Critical Focus

# Computational and Neuroscientific Objections to Integrated Information Theory

**Michael Egnor**[1*], **Robert J. Marks**[2]

1. Stony Brook University, Stony Brook, NY, USA
2. Baylor University, Waco, TX, USA

## Abstract

Integrated information theory (IIT) equates consciousness to integrated information, a concept that quantifies the interconnectedness of a system's physical causal properties. However, IIT relies on a metric of complexity for integrated information—the *phi* ($\Phi$) metric—that is both impractical to compute and insufficient to predict consciousness. Evaluating the Kullback-Leibler divergence for all system bifurcations causes the number of states to grow rapidly, making the calculation of $\Phi$ computationally impractical. And because high complexity, non-conscious systems exist (for example, some communication error correction codes have a large $\Phi$ but are not conscious) the proposed metric does not reliably indicate consciousness. There are also neuroscientific challenges to IIT. The theory relies on the assumption that all mental states supervene on brain states, implying that consciousness arises solely from physical processes in the brain. We propose that many studies in classical neuroscience point to the immateriality of the intellect and will, challenging the IIT of consciousness. Unlike locomotion, perception, memory, and emotion, higher cognitive powers such as intellect and will lack precise localization in the brain. Intellectual seizures do not happen, and abstract thought cannot be triggered by intraoperative brain stimulation. Intellect remains coherent in split-brain patients, and awareness persists in deep coma and near-death experiences. Studies showing free will operates separately from brain processes also support the immateriality of intellect and will. These observations challenge IIT's explanation of consciousness. In conclusion, while IIT presents an elegant mathematical framework, it does not fully capture the complexities of human consciousness. Explaining consciousness requires exploring beyond the modeling of material processes in the brain or the integration of information these processes convey.

**Cite as:** Egnor M, Marks RJ (2026). Computational and neuroscientific objections to integrated information theory. BIO-Complexity 2026 (2):1–7. **doi:**10.5048/BIO-C.2026.2.

**Editor:** Douglas D. Axe

**Received:** June 11, 2025; **Accepted:** June 24, 2025; **Published:** February 12, 2026

**Copyright:** © 2026 Egnor & Marks. This open-access article is published under the terms of the Creative Commons Attribution License, which permits free distribution and reuse in derivative works provided the original author(s) and source are credited.

**Notes:** A *Critique* of this paper, when available, will be assigned **doi:**10.5048/BIO-C.2026.2.c.

*Email: michael.egnor@stonybrookmedicine.edu

## INTRODUCTION

Integrated Information Theory (IIT) of consciousness posits that consciousness equates to integrated information. Initially proposed by Tononi in 2004, IIT aims to explain why certain physical systems, such as human beings, possess consciousness, the extent of their consciousness, and the nature of their conscious experience. According to IIT, the causal properties of a system are the same thing as the consciousness of the system. IIT accepts the phenomenology of consciousness as fundamental and, based on the nature of our experience of certain entities as conscious, infers the mechanism by which consciousness arises—i.e., IIT is a subjective-to-objective inference. Notably, IIT proposes that there are greater and lesser degrees of consciousness in a spectrum.

In IIT, the integrated information of a system, *phi* ($\Phi$), is a measure of its complexity. Integrated information applies Shannon information theory and is measured in bits. The term "bits" was coined by Claude Shannon in a classic 1948 paper that introduced the field of information theory [1]. The mathematical tools used in Integrated Information Theory are all found in Shannon's paper.

Shannon's motivation behind his information theory was to develop a theory of communication over a channel like a radio broadcast channel or a streaming service, in which a signal is transmitted and received over a noisy channel. Remarkably, Shannon showed that, even in these conditions, communication could be nearly flawless if the rate of information transfer

did not exceed the so-called *channel capacity*. If a channel capacity is 1,000 bits per second, nearly flawless communication could be achieved if a signal was transmitted at a rate less than 1,000 bits per second. Shannon's theory, when applied, allows a cell phone to receive perfect pictures even when it is raining outside and signal strength is low.

Operating under channel capacity maximizes *mutual information* between the transmitted signal and the received signal. Mutual information is a measure of the probabilistic distance between the transmitted and received signals. Mutual information is a special case of the Kullback-Leibler (KL) divergence—that is, a measure, in bits, of the degree to which two probability distributions differ. Integrated information is not concerned with communication but rather with the causal impact of one part of a system on another. Measuring the KL divergence of this pairwise impact is the key to measuring integrated information.

### The Model

There are a number of tutorials on the computations involved in Integrated Information Theory [2, 3, 4],[1] including those contained in the introductions co-authored by Tononi [5, 6]. The mathematics describing IIT can change from paper to paper, but the basic idea is as follows:

Pixels in a camera have low integrated information because the pixels function independently, with no causal influence on each other. Unlike a conscious system, which requires feedback loops and interdependence, a camera simply records incident light. One pixel does not affect another pixel. There is no connectivity, and thus no integration. Therefore, its IIT measure of consciousness, Φ, is zero.

Contrast this to the human brain where neurons, rather than pixels, are highly interconnected. One neuron, indeed numerous neurons, can affect what other neurons do. Neural feedback also allows a neuron at a given time to affect its own subsequent activity.

To compute the Φ metric of integrated information used in Integrated Information Theory, consider any system consisting of connected elements like neurons that interact with themselves. At any point in time, the system is in some state as defined by the activity of the elements. The system's past states contribute to its current state which, in turn, contributes to future states. Different past states can contribute to the same current state. Consider, for example, the state of a man being on second base in a baseball game. He could be on second because he walked and stole second base or because he hit a double, among other possibilities. Past states relate to current states probabilistically. There is a certain probability the player walked and stole second and a certain probability the player hit a double, etc.

Integrated Information Theory considers such probabilities in two ways: first, the probability of the system as a whole and, then, of the system divided into two parts severed from each other. Both are modeled by transitional probability distributions. The degree to which one distribution is separated from another is measured mathematically by the KL divergence.

There are many ways to divide a system into two parts. Each bifurcation has a KL divergence. The division that yields the smallest KL divergence, *i.e.,* the least disruptive separation, is called the *minimum information partition* (MIP). The KL divergence of the MIP is the *integrated information*, Φ, of the system. If Φ = 0 as in the case of camera pixels, the system's information can be fully accounted for by its parts individually. There is no cause-effect interaction between one pixel group and another. Φ quantifies the irreducibility of the system's cause-effect structure by analyzing all of the two-way subdivisions of the system.

Changing states is modeled by probability, yet Φ is measured in bits. In the most basic interpretation, bits are an alternative to measuring probability. According to Shannon information theory [7], the self-information of an event (B) with probability $p$ in bits is $B = -\log_2 p$.[2]

## CRITICAL OBJECTIONS TO INTEGRATED INFORMATION THEORY

There are several important objections to IIT, which we classify as either computational or neuroscientific.

### Computational objections to IIT

First, there are significant computational barriers to calculating Φ, the measure of high complexity used in IIT. IIT requires examining the KL divergence of all bifurcations of a system to calculate Φ. This number grows so quickly with respect to the number of states that calculating Φ becomes computationally prohibitive.

Consider a four-state system with states (ABCD). The binary partitions are (A|BCD), (B|ACD), (C|ABD), (D|ABC), (AB|CD), (AC|BD). And (AD|BC). For $n$ elements, there are $(2^{n-1}-1)$ such partitions. For (ABCD), $n = 4$ so there are seven partitions. Each partition requires a KL divergence analysis. For large $n$, this is computationally prohibitive because there are so many binary partitions.

To give an idea of a large number, the number of atoms in the universe is about $10^{80}$. This represents the number of partitions for just 28 states[3] (ABCD…XYZab), a number that is dwarfed compared to the complexity of the human brain. Assuming there are eighty-six billion neurons [8], the number of two-way partitions between neurons is about one followed by about twenty-six billion zeros. It is immediately obvious that, because of the exponential growth of the number of partitions, evaluation of all pairings quickly becomes computationally prohibitive.

Secondly, high complexity as measured by IIT is necessary but not sufficient for consciousness. The belief that high

[1] Kelley, L (2025) "What Is the Integrated Information Theory or IIT?." *Humanities*. https://discover.hubpages.com/education/What-is-the-Integrated-Information-Theory-or-IIT

[2] For example, the probability of flipping a coin three times and getting a tails followed by two heads is $p = \frac{1}{2}\times\frac{1}{2}\times\frac{1}{2} = \left(\frac{1}{2}\right)^3$. With heads a one and tails a zero, this corresponds to (011). And, accordingly, $B = -\log_2\left(\frac{1}{2}\right)^3 = 3$ bits of information.

[3] Set $10^{80} = 2^{n-1} - 1$ and solve for $n$

complexity measured by IIT leads to consciousness is an explicit assumption of IIT and requires faith.

Computer science professor Scott Aaronson points this out:[4]

> … having a large Φ-value is certainly not a sufficient condition for consciousness, or even for the appearance of consciousness. As a consequence, Φ can't possibly capture the essence of what makes a physical system conscious, or even of what makes a system look conscious to external observers.

It is fallacious to assert that a difference in kind is merely a tremendous difference in degree. An unbridgeable explanatory gap cannot be crossed by multiplying the things that are not explainable.

Thirdly, IIT is falsifiable. Aaronson, quoted above, also formulated a "conscious" grid of logic gates that IIT would predict as conscious but which most people would intuitively reject as such. Some error correction codes used in communication would also have a large Φ but obviously not be conscious. He argues that these thought experiments highlight flaws in IIT's framework. Although he applauds the boldness of IIT, he says:[5]

> The danger of sticking your neck out is that it can get cut off. IIT is not only falsifiable but has been falsified.

Fourthly, IIT has been critiqued as pseudoscience. In September 2023, 124 scholars argued that integrated information theory (IIT) should be labelled as pseudoscience [9]. Falsifiability and sufficiency were cited as motivation by some of the signatories.[6] Others were concerned that the popularity of IIT, often presented in the media as proven science, was eclipsing competing research into consciousness.[7]

Lastly, Claude Shannon expressed concern over the misuse of his theory of information, and this concern applies to IIT. Claude Shannon's theory of information was soon being applied everywhere, sometimes inappropriately. This bothered him. Half a decade after his classic paper was published, Shannon became concerned enough to address the hype surrounding his theory. In an article titled "The Bandwagon," he wrote [10]:

> Information theory has, in the last few years, become something of a scientific bandwagon. Starting as a technical tool for the communication engineer, it has received an extraordinary amount of publicity in the popular as well as the scientific press.
>
> Although this wave of popularity is certainly pleasant and exciting for those of us working in the field, it carries at the same time an element of danger.
>
> It will be all too easy for our somewhat artificial prosperity to collapse overnight when it is realized that the use of a few exciting words like information, entropy, redundancy, do not solve all our problems.

Shannon's concern applies to IIT where measurement of consciousness is equated to Φ which, in turn, is measured in bits using Shannon information theory.

## Neuroscientific objections to IIT

In addition to the substantial computational objections, there are neuroscientific objections to IIT. IIT depends on the assumption that mind states supervene on brain states, i.e., that consciousness arises from physical processes in the brain, without remainder. There is, however, a substantial body of neuroscientific evidence that casts doubt on that inference and that points to the immateriality of intellect and will.

IIT asserts that the informational content of a brain state causes or is identical to the corresponding mental state. What does it mean to say that a specific brain state causes or is identical to a specific mental state? It means that any change in mental state is accompanied by a corresponding change in brain state. Another way of expressing this is that the brain state is necessary and sufficient for the mental state.

In neuroscience, we can test supervenience— "i.e., causes or is identical to"—experimentally. We propose three criteria to be met together to establish causation/identity: 1) the brain state is temporally and invariably associated with the mental state, 2) ablation of the brain state ablates the mental state, and 3) activation of the brain state activates the mental state. Several landmark studies in modern neuroscience address the association, ablation and activation of brain states and mental states.

## Cerebral localization

A salient discovery of the past two centuries of neuroscience is that brain states are to some extent modular [11]. There are distinct regions of the brain that subserve distinct mental and physical activities. Neuroscientists in the 18th and 19th century noted that certain cortical regions controlled discrete abilities—regions in the posterior frontal lobes control movement on the opposite side of the body, regions in the dominant frontal and temporal lobes (usually in the left hemisphere) control speech, regions in the anterior parietal lobes control tactile sensation on the skin, and regions in the occipital lobes control vision. They inferred, not entirely unreasonably, that other regions of the brain control reason and will. They had no ability to see the brain in life, but they inferred that the shape of the skull reflected the shape of the underlying brain (which is, to a limited extent, true). Based on this inference, phrenologists created maps of personality and intellectual ability based on subtle bumps and dips on the skull.

The phrenological enterprise based on the premise that intellectual capacities are localized in the brain in a manner similar to locomotion, perception, etc. has been discredited as pseudoscience. We think pseudoscience is a bit too harsh a term for it—the better understanding is that phrenology was wrong, but it was a scientific enterprise in the sense that it made hypotheses

[4] Scott Aaronson, "Why I Am Not An Integrated Information Theorist (or, The Unconscious Expander)" Shtetl-Optimized, https://scottaaronson.blog/?p=1799

[5] Scott Aaronson " A Critique of Integrated Information Theory (IIT)" Nomen Nominandum, YouTube. https://youtu.be/DQa5cu7Ao9M?si=HL2CYkweDkwbbRf

[6] Brian Odegaard, "Statement on the IIT Letter" University of Florida, September 21, 2023. https://paclab.psych.ufl.edu/consciousness-letter/

[7] Ihm Curious, "Consciousness Theory Declared "Pseudoscience" by 124 Researchers: IIT's Adversarial Collaboration" YouTube: https://youtu.be/uPqWgB8O-Fs?si=cxV9hVGJRSm5aJuG

that could be tested. Phrenology was nearly all untrue, but the way that it was untrue is very important for our understanding of the relationship between the mind and the brain.

The inference that certain neurological activities (movement, sensation, emotion, etc.) arise from discrete regions of the brain was true, but the inference that higher intellectual function—abstract thought, reasoning, judgement, moral discernment, etc. arise from discrete regions of the brain has not been borne out by modern neuroscience. There is a sharp dichotomy between cerebral localization of movement, the generation of language, sensation, motion etc. on the one hand and abstract thought and will on the other hand. Intellect and will are not localized in the brain, at least not localized in the same way that movement, sensation, language generation etc. are localized. There is not a spot on the cerebral cortex that subserves the ability to do arithmetic, for example, whereas there is a very specific spot in the cortex that subserves the movement of the hand to write arithmetic. It is easy to elide this remarkable observation, to insist that intellect and will are too complex to be localized precisely on the cortex, but we believe that is a mistake. We should take this striking dichotomy of localization that emerged from the falsification of phrenology seriously. Some mind states are localized precisely to the brain, and some are not. We should not merely dismiss phrenology's errors; we should learn from them.

IIT inherently predicts that all aspects of consciousness arise from integrated information processing in the brain, and implicitly predicts that all aspects of consciousness—intellect and will as well as locomotion, perception, etc. should correspond to the same basic brain architecture. The paradox of cerebral localization—that higher cognitive abilities are not localized in the same fashion that motor and perceptual abilities are localized, points to a weakness in IIT.

It may be argued that the analogy between neuroscientific inferences based on phrenology and neuroscientific inferences based on fMRI imaging is weak and unfair, but the analogy between phrenology and fMRI-based neuroscience, whether weak or strong, is the topic of a major discussion in the scientific literature and in the lay press. The attribution of mental states to measurable physical states of the brain (i.e., to skull shape, which is to some degree determined by the shape of the brain, and to patterns of cerebral blood flow, which is to some degree determined by local neuronal activity is a fundamental feature of both phrenology and fMRI-based neuroscience, and both face analogous and substantial interpretative challenges.

We were led astray by phrenology not merely because of its methodological inadequacy but because of our interpretive misconceptions. We are mistaken to believe that the 19$^{th}$ century philosophical framework of phrenology, which is mechanistic materialism, is dead. It is not even past.

## Penfield—why are there no intellectual seizures?

Wilder Penfield, a pioneering epilepsy neurosurgeon working in the mid-20$^{th}$ century, reviewed his experience with thousands of seizures and the literature of epilepsy management and asked a remarkable question: why are there no mind seizures [12]? By 'mind seizures' Penfield meant why are there no seizures that evoke abstract thought—reason, free will, etc. He noted that all seizures he had studied that evoked brain activity fell into one of four categories—movement, sensations, emotion or memories. Seizures that evoke involuntary movements are common, shaking of a limb, facial movements, etc. Some seizures evoke sensations such as tingling of skin, flashes of light, unpleasant odors, etc. Rarely, seizures can evoke strong emotions (characterized by laughter or fear) or memories resembling a déjà vu experience. Some patients during seizures will do complex activities like walking, opening doors, even driving cars. There is a rare kind of seizure called a forced thinking seizure, in which a patient will obsess about a discrete activity such as checking a lock or examining an object, similar to obsessive-compulsive behavior. But no patient in Penfield's experience or in the medical literature experienced a seizure in which forced contemplation of mathematics or logic or philosophy was part or all of the ictus. There are no 'abstract thinking' seizures, and Penfield thought this odd. If a sizeable portion of the cerebral cortex were devoted to the generation of abstract thought, and generally about half of the cortical area is ascribed to 'association' areas which presumably generate such mental states, then there should at least occasionally be a random seizure which activates brain circuitry that gives rise to mathematical, logical or rational thought. But this never occurs.

It is important to clarify what is meant by abstract thought. By abstract thought we mean universals, in contrast to particular thought, which is thought about a material thing. If I see a picture of a slave laboring in a field, that is a particular thought—thought about a particular person doing something in a picture. If I contemplate the injustice of slavery, that is a universal thought—thought about a concept without any particular physical object. Of course, universal and particular thoughts are commonly intertwined—I am inspired to think of the injustice of slavery by the sight of a slave—but the distinction between particular thought and universal thought is important and has a long history in philosophy, dating back to Plato.

It is noteworthy that this striking dichotomy between the common evocation of movement, sensation, emotion and memory by focal seizures and the failure of seizures to evoke abstract thought is not a consequence of complexity. Patients experiencing partial seizures may do complex things—walk around the house, operate machinery, remember events from the distant past, etc.—whereas they never experience during the ictus simple mental states such as adding 1+1=2 or thinking logically that A is A. It is the abstract nature of mental states—the contemplation of universals rather than particulars—that never occurs during an epileptic seizure.

Penfield concluded the most parsimonious explanation for the failure to evoke abstract thought is that while movement, sensation, emotion and memory clearly have their origin in brain activity, abstract thought—intellect and will—is not caused by the brain. Undoubtedly, normal exercise of reason or free will depends on proper functioning of the brain—reasoning and will are impaired by alcohol or brain injury—but brain

activity, while necessary for ordinary abstract thought, is not sufficient for abstract thought. Penfield referred to this abstract agency partially independent of brain activity as 'mind activity', and he considered the brain's function to connect abstract mental states with the physical world.

It is worth noting that Penfield's observation that seizures never evoke abstract thought coincides with the insight that the study of brain localization provided—there seems to be a disconnect between mental states that entail abstract thought and mental states that entail concrete action and experience. Brain states correlate closely with and seem to cause concrete actions and experience, but do not seem to give rise to abstract thought, although normal brain states may be permissive for abstract thought. This disconnect between abstract mental states and concrete mental states. is not predicted by IIT.

### Penfield—why can only movement, sensation, emotion and memory be evoked by brain stimulation?

Penfield carried his research on epilepsy and the outcome of brain stimulation into the operating room. He pioneered awake brain surgery, in which patients with epilepsy and tumors would be awake during the critical parts of the operation. The scalp was anesthetized with local anesthetic and the brain is insensate to pain. Penfield did this to map the brain—he would methodically stimulate large expanses of cortex with an electrical probe and determine the functional anatomy of the cortex so as to safely remove epileptic foci. He performed over 1100 awake brain operations, entailing hundreds of thousands of individual stimulations, and found that, like epileptic seizures, he could stimulate only four neurological responses: movement, sensation, memory and emotion. He could not stimulate abstract thought, such as reasoning, logic, etc. He concluded from this, in accordance with his research on the manifestations of epilepsy, that the brain caused movement, sensation, memory and emotion, but it did not cause intellectual activity involving what he called 'mind' activity, which entails abstract thought [12].

As research on cerebral localization and the ictal manifestations of epilepsy, direct intraoperative brain stimulation demonstrates a remarkable dichotomy between concrete/particular neurological functions and mental states such as movement, sensation, memory and emotion and abstract mental states, which clearly depend in many ways on brain function, but cannot be elicited by brain stimulation. This dichotomy is not predicted or explained by IIT.

### Split brain—Sperry

In the mid-20$^{th}$ century neurophysiologist Roger Sperry studied hemispheric brain localization in patients who had undergone corpus callosotomy (split brain surgery)[13], which is a neurosurgical operation in which the corpus callosum (a large axonal bundle that connects the cerebral hemispheres) is cut to prevent the propagation of seizures. Sperry confirmed evidence for motor and sensory localization of function in the cerebral hemispheres—each hemisphere subserves movement and sensation on the opposite side of the body. Most patients have left hemispheric dominance for language and the right hemisphere mediates perception of spatial relationships. Sperry also noted an outcome from split brain surgery that surprised him—the patients were for ordinary purposes neurologically normal despite surgical disconnection of the axons directly connecting their brain hemispheres. Occasionally odd movements of limbs occurred. A few patients would deliberately unbutton their shirt with one hand while the other hand would button it (alien hand syndrome). Yet when considering the unity of consciousness and abstract thought, patients remained with singular unified consciousness. The unity of consciousness, combined with split perception, is neither predicted nor explained by IIT.

### Split brain—Sergent and Pinto

In the late 20$^{th}$ and early 21$^{st}$ centuries, Justine Sergent and Yair Pinto extended Sperry's work. Sergent [14] asked patients who had split brain surgery to compare images presented simultaneously to each separate hemisphere, despite the fact that no part of the patients' brains 'saw' both images (e.g., patients were asked to compare the directions of two different arrows in the right and left visual field). They were generally able to do so. Over the past few decades, Yair Pinto [15] has extended Sergent's work. He has studied the cognitive abilities of split-brain patients by presenting different images to each hemisphere simultaneously in which sense could only be made of the situation if both images were seen and understood. As in Sergent's work, Pinto found that patients were able to integrate information presented simultaneously to separated hemispheres despite the fact that no region of the brain had both images presented to it. He referred to this remarkable cognitive phenomenon as "unified consciousness with split perception" [15].

As with research on cerebral localization, as well epilepsy phenomenology and brain stimulation by Penfield, Sperry, Sergent and Pinto demonstrated a sharp dichotomy between perception and abstract thought. Perception is caused directly by brain activity and can be elicited and suppressed surgically. Yet there exists an ability to integrate perceptions (i.e., reason, judgement and abstract thought) for which brain activity is necessary but not sufficient.

Their work points to intellectual abilities (comparison of images presented to separated hemispheres) that are independent of brain function—no part of the brain has access to both images—and is inconsistent with IIT.

### Awareness in PVS—Owen

In 2006, Arian Owen published a landmark paper titled *Detecting Awareness in the Vegetative State* [16]. Owen reported fMRI imaging on a patient diagnosed with persistent vegetative state (PVS). The traditional understanding of PVS is that it represented a state in which internal mental states are absent. Owen performed fMRI brain imaging while the woman was asked to imagine doing various activities, such as playing tennis, walking across the room, etc. Her brain showed patterns of fMRI activation similar to that of 15 normal controls, and the fMRI activation disappeared when the words in the requests

were scrambled nonsensically. There was clear evidence that despite the massive brain damage, she was capable of thought at a reasonably high level. Subsequent studies confirm that at least 40% of patients reliably diagnosed in PVS are capable of imagination, reasoning, and communication. Again, there is a striking dichotomy between abstract thought and concrete activities such as movement, perception, etc., all of which was profoundly impaired in these PVS patients. The retention of higher-level cognitive abilities in patients with severe brain damage and markedly impaired motor and perceptual abilities is difficult to explain from the perspective of IIT.

### Near Death Experiences

Reports of near-death experiences (NDEs) are increasing due to medical advances that revive patients from brain death. An analysis of 11 studies shows NDEs occur in 18 to 47% of resuscitations and University of Virginia psychiatrist Bruce Greyson suggests reliable estimates range between 9% and 18% [17]. Given the number of annual resuscitations, this indicates many NDEs happen. Near-death experiences are reported across different cultures and time periods. People having near-death experiences (NDEs) sometimes report events they should not be able to observe. These veridical NDEs are later confirmed by others. Researchers define these as any perception during an NDE—visual, auditory, kinesthetic, or olfactory—that is later corroborated with reality. Over the past 50 years, near-death experiences have been extensively studied. In 2009, Janice Holden and colleagues reported 107 cases of NDE patients with verified observations [18].

The frequency and common occurrence of NDEs suggest the possibility of consciousness that is not entirely dependent on brain function. Though not universally accepted by scientists, NDEs challenge theories of consciousness solely based on brain function.

### Free will—Penfield

In addition to his ground-breaking research on epilepsy and intraoperative brain mapping, Wilder Penfield conducted a remarkable investigation of free will [12]. During awake brain mapping, he told patients that he was going to occasionally stimulate the arm area of their brain to cause their arm to raise. He also asked them voluntarily raise their arm of their own free will. They could not see what Penfield was doing in the operative field, and the brain has no sensation of its own. When their arm went up, he asked them if their arm went up because they freely chose to do so or because he stimulated the arm area of their brain. Patients always got it right—they always retained the sense of free will and the ability to distinguish between freely chosen actions and brain stimulation without their will. He showed that, like the intellect, the will cannot be triggered by brain stimulation, suggesting it does not originate from the brain. If free will does not originate from the brain, it cannot be explained by IIT.

### Free will—Libet

Benjamin Libet pioneered work in the neurophysiology of free will [19]. He asked volunteers to decide to press a button, and then immediately press it, while EEG recordings were ongoing and they were able to note the moment they decided to push the button on a clock in front of them. Libet found that the conscious awareness of the decision to push the button was preceded by a spike in brain activity by about half a second, but the decision to veto the initial decision was not associated with a change in brain activity. He famously called this phenomenon "free won't" and concluded that while brain activity provided pre-conscious proclivity to act, the final decision to act was often independent of brain function.

## CONCLUSION

Integrated Information Theory (IIT) proposes that consciousness is identical to integrated information, quantified as $\Phi$ (phi), a measure derived from Shannon information theory that captures the interdependence within a system. While IIT offers an ambitious, mathematically grounded framework claiming that greater causal integration entails greater degrees of consciousness, it faces serious challenges on both computational and neuroscientific grounds. Computationally, calculating $\Phi$ is intractable for systems of realistic size, and high $\Phi$ appears neither sufficient nor uniquely indicative of consciousness, as illustrated by counterexamples and critiques from computer scientists such as Scott Aaronson. Neuroscientifically, IIT presupposes that all mental states supervene fully on brain states, yet extensive evidence from cerebral localization, epilepsy, direct brain stimulation, split-brain research, studies of persistent vegetative state, near-death experiences, and investigations of free will suggests a persistent dichotomy: concrete functions like movement, sensation, memory, and emotion are reliably elicited by brain activity, whereas abstract thought, intellect, and will are not. This dissociation—observed repeatedly across diverse experimental paradigms—poses a fundamental problem for IIT, which predicts a uniform brain-based explanation of all aspects of consciousness and thus struggles to account for the apparent independence of higher cognitive and volitional capacities from integrated neural information alone.

We argue that there is no logical or evidence-based reason to presume consciousness in any and all systems that exhibit complex interactions in a cause-effect structure. All of nature and everything in nature exhibits complex interactions and cause-effect structures. Unless one embraces panpsychism, one has no a priori or evidence- based reason to attribute consciousness to all such systems. Theories such as IIT that attribute consciousness to such systems have yet to make a cogent metaphysical argument as to why such a radical assumption should be accepted. We point this out in our paper.

The foundation of IIT and much of modern materialist neuroscience of consciousness is flawed. Consciousness is not a product of computation, which involves matching inputs to

outputs based on predefined rules. Intentionality (the aboutness of mental states), which is a core feature of consciousness, is not rule-based. A word processing program, a computational system, functions irrespective of the meaning of the essay typed on it. One can use the same word processing program to type an essay defending ITT as one can use to critique IIT, just as a digital camera can photograph a summer scene or a winter scene without holding an opinion on weather. All thought has meaning, and computation is blind to meaning.

A particularly striking example of IIT's incongruity with neuroscience is the undeniable presence of consciousness in children afflicted with hydranencephaly, in which intrauterine strokes lead to the loss of both cerebral hemispheres. Despite their severe motor and sensory disabilities, these children are clearly conscious and show various cognitive interactions with their surroundings, such as fear, attention, and laughter. Theories linking consciousness to brain complexity fail because patients without brain hemispheres still experience consciousness.

The fundamental weaknesses of IIT as an explanation for consciousness are its computational deficiencies and its inconsistencies with modern neuroscience. IIT is a computationally suspect model of the fallacies that pervade contemporary neuroscience. The elegance of the mathematics of IIT obscures the misconceptions at its core. Despite the ingenuity of many recent materialist theories of mind [20–22], any attempt to explain human consciousness by the modeling of material processes in the brain or of the integration of information conveyed by such processes is doomed to failure.


1. Shannon CE (1948) A mathematical theory of communication. The Bell System Technical Journal 27(3): 379–42. **doi:**10.1002/j.1538-7305.1948.tb01338.x
2. Hunt T (2016) Taking time seriously in Tononi's Integrated Information Theory. J Conscious Stud 23(9-10): 88–110.
3. Lundbak Olesen C, Thestrup Waade P, Albantakis L, Mathys C (2023) Phi fluctuates with surprisal: An empirical pre-study for the synthesis of the free energy principle and integrated information theory. PLOS Comput Biol 19(10): e1011346. **doi:**10.1371/journal.pcbi.1011346
4. Krohn S, Ostwald D (2017) Computing integrated information. Neurosci Conscious 2017(1): nix017. **doi:**10.1093/nc/nix017
5. Tononi G, Sporns O (2003) Measuring information integration. BMC Neurosci 4: 1–20. **doi:**10.1186/1471-2202-4-31
6. Oizumi M, Albantakis L, Tononi G (2014) From the phenomenology to the mechanisms of consciousness: integrated information theory 3.0. PLoS Comput Biol 10(5): e1003588. **doi:**10.1371/journal.pcbi.1003588
7. Cover, TM (1999) Elements of information theory. John Wiley & Sons (Hoboken, NJ).
8. Herculano-Houzel S (2009) The human brain in numbers: a linearly scaled-up primate brain. Front Hum Neurosci 3: 857. **doi:**10.3389/neuro.09.031.2009
9. Lenharo M (2023) Consciousness theory slammed as 'pseudoscience'—sparking uproar. Nature 20. **doi:**10.1038/d41586-023-02971-1
10. Shannon CE (1956) The Bandwagon. IRE Transactions on Information Theory 2(1): 3. **doi:**10.1109/TIT.1956.1056774
11. Sabbatini RME (1997) Phrenology: The history of brain localization. Brain and Mind 1: 69–98. https://www.thebrain.mcgill.ca/flash/capsules/pdf_articles/phrenology.pdf
12. Penfield W (1975) The Mystery of the Mind. Princeton University Press (Princeton, NJ). **doi:**10.1515/9781400868735
13. Sperry RW (1968) Hemisphere deconnection and unity in conscious awareness. Am Psychol 23(10): 722–3. **doi:**10.1037/h0026839
14. Sergent J (1987) A new look at the human split brain. Brain 110(5): 1375–1392. **doi:**10.1093/brain/110.5.1375
15. Pinto Y, Neville DA, Otten M, Corballis PM, Lamme VAF, et al (2017). Split brain: divided perception but undivided consciousness. Brain 140(5): 1231–1237. **doi:**10.1093/brain/aww358
16. Owen AM, Coleman MR, Boly M, Davis MH, Laureys S, et al (2006) Detecting awareness in the vegetative state. Science 313(5792): 1402. **doi:**10.1126/science.1130197
17. Greyson B (1998) The incidence of near-death experiences. Medicine and Psychiatry 1: 92–99
18. Holden JM (2009) Veridical perception in near-death experiences. In: Holden JM, Greyson B, James D, eds. The Handbook of Near-Death Experiences: Thirty Years of Investigation. Praeger (Santa Barbara) pp. 185–211. **doi:**10.5040/9798400661389.ch-009
19. Libet B (1999) Do we have free will? J Conscious Stud 6 (8–9): 47–57.
20. Poznanski RR, Alemdar E (2024) The 'hidden' structure of uncertainties unfolding through poststructural dynamics of the entropic brain. Journal of Multiscale Neuroscience 4: 264–273. **doi:**10.56280/1667295098
21. Hoel EP, Albantakis L, Marshall W, Tononi G (2016) Can the macro beat the micro? Integrated information across spatiotemporal scales. Neurosci Conscious 2016(1): niw012. **doi:**10.1093/nc/niw012
22. Noble D (2011) A theory of biological relativity: no privileged level of causation. Interface Focus 2: 55–64. **doi:**10.1098/rsfs.2011.0067