by neuromodulatory changes

(simulated as an increase in noise), and finally breaks down into small complexes of low $\phi^{\max}$.

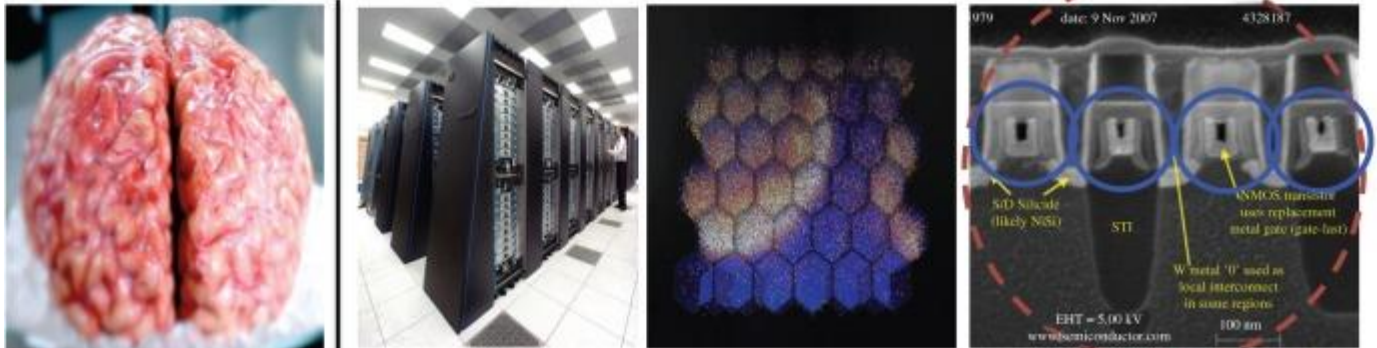
Figure 5.



(b)



(c)



[Open in a new tab](#)

IIT makes several predictions about which systems can experience anything—how much and in which way—and which systems, even complicated ones, have no experience, remaining ‘in the dark’. IIT implies that consciousness is graded (*a*); that aggregates are not conscious (*a*, right panel); that strictly feed-forward systems are not conscious (*b*, right panel), even if they are functionally equivalent in terms of their input–output operations to feedback networks that are conscious (*b*, left panel); that even accurate biophysical simulations of the human brain running on digital machines would not be conscious like us, but would be mere aggregates of much simpler systems (transistors and the like) having minimal $\Phi^{\max}$ (*c*). The last row (*c*) shows, from left to right, a human brain (Allen Institute), the IBM Blue Gene P supercomputer, a columnar model of mouse cortex (Blue Brain Project) and a scanning electron micrographic cross-section of 4 NMOS INTEL transistors in a grid.

A corollary of IIT that violates common intuitions is that even circuits as simple as a ‘photodiode’ made up of a sensor and a memory element can have a modicum of experience [80] (see also [figure 5a](#), right panel). It is nearly impossible to imagine what it would ‘feel like’ to be such a circuit, for which the only phenomenal distinction would be between ‘this rather than not this’ (unlike a photodiode, when we are conscious of ‘light’ or of ‘dark,’ our experience is what it is because it includes scores of negative concepts, such as no colours, no shapes, no thoughts and so on, that are all available to us). But consider that normal matter at -272.15°C , one degree above absolute zero, still contains some heat. However, in practice its temperature is as cold as it gets. Similarly, there may well be a practical threshold for $\Phi^{\max}$ below which people do not report feeling much of anything, but this does not mean that consciousness has reached its absolute minimum, zero. Indeed, when we fall into a deep, dreamless sleep and don’t report any experience upon being awoken, some small complex of neurons within our sleeping brain will likely have a $\Phi^{\max}$ value greater than zero, yet that may not amount to much compared to that of our rich, everyday experience.

(e). Multiple consciousnesses

IIT also allows for the possibility of two or more complexes coexisting within a single system [80]. Depending on the exact connectivity, these are likely to have quite different values of $\Phi^{\max}$. Indeed, in the brains of both vertebrates and invertebrates, there may well exist, at least under some conditions, a major complex and one or more minor complexes. In humans, the complex that supports our day-to-day stream of conscious experience should have by far the highest value of integrated information—it should be the *major* complex. In split-brain patients the speaking, major complex is unaware of the presence of another consciousness, one that typically lacks speech, but which can be revealed by clever experimental paradigms [102,114]. It is conceivable that at least some cases of ‘high-level’ performance found in normal subjects [64,115]), while unconscious from the perspective of the major complex, may be due to the presence of minor complexes (of course, some of these behaviours may be mediated by purely feed-forward circuits). This counterintuitive scenario of ‘many conscious minds within a single brain’ could be assessed, at least in principle, by measurements of integrated

information at the neuronal level. Major and minor complexes may also occur in patients with Marchiafava–Bignami disease [116] and other disconnection syndromes, in patients with identity and conversion disorders [63], and in other neurological and psychiatric conditions.

(f). Aggregates are not conscious

‘Take a sentence of a dozen words, and take twelve men and tell to each one word. Then stand the men in a row or jam them in a bunch, and let each think of his word as intently as he will; nowhere will there be a consciousness of the whole sentence’. This is how William James illustrated the combination problem of panpsychism [110]. Or take John Searle: ‘Consciousness cannot spread over the universe like a thin veneer of jam; there has to be a point where my consciousness ends and yours begins’ [117]. Indeed, if consciousness is everywhere, why should it not animate the United States of America? IIT deals squarely with this problem by stating that only maxima of integrated information exist. Consider two people talking: within each brain, there will be a major complex—a set of neurons that form a maximally irreducible cause–effect structure with definite borders and a high value of $\Phi^{\max}$. Now let the two speak together. They will now form a system that is also irreducible ($\Phi > \text{zero}$) due to their interactions. However, it is not maximally irreducible, since its value of integrated information will be much less than that of each of the two major complexes it contains. According to IIT, there should indeed be two separate experiences, but no superordinate conscious entity that is the union of the two. In other words, there is nothing-it-is-like-to-be two people, let alone the 300 plus million citizens making up the USA.¹³ Again, this point can be exemplified schematically by the system of [figure 5a](#), right panel. While the five small complexes do interact, forming a larger integrated system, the larger system is not a complex: by the exclusion postulate, only the five smaller complexes exist, since they are local maxima of integrated information ($\Phi^{\max} = 0.19$), while the larger system is not a complex ($\Phi = 0.03$). Worse, a dumb thing with hardly any intrinsically distinguishable states, say a grain of sand for the sake of the argument, has no experience whatsoever. And heaping a large number of such zero- Φ systems on top of each other would not increase their Φ to a non-zero value: to be a sand dune does not feel like anything either—aggregates have no consciousness.

(g). Complicated systems can be unconscious

A second class of zero- Φ systems are purely feed-forward computational networks in which one layer feeds the next one without any recurrent connections. In a feed-forward network, the input layer is always determined entirely by external inputs and the output layer does not affect the rest of the system, hence neither layer can be part of a complex, and the same is true recursively for the next layers downstream and upstream. According to IIT, then, a feed-forward network does not exist intrinsically—for itself—but is a zombie—carrying out tasks unconsciously [118]. Yet from the extrinsic perspective of a user, feed-forward networks, like those used in deep learning, perform plenty of useful computational functions, such as finding faces or cats in images [119], labelling images, reading zip codes and detecting credit card fraud.

This has a rather startling consequence. Consider that any neural network with feedback circuits can be mapped onto a purely feed-forward network in such a manner that the latter approximates its input–output relationships (for computations bounded by a maximal time step [120]). That is, for the same inputs, the two networks will yield the same output (in general, the equivalent feed-forward network will have many more nodes and connection than the feedback network). Therefore, a purely feed-forward system able to replicate the input–output behaviour of the human brain (under the limited time-step constraint) would be behaviourally indistinguishable from us, and certainly capable of passing the Turing test, yet it would have zero Φ and would thus be a ‘perfect’ zombie. A simple example of two functionally equivalent systems, one with recurrent connections and non-zero Φ , and one purely feed-forward with zero Φ , is shown in [figure 5b](#) [80].

In people and organisms that evolved through natural selection, input–output behaviour provides a good first guess about the presence of consciousness. However, as demonstrated by the example in [figure 5b](#), this may not always be the case for radically different computational architectures. In the general case, and certainly with machines, it becomes essential to consider the internal circuitry—not just what the machine does, but how it does it. This also means that there cannot be an ultimate Turing test for consciousness (although, there may be some practical CAPTCHA-like tests [121]). According to many functionalist notions [122], if a machine reproduces our input–output behaviour in every circumstance, it would have to be granted consciousness just as much as us. IIT could not disagree more—no Turing test (e.g. Samantha in the Hollywood movie *She*) can be a sufficient criterion for consciousness, human or otherwise.

(h). Simulations of conscious neural systems can be unconscious

Finally, what about a computer whose software simulates in detail not just our behaviour, but even the biophysics of neurons, synapses and so on, of the relevant portion of the human brain [123]? Could such a digital simulacrum ever be conscious? Functionalism again would say yes, even more forcefully. For in this case *all* the relevant functional roles within our brain, not just our input–output behaviour, would have been replicated faithfully. Why should we not grant to this simulacrum the same consciousness we grant to a fellow human? According to IIT, however, this would not be justified, for the simple reason that the brain is real, but a simulation of a brain is virtual. For IIT, consciousness is a fundamental property of certain physical systems, one that requires having *real* cause–effect power, specifically the power of shaping the space of possible past and future states in a way that is maximally irreducible intrinsically. In the same way, mass is an intrinsic property of systems of particles, a property that has real causal power, specifically that of bending space–time. Therefore, just like a computer simulation of a giant star will not bend space–time around the machine, a simulation of our conscious brain will not have consciousness.¹⁴ Of course, the physical computer that is running the simulation is just as real as the brain. However, according to the principles of IIT, one should analyse its real physical components—identify elements, say transistors, define their cause–effect repertoires,

find concepts, complexes and determine the spatio-temporal scale at which Φ reaches a maximum. In that case, we suspect that the computer would likely *not* form a large complex of high $\Phi^{\max}$, but break down into many mini-complexes of low $\Phi^{\max}$. This is due to the small fan-in and fan-out of digital circuitry ([figure 5c](#)), which is likely to yield maximum cause–effect power at the fast temporal scale of the computer clock.^{[15](#)}

6. Conclusion

In summary, there are some aspects of IIT that definitely do not fit with panpsychism, and others that vindicate some of its intuitions. In this respect, it is natural to consider how one should regard some of the inferences derived from IIT for which it is hard even to imagine a direct test at the present time. Our position is that, as is often the case in science,^{[16](#)} a theory is first tested and validated in situations that are close to ideal, and then extrapolated to more remote cases. Ideally, whether consciousness varies with integrated information, and other predictions of IIT, would first be validated *here*—on my own consciousness: for example, does $\Phi^{\max}$ collapse when I undergo general anaesthesia or a seizure, or when I fall into dreamless sleep, and return to high values when I dream? Does my experience change if one temporarily inactivates a region of my brain that is part of the major complex, but not one that is outside it? Does it change if one succeeds in connecting a neuromorphic microcircuit that becomes part of my major complex and not otherwise? Then one can extrapolate to *there*, at first in situations involving other healthy humans, then in slightly more difficult cases, say monkeys with a brain similar to ours who are trained to give reports similar to ours. Finally, insofar as the theory has been validated and has shown good predictive and explanatory power, one can try and extrapolate to *everywhere*, unresponsive patients with just a small ‘island’ of functioning brain tissue, newborn babies, animals very different from us, photodiodes, machines, and computer simulations. After all, often in science the most we can do is to draw our best inferences about unknown instances based on a theory that works well in many known instances. And that is much better than to make arbitrary claims or to draw no inference whatsoever.

Acknowledgements

We thank Larissa Albantakis, Melanie Boly, Chiara Cirelli, Lice Ghilardi and Marcello Massimini for their many contributions to the work presented here.

Endnotes

1

Note that we consider reflective–consciousness, highly developed in adult humans, to be a subclass of conscious experiences. Likewise, the feeling of freely willing an action—such as raising one’s arm—sometimes also referred to as *agency* [[6,7](#)—is another subclass of conscious experiences. While their content differs from the content associated with feeling pain or seeing red, subjectivity is common to all.

Consciousness can be dissociated from many other cognitive processes that have traditionally been closely linked to it, including memory, emotions and selective attention (for reviews see [19,20]). It can persist if the recall of long-term memories is impaired, it can be present in patients who lack affect, and it can be dissociated from attention. The last point is particularly counterintuitive but is well supported—subjects can attend to invisible objects [21]. The extent to which it is possible to become conscious of something without also attending to it is more controversial [21,22].

3

Not to mention the question of whether it feels-like-something to be a Venus flytrap or a single-cell organism.

4

If it is not outright wrong, IIT most likely will have to be refined, expanded and adjusted. However, in its current form (IIT 3.0), it explains and predicts a wide range of phenomena, including a number of counterintuitive predictions amenable to empirical falsification. For the latest formulation of the theory, see [80]; for earlier versions, see [76,77,81,82]; for a literary account, see [77,83]. The main differences between IIT 3.0 and earlier versions are listed in the appendix of [80].

5

For instance, the unified nature of conscious experiences has been questioned by psychophysical experiments demonstrating temporal asynchrony [85,86]. See also [87].

6

For example, the notion of the aether was introduced in the late nineteenth century to explain the propagation of light. When more and more experiments concluded that, whatever the aether might be, it had no effects whatsoever, it finally fell under Occam's razor, and it plays no role in modern physics.

7

Importantly, this may be a macro- rather than a micro-spatio-temporal scale [79]. For example, the relevant level for human consciousness is likely to be

neurons at the scale of 100 ms, rather than molecules at the nanosecond scale. Note that it is possible for a single physical system, such as the brain, to contain two or more independent yet causally interacting complexes, each with their own $\Phi^{\max}$ (see section on multiple consciousnesses). Indeed, it is even possible for a physical system to contain complexes at different spatio-temporal grains, such as a mitochondrion forming a complex inside a neuron, as long as there is no causal overlap at the relevant scales.

8

Requiring that only the maximum of Φ over elements, spatial and temporal grain must be considered is not exceptional in science: many of the laws of physics are formulated as extremum principles, e.g. the principle of least action.

9

IIT postulates that experience is a fundamental, intrinsic property of the world. Different experiences must be specified by different physical substrates, although different physical substrates may specify the same experience (for example, by differing in micro-properties that are causally irrelevant at the macro-scale that achieves a maximum of cause-effect power, see also *metamers*) [89]. Note that IIT is compatible with quantum mechanics. In principle, Φ and related quantities can be assessed also in quantum system, although it has been suggested that at the quantum level Φ values may be very small [90].

10

Here we do not elaborate about particular cortical areas, cortical layers or particular population of neurons.

11

The exclusion postulate requires that the set of mechanisms that specify one particular experience do so over the time window at which Φ reaches a maximum. If the next experience involves an overlapping set of mechanisms, it would seem that, to avoid multiple causation, it should be specified over a non-overlapping time window. Accordingly, the seemingly continuous 'stream' of consciousness would actually be constituted by a discrete succession of 'snapshots', in line with some psychophysical evidence [91–94]. Note that each snapshot has motion and other dynamic percepts associated with it.

12

It is instructive to consider ‘the perfect experiment’ hypothesized by Cohen & Dennett [99]: a subject looks at a red apple, and neurons in his cortical colour region (say V4) fire. However, imagine one could selectively block their projections to further cortical regions, so that the subject cannot access and report the colour of the apple. According to Cohen and Dennett, any theory claiming that ‘phenomenal’ consciousness can be dissociated from cognitive access would have to claim that, as long as the colour neurons are active, the subject would be conscious of a red apple, while at the same time he would be denying that it is red. According to IIT, however, blocking the projections of V4 neurons destroys their cause–effect repertoires no matter whether they are firing or not, leading to the collapse of the ‘colour section’ (Q-fold) of the conceptual structure that corresponds to the subject's experience. As a consequence, the subject would not be conscious of colours (*cerebral achromatopsia*) and would not even understand what he has lost (*colour anosognosia*), just like a patient described by von Arx *et al.* [100].

13

By the same token, the exclusion postulate predicts a scenario that is the mirror image of the prediction that consciousness will suddenly split in two when the corpus callosum is ‘cooled’ below a critical point: if two people speaking were to increase their effective causal interactions by some, yet to be invented, direct brain-to-brain connectivity booster, to the point where the $\Phi^{\max}$ of the two interacting brains would exceed $\Phi^{\max}$ of the individual brains, their individual conscious mind would disappear and its place would be taken by a new Über-mind that subsumes both.

14

A similar point was made by John Searle with his Chinese Room Argument [124] and by Leibniz 300 years earlier with his mill [125].

15

Ultimately, any digital computer running software can be mimicked by a Turing Machine with a large state-transition matrix, a moving head that writes and erases, and a very, very long memory tape—in that case, cause–effect power would reside in the moving head that follows one out of a few instructions at a time. On the other hand, there is no reason why hardware-level, neuromorphic models of the human brain that do not rely on software running on a digital computer, could not approximate, one day, our level of consciousness [126].

A related question has to do with the Internet and whether it could be conscious [127]. One way to think about this is to assume that each computer connected to the Internet is an element having real cause–effect power at a macro-level (by ‘black-boxing’ its internal mechanisms). For example, each computer could send an ON signal when it is ON and an OFF signal when it is OFF. One could then make sure that each computer increased or decreased the likelihood of being ON depending on how many ON signals it receives. In principle, this kind of organization could be arranged so that it gives rise to a complex of high Φ , although this is certainly not the way the Internet works right now. On the other hand, if one considers the micro-elements inside each computer (say its transistors) as having real cause–effect power, we are back to the situation in which they most likely would not form any large complex within each computer, let alone across connected computers.

16

A well-known instance of such an extrapolation is the inference of singularities in space–time due to the extreme mass of a stellar object. Such black holes were pure conjectures, based on a solution of Einstein's theory of General Relativity, until they were subsequently confirmed observationally.

References

- 1.Koch C. 2004. The quest for consciousness: a neurobiological approach. Englewood, CO: Roberts and Co. [[Google Scholar](#)]
- 2.Baars B, Gage NM. 2010. Cognition, brain, and consciousness. New York, NY: Academic Press. [[Google Scholar](#)]
- 3.Dehaene S, Changeux J-P. 2011. Experimental and theoretical approaches to conscious processing. *Neuron* 70, 200–227. (10.1016/j.neuron.2011.03.018) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 4.Boly M, Baars B, Seth AK, Wilke M, Ingmundson P, Laureys S, Edelman D, Tsuchiya N. 2013. Consciousness in humans and non-human animals: recent advances and future directions. *Front. Psychol.* 4, 625 (10.3389/fpsyg.2013.00625) [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 5.Laureys S, Tononi G. 2009. The neurology of consciousness: cognitive neuroscience and neuropathology. Amsterdam, The Netherlands: Academic Press. [[Google Scholar](#)]
- 6.Fried I, Katz A, McCarthy G, Sass KJ, Williamson P, Spencer SS, Spencer DD. 1991. Functional organization of human supplementary motor cortex

studied by electrical stimulation. *J. Neurosci.* 11, 3656–3666. [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]

- 7. Wegner DM. 2002. The illusion of conscious will. Cambridge, MA: MIT Press. [[Google Scholar](#)]
- 8. Lancaster MA, et al. 2013. Cerebral organoids model human brain development and microcephaly. *Nature* 501, 373–379. (10.1038/nature12517) [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 9. Tononi G, Srinivasan R, Russell DP, Edelman GM. 1998. Investigating neural correlates of conscious perception by frequency-tagged neuromagnetic responses. *Proc. Natl Acad. Sci. USA* 95, 3198–3203. [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 10. Cheesman J, Merikle PM. 1984. Priming with and without awareness. *Percept. Psychophys.* 36, 387–395. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 11. Ramsoy TZ, Overgaard M. 2004. Introspection and subliminal perception. *Phenomenology Cogn. Sci.* 3, 1–23. [[Google Scholar](#)]
- 12. Sandberg K, Timmermans B, Overgaard M, Cleeremans A. 2010. Measuring consciousness: is one measure better than the other? *Conscious Cogn.* 19, 1069–1078. (10.1016/j.concog.2009.12.013) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 13. Persaud N, McLeod P, Cowey A. 2007. Post-decision wagering objectively measures awareness. *Nat. Neurosci.* 10, 257–261. (10.1038/nn1840) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 14. Cowey A, Stoerig P. 1995. Blindsight in monkeys. *Nature* 373, 247–249. (10.1038/373247a0) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 15. Kepecs A, Uchida N, Zariwala HA, Mainen ZF. 2008. Neural correlates, computation and behavioral impact of decision confidence. *Nature* 455, 227–231. (10.1038/nature07200) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 16. Kornell N. 2014. Where is the ‘meta’ in animal metacognition. *J. Comp. Psychol.* 128, 143–149. (10.1037/a0033444) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 17. Lamme VAF. 2006. Towards a true neural stance on consciousness. *Trends Cogn. Sci.* 10, 494–501. (10.1016/j.tics.2006.09.001) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 18. Block N. 2007. Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behav. Brain Sci.* 30, 481–499, discussion 499–548. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]

- 19.Koch C, Tsuchiya N. 2012. Attention and consciousness: related yet different. *Trends Cogn. Sci.* 16, 103–105. (10.1016/j.tics.2011.11.012) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 20.Tononi G, Koch C. 2008. The neural correlates of consciousness: an update. *Ann. NY Acad. Sci.* 1124, 239–261. (10.1196/annals.1440.004) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 21.Tsuchiya N, Koch C. 2014. On the relationship between consciousness and attention. In *The cognitive neurosciences* (ed. Gazzaniga MS.). Cambridge, MA: MIT Press. [[Google Scholar](#)]
- 22.Cohen MA, Cavanagh P, Chun MM, Nakayama K. 2012. The attentional requirements of consciousness. *Trends Cogn. Sci.* 16, 411–417. (10.1016/j.tics.2012.06.013) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 23.Koch C, Crick F. 1990. Some reflections on visual awareness. *Cold Spring Harb. Symp. Quant. Biol.* 55, 953–962. (10.1101/SQB.1990.055.01.089) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 24.Crick F, Koch C. 1998. Consciousness and neuroscience. *Cereb. Cortex* 8, 97–107. (10.1093/cercor/8.2.97) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 25.Moruzzi G, Magoun HW. 1949. Brain stem reticular formation and activation of the EEG. *Electroencephalogr. Clin. Neurophysiol.* 1, 455–473. [[PubMed](#)] [[Google Scholar](#)]
- 26.Parvizi J, Damasio AR. 2003. Neuroanatomical correlates of brainstem coma. *Brain* 126, 1524–1536. (10.1093/brain/awg166) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 27.Zeman A. 2001. Consciousness. *Brain* 124, 1263–1289. (10.1093/brain/124.7.1263) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 28.Perry E, Ashton H, Young A. (eds). 2002. *Neurochemistry of Consciousness*. Amsterdam, The Netherlands: John Benjamins. [[Google Scholar](#)]
- 29.Tononi G, Laureys S. 2009. The neurology of consciousness: an overview. In *The neurology of consciousness: cognitive neuroscience and neuropathology* (eds Laureys S, Tononi G.), pp. 375–412. Amsterdam, The Netherlands: Academic Press. [[Google Scholar](#)]
- 30.Baars BJ, Ramsoy TZ, Laureys S. 2003. Brain, conscious experience and the observing self. *Trends Neurosci.* 26, 671–675. (10.1016/j.tins.2003.09.015) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]

- 31. Myerson J, Miezin F, Allman JM. 1981. Binocular rivalry in macaque monkeys and humans: a comparative study in perception. *Behav. Anal. Lett.* 1, 149–159. [[Google Scholar](#)]
- 32. Baars BJ. 1988. A cognitive theory of consciousness. Cambridge, UK: Cambridge University Press. [[Google Scholar](#)]
- 33. Logothetis NK. 1998. Single units and conscious vision. *Phil. Trans. R. Soc. Lond. B* 353, 1801–1818. (10.1098/rstb.1998.0333) [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 34. Mudrik L, Koch C. 2013. Differential processing of invisible congruent and incongruent scenes: a case for unconscious integration. *J. Vis.* 13, 24 (10.1167/13.13.24) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 35. Mataró M, Jurado MA, García-Sánchez C, Barraquer L, Costa-Jussà FR, Junqué C. 2001. Long-term effects of bilateral frontal brain lesion: 60 years after injury with an iron bar. *Arch. Neurol.* 58, 1139–1142. (10.1001/archneur.58.7.1139) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 36. Goldberg II, Harel M, Malach R. 2006. When the brain loses its self: prefrontal inactivation during sensorimotor processing. *Neuron* 50, 329–339. (10.1016/j.neuron.2006.03.015) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 37. Frässle S, Sommer J, Jansen A, Naber M, Einhäuser W. 2014. Binocular rivalry: frontal activity relates to introspection and action but not to perception. *J. Neurosci.* 34, 1738–1747. (10.1523/JNEUROSCI.4403-13.2014) [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 38. Engel AK, Singer W. 2001. Temporal binding and the neural correlates of sensory awareness. *Trends Cogn. Sci. (regul. ed.)* 5, 16–25. (10.1016/S1364-6613(00)01568-0) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 39. Imas OA, Ropella KM, Ward BD, Wood JD, Hudetz AG. 2005. Volatile anesthetics enhance flash-induced gamma oscillations in rat visual cortex. *Anesthesiology* 102, 937–947. (10.1097/00000542-200505000-00012) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 40. Murphy MJ, et al. 2011. Propofol anesthesia and sleep: a high-density EEG study. *Sleep* 34, 283–291. [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 41. Cote KA, Etienne L, Campbell KB. 2001. Neurophysiological evidence for the detection of external stimuli during sleep. *Sleep* 24, 791–803. [[PubMed](#)] [[Google Scholar](#)]

- 42.Takahara M, Nittono H, Hori T. 2002. Comparison of the event-related potentials between tonic and phasic periods of rapid eye movement sleep. *Psychiatry Clin. Neurosci.* 56, 257–258. (10.1046/j.1440-1819.2002.00999.x) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 43.Aru J, Axmacher N, Do Lam ATA, Fell J, Elger CE, Singer W, Melloni L. 2012. Local category-specific gamma band responses in the visual cortex do not reflect conscious perception. *J. Neurosci.* 32, 14 909–14 914. (10.1523/JNEUROSCI.2051-12.2012) [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 44.Pitts MA, Martínez A, Hillyard SA. 2012. Visual processing of contour patterns under conditions of inattention blindness. *J. Cogn. Neurosci.* 24, 287–303. (10.1162/jocn_a_00111) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 45.Chalmers DJ. 2000. What is a neural correlate of consciousness? In *Neural correlates of consciousness: empirical and conceptual questions* (ed. Metzinger T.), pp. 17–40. Cambridge, MA: MIT Press. [[Google Scholar](#)]
- 46.Miller SM. 2007. On the correlation/constitution distinction problem (and other hard problems) in the scientific study of consciousness. *Acta Neuropsychiatrica* 19, 159–176. (10.1111/j.1601-5215.2007.00207.x) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 47.Owen AM, Coleman MR, Boly M, Davis MH, Laureys S, Pickard JD. 2006. Detecting awareness in the vegetative state. *Science* 313, 1402 (10.1126/science.1130197) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 48.King JR, Sitt JD, Faugeras F, Rohaut B, El Karoui I, Cohen L, Naccache L, Dehaene S. 2013. Information sharing in the brain indexes consciousness in noncommunicative patients. *Curr. Biol.* 23, 1914–1919. (10.1016/j.cub.2013.07.075) [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 49.Schiff N, et al. 2002. Residual cerebral activity and behavioural fragments can remain in the persistently vegetative brain. *Brain* 125, 1210–1234. (10.1093/brain/awf131) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 50.Lagercrantz H, Changeux J-P. 2009. The emergence of human consciousness: from fetal to neonatal life. *Pediatr. Res.* 65, 255–260. (10.1203/PDR.0b013e3181973b0d) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 51.Kouider S, Stahlhut C, Gelskov SV, Barbosa LS, Dutat M, de Gardelle V, Christophe A, Dehaene S, Dehaene-Lambertz G. 2013. A neural marker of perceptual consciousness in infants. *Sci. Am. Assoc. Adv. Sci.* 340, 376–380. (10.1126/science.1232509) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]

- 52.Smith JD, Couchman JJ, Beran MJ. 2014. Animal metacognition: a tale of two comparative psychologies. *J. Comp. Psychol.* 128, 115–131. (10.1037/a0033105) [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 53.Hawrylycz MJ, et al. 2012. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* 489, 391–399. (10.1038/nature11405) [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 54.Herculano-Houzel S. 2012. The remarkable, yet not extraordinary, human brain as a scaled-up primate brain and its associated cost. *Proc. Natl Acad. Sci. USA* 109(Suppl. 1), 10 661–10 668. (10.1073/pnas.1201895109) [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 55.Huxley TH. 1863. Evidence as to mankind's place in nature. London, UK: Williams & Northgate. [[Google Scholar](#)]
- 56.Dawkins MS. 1998. Through our eyes only? Oxford, UK: Oxford University Press. [[Google Scholar](#)]
- 57.Griffin DR. 2001. Animal minds. Chicago, IL: University of Chicago Press. [[Google Scholar](#)]
- 58.Edelman D, Seth AK. 2009. Animal consciousness: a synthetic approach. *Trends Neurosci.* 32, 476–484. (10.1016/j.tins.2009.05.008) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 59.Darwin C. 1881. The formation of vegetable mould through the action of worms: with observations on their habits. Cambridge, UK: Cambridge Library Collection. [[Google Scholar](#)]
- 60.Koch C, Laurent G. 1999. Complexity and the nervous system. *Science* 284, 96–98. (10.1126/science.284.5411.96) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 61.Hassin RR, Uleman JS, Bargh JA. 2005. The new unconscious. Oxford, UK: Oxford University Press. [[Google Scholar](#)]
- 62.Kouider S, Dehaene S. 2007. Levels of processing during non-conscious perception: a critical review of visual masking. *Phil. Trans. R. Soc. B* 362, 857–875. (10.1098/rstb.2007.2093) [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 63.Berlin HA. 2011. The neural basis of the dynamic unconscious. *Neuropsychanalysis* 13, 1–68. [[Google Scholar](#)]

- 64.Mudrik L, Breska A, Lamy D, Deouell LY. 2011. Integration without awareness: expanding the limits of unconscious processing. *Psychol. Sci.* 22, 764–770. (10.1177/0956797611408736) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 65.Sklar AY, Levy N, Goldstein A, Mandel R, Maril A, Hassin RR. 2012. Reading and doing arithmetic nonconsciously. *Proc. Natl Acad. Sci. USA* 109, 19 614–19 619. (10.1073/pnas.1211645109) [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 66.Hassin RR. 2013. Yes it can: on the functional abilities of the human unconscious. *Perspect. Psychol. Sci.* 8, 195–207. (10.1177/1745691612460684) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 67.Lemon RN, Edgley SA. 2010. Life without a cerebellum. *Brain* 133, 652–654. (10.1093/brain/awq030) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 68.Kurzweil R. 2012. How to create a mind: the secret of human thought revealed. New York, NY: Viking. [[Google Scholar](#)]
- 69.Crick FC, Koch C. 2005. What is the function of the claustrum? *Phil. Trans. R. Soc. B* 360, 1271–1279. (10.1098/rstb.2005.1661) [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 70.Koubeissi MZ, Bartolomei F, Beltagy A, Picard F. 2014. Electrical stimulation of a small brain area reversibly disrupts consciousness. *Epilepsy Behav.* 37, 32–35. (10.1016/j.yebeh.2014.05.027) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 71.Chalmers DJ. 1996. The conscious mind: in search of a fundamental theory. New York, NY: Oxford University Press. [[Google Scholar](#)]
- 72.Block NJ, Flanagan OJ, Güzeldere G. 1997. The nature of consciousness. Cambridge, MA: MIT Press. [[Google Scholar](#)]
- 73.Shear J. 1999. Explaining consciousness. Cambridge, MA: MIT Press. [[Google Scholar](#)]
- 74.McGinn C. 2000. The mysterious flame: conscious minds in a material world. New York, NY: Basic Books. [[Google Scholar](#)]
- 75.Tononi G. 2004. An information integration theory of consciousness. *BMC Neurosci.* 5, 42 (10.1186/1471-2202-5-42) [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 76.Tononi G. 2008. Consciousness as integrated information: a provisional manifesto. *Biol. Bull.* 215, 216–242. (10.2307/25470707) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]

- 77.Tononi G. 2012. The integrated information theory of consciousness: an updated account. Arch. Ital. Biol. 150, 56–90. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 78.Tononi G. 2012. Phi: a voyage from the brain to the soul. New York, NY: Pantheon Books. [[Google Scholar](#)]
- 79.Hoel EP, Albantakis L, Tononi G. 2013. Quantifying causal emergence shows that macro can beat micro. Proc. Natl Acad. Sci. USA 110, 19 790–19 795. (10.1073/pnas.1314922110) [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 80.Oizumi M, Albantakis L, Tononi G. 2014. From the phenomenology to the mechanisms of consciousness: integrated information theory 3.0. PLoS Comput. Biol. 10, e1003588 (10.1371/journal.pcbi.1003588) [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 81.Balduzzi D, Tononi G. 2008. Integrated information in discrete dynamical systems: motivation and theoretical framework. PLoS Comput. Biol. 4, e1000091 (10.1371/journal.pcbi.1000091) [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 82.Balduzzi D, Tononi G. 2009. Qualia: the geometry of integrated information. PLoS Comput. Biol. 5, e1000462 (10.1371/journal.pcbi.1000462) [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 83.Tononi G. 2015. Integrated information theory. Scholarpedia 10, 4164 (10.4249/scholarpedia.4164) [[DOI](#)] [[Google Scholar](#)]
- 84.Mach E. 1959. The analysis of sensations and the relation of the physical to the psychical [Transl. 1st German Ed. by CM Williams. Rev. and supplemented from the 5th German edn, 1906 by Sydney Waterlow. With a New Introd. by Thomas S. Szasz]. Mineola, NY: Dover Publications. [[Google Scholar](#)]
- 85.Moutoussis K, Zeki S. 1997. Functional segregation and temporal hierarchy of the visual perceptive systems. Proc. R. Soc. Lond. B 264, 1407–1414. (10.1098/rspb.1997.0196) [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 86.Zeki S, Moutoussis K. 1997. Temporal hierarchy of the visual perceptive systems in the Mondrian world. Proc. R. Soc. Lond. B 264, 1415–1419. (10.1098/rspb.1997.0197) [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 87.Seth AK, Baars BJ, Edelman DB. 2005. Criteria for consciousness in humans and other mammals. Conscious. Cogn. 14, 119–139. (10.1016/j.concog.2004.08.006) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]

- 88. Bateson G. 1972. Steps to an ecology of mind. Chicago, IL: University of Chicago Press. [[Google Scholar](#)]
- 89. Palmer SE. 1999. Vision science: photons to phenomenology. Cambridge, MA: MIT Press. [[Google Scholar](#)]
- 90. Tegmark M. 2014. Consciousness as a state of matter. (<http://arxiv.org/abs/1401.1219>)
- 91. Crick F, Koch C. 2003. A framework for consciousness. Nat. Neurosci. 6, 119–126. (10.1038/nn0203-119) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 92. Stroud JM. 1956. The fine structure of psychological time. In Information theory in psychology: problems and methods, pp. 174–207. New York, NY: Free Press. [[Google Scholar](#)]
- 93. VanRullen R, Koch C. 2003. Is perception discrete or continuous? Trends Cogn. Sci. 7, 207–213. (10.1016/S1364-6613(03)00095-0) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 94. VanRullen R, Zoefel B, Ilhan B. 2014. On the cyclic nature of perception in vision versus audition. Phil. Trans. R. Soc. B 369, 20130214 (10.1098/rstb.2013.0214) [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 95. Massimini M, Ferrarelli F, Huber R, Esser SK, Singh H, Tononi G. 2005. Breakdown of cortical effective connectivity during sleep. Science 309, 2228–2232. (10.1126/science.1117256) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 96. Casali AG, et al. 2013. A theoretically based index of consciousness independent of sensory processing and behavior. Sci. Transl. Med. 5, 198ra105 (10.1126/scitranslmed.3006294) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 97. Sullivan PR. 1995. Contentless consciousness and information-processing theories of mind. Philos. Psychiatry Psychol. 2, 51–59. [[Google Scholar](#)]
- 98. Blackmore S. 2009. Ten Zen questions. Oxford, UK: Oneworld Publications. [[Google Scholar](#)]
- 99. Cohen MA, Dennett DC. 2011. Consciousness cannot be separated from function. Trends Cogn. Sci. 15, 358–364. (10.1016/j.tics.2011.06.008) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 100. von Arx SW, Muri RM, Heinemann D, Hess CW, Nyffeler T. 2010. Anosognosia for cerebral achromatopsia: a longitudinal case study. Neuropsychologia 48, 970–977. (10.1016/j.neuropsychologia.2009.11.018) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]

- 101.Sperry RW. 1974. Lateral specialization in the surgically separated hemispheres. In Neuroscience 3rd Study Program (eds Schmitt FO, Worden FG.). Cambridge, MA: MIT Press. [[Google Scholar](#)]
- 102.Gazzaniga MS. 2005. Forty-five years of split-brain research and still going strong. *Nat. Rev. Neurosci.* 6, 653–659. (10.1038/nrn1723) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 103.Herculano-Houzel S. 2009. The human brain in numbers: a linearly scaled-up primate brain. *Front. Hum. Neurosci.* 3, 31 (10.3389/neuro.09.031.2009) [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 104.Cohen D. 1998. Patches of synchronized activity in the cerebellar cortex evoked by mossy-fiber stimulation: questioning the role of parallel fibers. *Proc. Natl Acad. Sci. USA* 95, 15 032–15 036. (10.1073/pnas.95.25.15032) [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 105.Bower JM. 2002. The organization of cerebellar cortical circuitry revisited. *Ann. NY Acad. Sci.* 978, 135–155. (10.1111/j.1749-6632.2002.tb07562.x) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 106.Skrbina D. 2009. Mind that abides: panpsychism in the new millennium. Amsterdam, The Netherlands: John Benjamins. [[Google Scholar](#)]
- 107.Chalmers D. 2013. The combination problem for panpsychism. *Panpsychism on the Reef*. (<http://consc.net/papers/combination.pdf>) [[Google Scholar](#)]
- 108.Strawson G, Freeman A. 2006. *Consciousness and its place in nature: does physicalism entail panpsychism?* Exeter, UK; Charlottesville, VA: Imprint Academic. [[Google Scholar](#)]
- 109.Strawson G, Freeman A. 2006. *Consciousness and its place in nature: does physicalism entail panpsychism?* Exeter, UK: Imprint Academic. [[Google Scholar](#)]
- 110.James W. 1890. *The principles of psychology*. New York, NY: H. Holt and Company. [[Google Scholar](#)]
- 111.Edlund JA, Chaumont N, Hintze A, Koch C, Tononi G, Adami C. 2011. Integrated information increases with fitness in the evolution of animats. *PLoS Comput. Biol.* 7, e1002236 (10.1371/journal.pcbi.1002236) [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]

- 112.Joshi NJ, Tononi G, Koch C. 2013. The minimal complexity of adapting agents increases with fitness. PLoS Comput. Biol. 9, e1003111 (10.1371/journal.pcbi.1003111) [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 113.Albantakis L, Hintze A, Koch C, Adami C, Tononi G. 2014. Evolution of integrated causal structures in animats exposed to environments of increasing complexity. PLoS Comput. Biol. 10, e1003966 (10.1371/journal.pcbi.1003966) [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 114.Bogen JE. 1986. Mental duality in the intact brain. Bull. Clin. Neurosci. 51, 3–29. [[PubMed](#)] [[Google Scholar](#)]
- 115.Mudrik L, Faivre N, Koch C. 2014. Information integration without awareness. Trends Cogn. Sci. 18, 488–496. (10.1016/j.tics.2014.04.009) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 116.Kohler CG, et al. 1998. Marchiafava-Bignami disease: literature review and case report. Neuropsychiatry Neuropsychol. Behav. Neurol. 13, 67–76. [[PubMed](#)] [[Google Scholar](#)]
- 117.Searle J. 2013. Can information theory explain consciousness? In New York Review of Books, vol. 60, pp. 30–41. New York, NY. [[Google Scholar](#)]
- 118.Koch C, Crick F. 2001. The zombie within. Nature 411, 893 (10.1038/35082161) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 119.Le QV, Ranzato MAA, Monga R, Devin M, Chen K, Corrado GS, Dean J, Ng AY. 2011. Building high-level features using large scale unsupervised learning. (<http://arxiv.org/abs/1112.6209>)
- 120.Hornik K, Stinchcombe M, White H. 1989. Multilayer feedforward networks are universal approximators. Neural Netw. 2, 359–366. (10.1016/0893-6080(89)90020-8) [[DOI](#)] [[Google Scholar](#)]
- 121.Koch C, Tononi G. 2011. A test for consciousness. How will we know when we've built a sentient computer? By making it solve a simple puzzle. Sci. Am. 304, 44–47. (10.1038/scientificamerican0611-44) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 122.Dennett DC. 1993. Consciousness explained. Harmondsworth, UK: Penguin. [[Google Scholar](#)]
- 123.Markram H. 2006. The blue brain project. Nat. Rev. Neurosci. 7, 153–160. (10.1038/nrn1848) [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]

- 124.Searle JR. 1980. Minds, brains, and programs. Behav. Brain Sci. 3, 417–424. (10.1017/S0140525X00005756) [[DOI](#)] [[Google Scholar](#)]
 - 125.Leibniz GW, Montgomery GR, Leibniz GW. 2005. Discourse on metaphysics, and the monadology. Mineola, NY: Dover Publications. [[Google Scholar](#)]
 - 126.Schmuker M, Pfeil T, Nawrot MP. 2014. A neuromorphic network for generic multivariate data classification. Proc. Natl Acad. Sci. USA 111, 2081–2086. (10.1073/pnas.1303053111) [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
 - 127.Koch C. 2014. Is consciousness universal? Sci. Am. Mind 25, 26–29. [[Google Scholar](#)]
-