getpocket.com

Is Consciousness Everywhere?

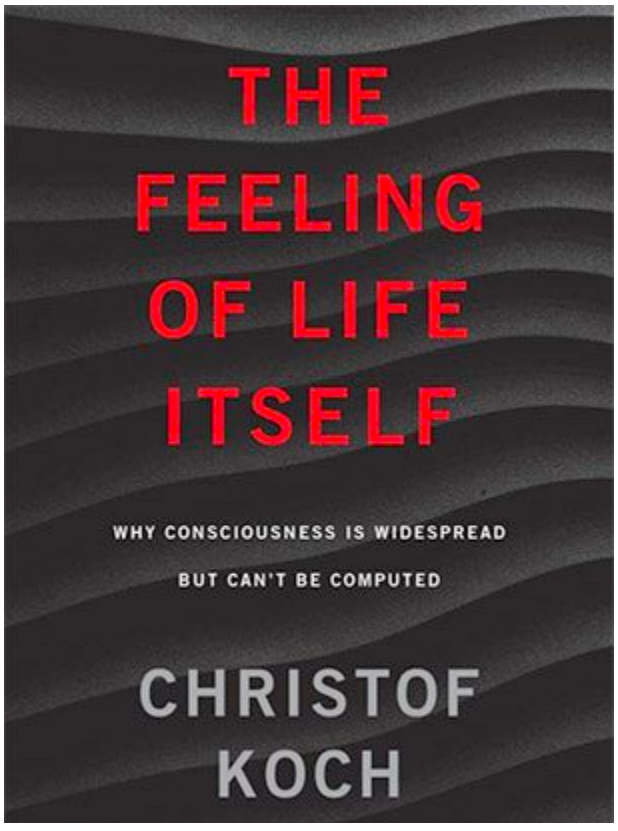
Christof Koch

28-35 minutes

Honey bees can recognize faces, communicate the location and quality of food sources to their sisters via the waggle dance, and navigate complex mazes with the help of cues they store in short-term memory. Image: Boba Jaglicic/Unsplash

What is common between the delectable taste of a favorite food, the sharp sting of an infected tooth, the fullness after a heavy meal, the slow passage of time while waiting, the willing of a deliberate act, and the mixture of vitality, tinged with anxiety, just before a competitive event?

All are distinct experiences. What cuts across each is that all are subjective states, and all are consciously felt. Accounting for the nature of consciousness appears elusive, with many claiming that it cannot be defined at all, yet defining it is actually straightforward. Here goes: Consciousness is experience.



This article is adapted from Christof Koch's book "The Feeling of Life Itself." Koch is chief scientist of the MindScope Program at the Allen Institute for Brain Science.

That's it. Consciousness is any experience, from the most mundane to the most exalted. Some distinguish *awareness* from consciousness; I don't find this distinction helpful and so I use these two words interchangeably. I also do not distinguish between *feeling* and *experience*, although in everyday use feeling is usually reserved for strong emotions, such as feeling angry or in love. As I use it, any feeling is an experience. Collectively taken, then, consciousness is lived reality. It is the feeling of life itself.

But who else, besides myself, has experiences? Because you are so similar to me, I abduce that you do. The same logic applies to other people. Apart from the occasional solitary solipsist this is uncontroversial. But how widespread is consciousness in the cosmos at large? How far consciousness extends its dominion within the tree

of life becomes more difficult to abduce as species become more alien to us.

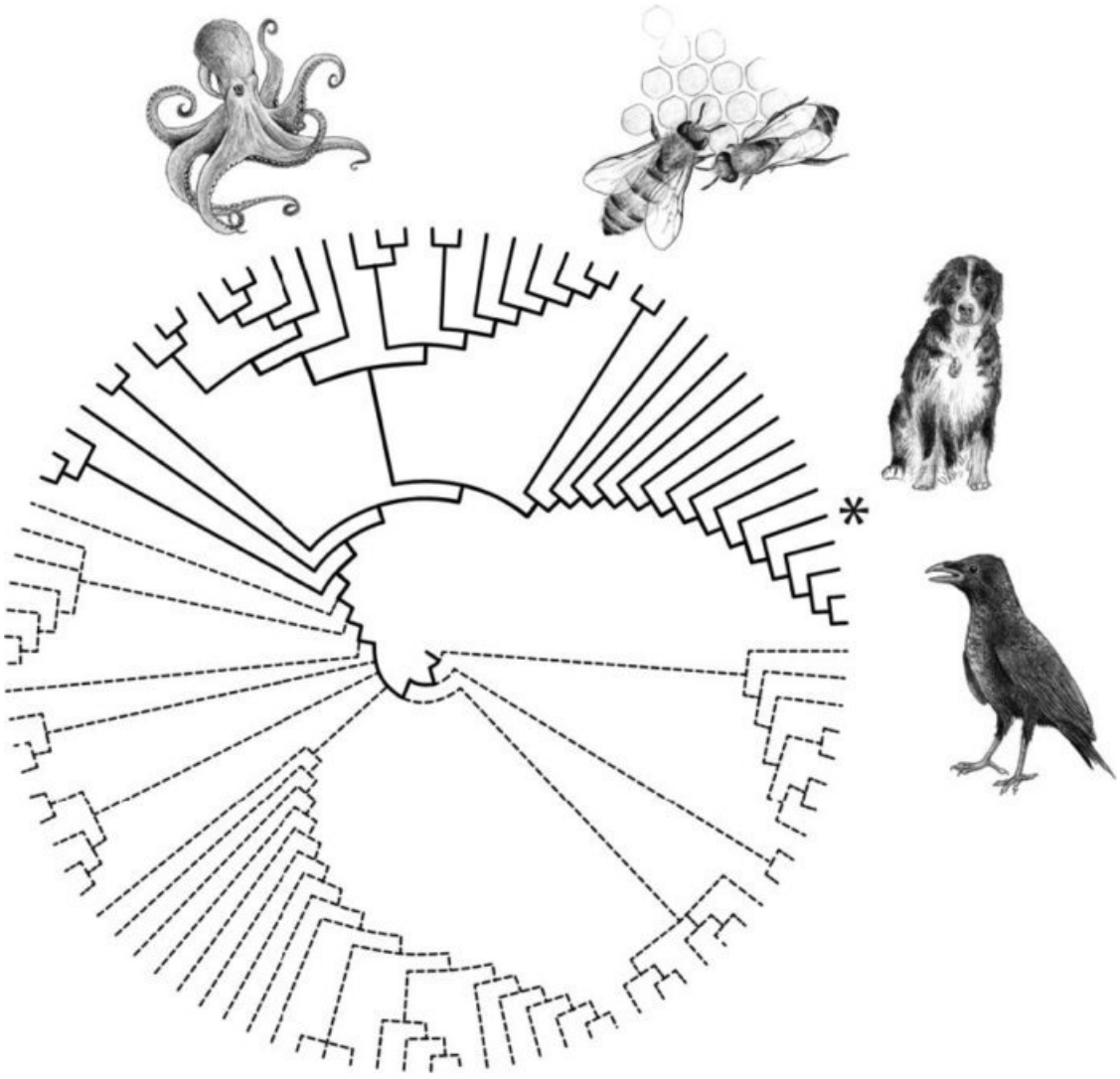
One line of argument takes the principles of integrated information theory (IIT) to their logical conclusion. Some level of experience can be found in all organisms, it says, including perhaps in *Paramecium* and other single-cell life forms. Indeed, according to IIT, which aims to precisely define both the quality and the quantity of any one conscious experience, experience may not even be restricted to biological entities but might extend to non-evolved physical systems previously assumed to be mindless — a pleasing and parsimonious conclusion about the makeup of the universe.

How Widespread Is Consciousness in the Tree of Life?

The evolutionary relationship among bacteria, fungi, plants, and animals is commonly visualized using the tree of life metaphor. All living species, whether fly, mouse, or person, lie somewhere on the periphery of the tree, all equally adapted to their

particular ecological niches.

Every living organism descends in an unbroken lineage from the last universal common ancestor (abbreviated to a charming LUCA) of planetary life. This hypothetical species lived an unfathomable 3.5 billion years ago, smack at the center of the tree-of-life mandala. Evolution explains not only the makeup of our bodies but also the constitution of our minds — for they don't get a special dispensation.



*The tree of life: Based on the complexity of their behavior and nervous systems, it is likely that it feels like something to be a bird, mammal (marked by *), insect, and cephalopod — represented here by a crow, dog, bee, and octopus. The extent to which consciousness is shared across the entire animal kingdom, let alone across all of life's vast domain, is at present difficult to establish. The last universal common ancestor of all living things is at the center, with time radiating outward.*

Given the similarities at the behavioral, physiological, anatomical, developmental, and genetic levels between *Homo sapiens* and other mammals, I have no reason to doubt that all of us experience the sounds and sights, the pains and pleasures of life, albeit not necessarily as richly as we do. All of us strive to eat and drink, to procreate, to avoid injury and death; we bask in the sun's warming rays, we seek the company of conspecifics, we fear predators, we sleep, and we dream.

While mammalian consciousness depends on a functioning six-layered neocortex, this does not imply that animals without a neocortex do not feel. Again, the similarities between the structure,

dynamics, and genetic specification of nervous systems of all tetrapods — mammals, amphibians, birds (in particular ravens, crows, magpies, parrots), and reptiles — allows me to abduce that they too experience the world. A similar inference can be made for other creatures with a backbone, such as fish.

But why be a vertebrate chauvinist? The tree of life is populated by a throng of invertebrates that move about, sense their environment, learn from prior experience, display all the trappings of emotions, communicate with others — insects, crabs, worms, octopuses, and on and on. We might balk at the idea that tiny buzzing flies or diaphanous pulsating jellyfish, so foreign in form, have experiences.

Yet honey bees can recognize faces, communicate the location and quality of food sources to their sisters via the waggle dance, and navigate complex mazes with the help of cues they store in short-term memory. A scent blown into a hive can trigger a return to the place where the bees previously encountered this odor, a type of associative memory. Bees have collective decision-making skills that, in

their efficiency, put any academic faculty committee to shame. This “wisdom of the crowd” phenomenon has been studied during swarming, when a queen and thousands of her workers split off from the main colony and chooses a new hive that must satisfy multiple demands crucial to group survival (think of that when you go house hunting). Bumble bees can even learn to use a tool after watching other bees use them.

Charles Darwin, in an 1881 book on earthworms, wanted “to learn how far the worms acted consciously and how much mental power they displayed.” Studying their feeding behaviors, Darwin concluded that there was no absolute threshold between complex and simple animals that assigned higher mental powers to one but not to the other. No one has discovered a Rubicon that separates sentient from nonsentient creatures.

Of course, the richness and diversity of animal consciousness will diminish as their nervous system becomes simpler and more primitive, eventually turning into a loosely organized neural net. As the pace of the underlying assemblies becomes more

sluggish, the dynamics of the organisms' experiences will slow down as well.

Does experience even require a nervous system? We don't know. It has been asserted that trees, members of the kingdom of plants, can communicate with each other in unexpected ways, and that they adapt and learn. Of course, all of that can happen without experience. So I would say the evidence is intriguing but very preliminary. As we step down the ladder of complexity rung by rung, how far down do we go before there is not even an inkling of awareness? Again, we don't know. We have reached the limits of abduction based on similarity with the only subject we have direct acquaintance with — ourselves.

Consciousness in the Universe

IIT offers a different chain of reasoning. The theory precisely answers the question of who can have an experience: anything with a non-zero maximum of integrated information; anything that has intrinsic causal powers is considered a Whole. What this

Whole feels, its experience, is given by its maximally irreducible cause-effect structure. How much it exists is given by its integrated information.

In other words, the theory doesn't stipulate that there is some magical threshold for experience to switch on. The degree of consciousness is instead measured with Φ, or *phi*. If phi is zero, then the system doesn't exist for itself; anything with Φ^{max} greater than zero exists for itself, has an inner view, and has some degree of irreducibility — the larger this number, the more conscious it is. And that means there are a lot of Wholes out there.

Certainly, this includes people and other mammals with neocortex, which we clinically know to be the substrate of experience. But fish, birds, reptiles, and amphibians also possess a telencephalon — the largest and most highly developed part of the brain — that is evolutionarily related to mammalian cortex. Given the attendant circuit complexity, the intrinsic causal power of the telencephalon is likely to be high.

When considering the neural architecture of creatures very different from us, such as the honey

bee, we are confronted by vast and untamed neuronal complexity — about one million neurons within a volume the size of a grain of quinoa, a circuit density 10 times higher than that of our neocortex of which we are so proud. And unlike our cerebellum, the bee's mushroom-shaped body is heavily recurrently connected. It is likely that this little brain forms a maximally irreducible cause-effect structure.

Integrated information is not about input–output processing, function or cognition, but about intrinsic cause-effect power. Having liberated itself from the myth that consciousness is intimately related to intelligence, the theory is free to discard the shackles of nervous systems and to locate intrinsic causal power in mechanisms that do not compute in any conventional sense.

A case in point is that of single-cell organisms, such as *Paramecium*, the *animalcules* discovered by the early microscopists in the late 17th century. Protozoa propel themselves through water by whiplash movements of tiny hairs, avoid obstacles, detect food, and display adaptive responses.

Because of their minuscule size and strange habitats, we don't think of them as sentient. But they challenge our presuppositions. One of the early students of such microorganisms, H. S. Jennings, expressed this well:

> The writer is thoroughly convinced, after long study of the behavior of this organism, that if Amoeba were a large animal, so as to come within the everyday experience of human beings, its behavior would at once call forth the attribution to it of states of pleasure and pain, of hunger, desire, and the like, on precisely the same basis as we attribute these things to the dog.

Among the best-studied of all organisms are the even smaller *Escherichia coli*, bacteria that can cause food-poisoning. Their rod-shaped bodies, about the size of a synapse, house several million proteins inside their protective cell wall. No one has modeled in full such vast complexity. Given this byzantine intricacy, the causal power of a bacterium upon itself is unlikely to be zero. Per IIT, it is likely that it feels like something to be a bacterium. It won't be upset about its pear-shaped body; no one will

ever study the psychology of a microorganism. But there will be a tiny glow of experience. This glow will disappear once the bacterium dissolves into its constituent organelles.

Let us travel down further in scale, transitioning from biology to the simpler worlds of chemistry and physics, and compute the intrinsic causal power of a protein molecule, an atomic nucleus or even a single proton. Per the standard model of physics, protons and neutrons are made out of three quarks with fractional electrical charge. Quarks are never observed by themselves. It is therefore possible that atoms constitute an irreducible Whole, a modicum of "enminded" matter. How does it feel to be a single atom compared to the roughly 10^{26} atoms making up a human brain? Given that its integrated information is presumably barely above zero, just a minute bagatelle, a this-rather-than-not-this?

To wrap your mind around this possibility that violates Western cultural sensibilities, consider an instructive analogy. The average temperature of the universe is determined by the afterglow left over from the Big Bang, the cosmic microwave

background radiation. It evenly pervades space at an effective temperature of 2.73° above absolute zero. This is utterly frigid, hundreds of degrees colder than any temperature terrestrial organisms can survive. But the fact that the temperature is non-zero implies a corresponding tiny amount of heat in deep space. This of course implies a corresponding tiny amount of experience.

To the extent that I'm discussing the mental with respect to single-cell organisms let alone atoms, I have entered the realm of pure speculation, something I have been trained all my life as a scientist to avoid. Yet three considerations prompt me to cast caution to the wind.

First, these ideas are straightforward extensions of IIT — constructed to explain human-level consciousness — to vastly different aspects of physical reality. This is one of the hallmarks of a powerful scientific theory — predicting phenomena by extrapolating to conditions far from the theory's original remit. There are many precedents — that the passage of time depends on how fast you travel, that spacetime can break down at singularities

known as black holes, that people, butterflies, vegetables, and the bacteria in your gut use the same mechanism to store and copy their genetic information, and so on.

Second, I admire the elegance and beauty of this prediction. (Yes, I'm perfectly cognizant that the last 40 years in theoretical physics have provided ample proof that chasing after elegant theories has yielded no new, empirically testable evidence describing the actual universe we live in.) The mental does not appear abruptly out of the physical. As Leibniz expressed it, *natura non facit saltus*, or nature does not make sudden leaps (Leibniz was, after all, the co-inventor of infinitesimal calculus). The absence of discontinuities is also a bedrock element of Darwinian thought.

Intrinsic causal power does away with the challenge of how mind emerges from matter. IIT stipulates that it is there all along.

Third, IIT's prediction that the mental is much more widespread than traditionally assumed resonates with an ancient school of thought: *panpsychism*.

Many but Not All Things Are Enminded

Common to panpsychism in its various guises is the belief that soul (*psyche*) is in everything (*pan*), or is ubiquitous; not only in animals and plants but all the way down to the ultimate constituents of matter — atoms, fields, strings, or whatever. Panpsychism assumes that any physical mechanism either is conscious, is made out of conscious parts, or forms part of a greater conscious whole.

Some of the brightest minds in the West took the position that matter and soul are one substance. This includes the pre-Socratic philosophers of ancient Greece, Thales, and Anaxagoras. Plato espoused such ideas, as did the Renaissance cosmologist Giordano Bruno (burned at the stake in 1600), Arthur Schopenhauer, and the 20th-century paleontologist and Jesuit Teilhard de Chardin (whose books, defending evolutionary views on consciousness, were banned by his church until his death).

Particularly striking are the many scientists and mathematicians with well-articulated panpsychist

views. Foremost, of course, is Leibniz. But we can also include the three scientists who pioneered psychology and psychophysics — Gustav Fechner, Wilhelm Wundt, and William James — and the astronomer and mathematicians Arthur Eddington, Alfred North Whitehead, and Bertrand Russell. With the modern devaluation of metaphysics and the rise of analytic philosophy, the last century evicted the mental entirely, not only from most university departments but also from the universe at large. But this denial of consciousness is now being viewed as the “Great Silliness,” and panpsychism is undergoing a revival within the academe.

Debates concerning what exists are organized around two poles: materialism and idealism. Materialism, and its modern version known as *physicalism*, has profited immensely from Galileo Galilei’s pragmatic stance of removing mind from the objects it studies in order to describe and quantify nature from the perspective of an outside observer. It has done so at the cost of ignoring the central aspect of reality — experience. Erwin Schrödinger, one of the founders of quantum mechanics, after

whom its most famous equation is named, expressed this clearly:

> The strange fact [is] that on the one hand all our knowledge about the world around us, both that gained in everyday life and that revealed by the most carefully planned and painstaking laboratory experiments, rests entirely on immediate sense perception, while on the other hand this knowledge fails to reveal the relations of the sense perceptions to the outside world, so that in the picture or model we form of the outside world, guided by our scientific discoveries, all sensual qualities are absent.

Idealism, on the other hand, has nothing productive to say about the physical world, as it is held to be a figment of the mind. Cartesian dualism accepts both in a strained marriage in which the two partners live out their lives in parallel, without speaking to each other (this is the *interaction* problem: how does matter interact with the ephemeral mind?). Behaving like a thwarted lover, analytic, logical-positivist philosophy denies the legitimacy and, in its more extreme version, even the very existence of one partner in the mental-physical relationship. It does

so to obfuscate its inability to deal with the mental.

Panpsychism is unitary. There is only one substance, not two. This elegantly eliminates the need to explain how the mental emerges out of the physical and vice versa. Both coexist.

But panpsychism's beauty is barren. Besides claiming that everything has both intrinsic and extrinsic aspects, it has nothing constructive to say about the relationship between the two. Where is the experiential difference between one lone atom zipping around in interstellar space, the hundred trillion trillion making up a human brain, and the uncountable atoms making up a sandy beach? Panpsychism is silent on such questions.

IIT shares many insights with panpsychism, starting with the fundamental premise that consciousness is an intrinsic, fundamental aspect of reality. Both approaches argue that consciousness is present across the animal kingdom to varying degrees.

All else being equal, integrated information, and with it the richness of experience, increases as the complexity of the associated nervous system grows, although sheer number of neurons is not a

guarantee, as shown by the cerebellum. Consciousness waxes and wanes diurnally with alertness and sleep. It changes across the lifespan — becoming richer as we grow from a fetus into a teenager and mature into an adult with a fully developed cortex. It increases when we become familiar with romantic and sexual relationships, with alcohol and drugs, and when we acquire appreciation for games, sports, novels, and art; and it will slowly disintegrate as our aging brains wear out.

Most importantly, though, IIT is a scientific theory, unlike panpsychism. IIT predicts the relationship between neural circuits and the quantity and quality of experience, how to build an instrument to detect experience, pure experience (consciousness without any content) and how to enlarge consciousness by brain-bridging, why certain parts of the brain have it and others not (the posterior cortex versus the cerebellum), why brains with human-level consciousness evolved, and why conventional computers have only a tiny bit of it.

When lecturing about these matters, I often get the

you've-got-to-be-kidding-stare. This passes once I explain how neither panpsychism nor IIT claim that elementary particles have thoughts or other cognitive processes. Panpsychism does, however, have an Achilles' heel — the *combination* problem, a problem that IIT has squarely solved.

On the Impossibility of Group Mind, or Why Your Neurons Are Not Conscious

William James gave a memorable example of the combination problem in the foundational text of American psychology, "The Principles of Psychology" (1890):

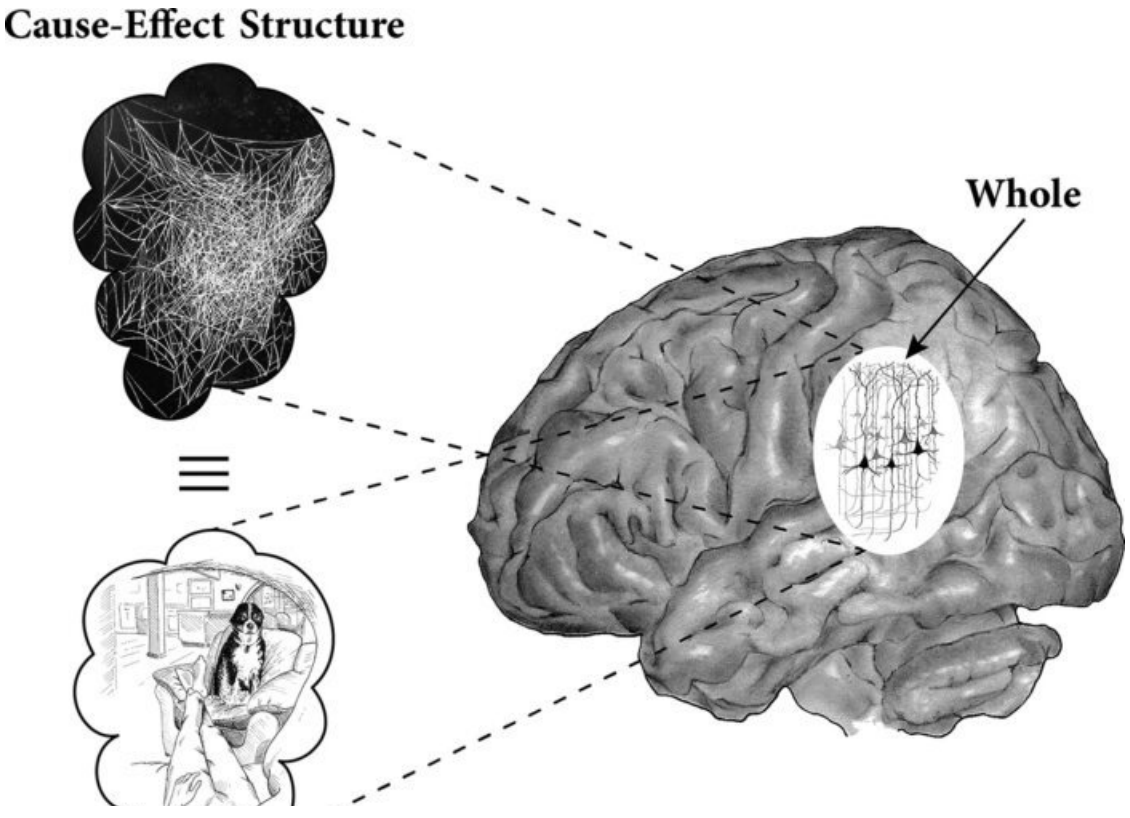
> Take a sentence of a dozen words, and take twelve men and tell to each one word. Then stand the men in a row or jam them in a bunch, and let each think of his word as intently as he will; nowhere will there be a consciousness of the whole sentence.

Experiences do not aggregate into larger, superordinate experiences. Closely interacting lovers, dancers, athletes, soldiers, and so on do not give rise to a group mind, with experiences above

and beyond those of the individuals making up the group. John Searle wrote:

> 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.

Panpsychism has not provided a satisfactory answer as to why this should be so. But IIT does. IIT postulates that only maxima of integrated information exist. This is a consequence of the exclusion axiom — any conscious experience is definite, with borders. Certain aspects of experience are in, while a vast universe of possible feelings are out.



The mind–body problem resolved? Integrated information theory posits that any one conscious experience, here that of looking at a Bernese mountain dog, is identical to a maximally irreducible cause-effect structure. Its physical substrate, its Whole, is the operationally defined neural correlate of consciousness. The experience is formed by the Whole but is not identical to it.

Consider the image above, in which I'm looking at my dog Ruby and have a particular visual experience, a maximally irreducible cause-effect structure. It is constituted by the underlying physical substrate, the Whole, here a particular neural correlate of consciousness within the hot zone in my posterior cortex. But the experience is not identical to the Whole. My experience is not my brain.

This Whole has definite borders; a particular neuron is either part of it or not. The latter is true even if this neuron provides some synaptic input to the Whole. What defines the Whole is a maximum of integrated information, with the maximum being evaluated over all spatiotemporal scales and levels of granularities,

such as molecules, proteins, subcellular organelles, single neurons, large ensembles of them, the environment the brain interacts with, and so on.

It is the irreducible Whole that forms my conscious experience, not the underlying neurons. So not only is my experience not my brain, but most certainly it is not my individual neurons. While a handful of cultured neurons in a dish may have an itsy-bitsy amount of experience, forming a mini-mind, the hundreds of millions neurons making up my posterior cortex do not embody a collection of millions of mini-minds. There is only one mind, my mind, constituted by the Whole in my brain.

Other Wholes may exist in my brain, or my body, as long as they don't share elements with the posterior hot zone Whole. Thus, it may feel like something to be my liver, but given the very limited interactions among liver cells, I doubt it feels like a lot.

The exclusion principle also explains why consciousness ceases during slow sleep. At this time, delta waves dominate the EEG and cortical neurons have regular hyperpolarized down-states during which they are silent, interspersed by active

up-states when neurons are more depolarized. These on- and off-periods are regionally coordinated. As a consequence, the cortical Whole breaks down, shattering into small cliques of interacting neurons. Each one probably has only a whit of integrated information. Effectively, “my” consciousness vanishes in deep sleep, replaced by myriad of tiny Wholes, none of which is remembered upon awakening.

The exclusion postulate also dictates whether or not an aggregate of conscious entities — ants in a colony, cells making up a tree, bees in a hive, starlings in a murmurating flock, an octopus with its eight semiautonomous arms, or the hundreds of Chinese dancers and musicians during the choreographed opening ceremony of the 2008 Olympic games in Beijing — exist as conscious entities. A herd of buffalo during a stampede or a crowd can act as if it had “one mind,” but this remains a mere figure of speech unless there is a phenomenal entity that feels like something above and beyond the experiences of the individuals making up the group. Per IIT, this would require the

extinction of the individual Wholes, as the integrated information for each of them is less than the Φ^{max} of the Whole. Everybody in the crowd would give up his or her individual consciousness to the mind of the group, like being assimilated into the hive mind of the Borg in the “Star Trek” universe.

IIT’s exclusion postulate does not permit the simultaneous existence of both individual and group mind. Thus, the *Anima Mundi* or world soul is ruled out, as it requires that the mind of all sentient beings be extinguished in favor of the all-encompassing soul. Likewise, it does not feel like anything to be the three hundred million citizens of the United States of America. As an entity, the United States has considerable extrinsic causal powers, such as the power to execute its citizens or start a war. But the country does not have maximally irreducible intrinsic cause-effect power. Countries, corporations, and other group agents exist as powerful military, economic, financial, legal, and cultural entities. They are aggregates but not Wholes. They have no phenomenal reality and no intrinsic causal power.

Thus, per IIT, single cells may have some intrinsic

existence, but this does not necessarily hold for the microbiome or trees. Animals and people exist for themselves, but herds and crowds do not. Maybe even atoms exist for themselves, but certainly not spoons, chairs, dunes, or the universe at large.

IIT posits two sides to every Whole: an exterior aspect, known to the world and interacting with other objects, including other Wholes; and an interior aspect, what it feels like, its experience. It is a solitary existence, with no direct windows into the interior of other Wholes. Two or more Wholes can fuse to give rise to a larger Whole but at the cost of losing their previous identity.

Finally, panpsychism has nothing intelligible to say about consciousness in machines. But IIT does. Conventional digital computers, built out of circuit components with sparse connectivity and little overlap among their inputs and their outputs, do not constitute a Whole. Computers have only a tiny amount of highly fragmented intrinsic cause-effect power, no matter what software they are executing and no matter their computational power. Androids, if their physical circuitry is anything like today's

CPUs, cannot dream of electric sheep. It is, of course, possible to build computing machinery that closely mimics neuronal architectures. Such neuromorphic engineering artifacts could have lots of integrated information. But we are far from those.

IIT can be thought of as an extension of physics to the central fact of our lives — consciousness. Textbook physics deals with the interaction of objects with each other, dictated by extrinsic causal powers. My and your experiences are the way brains with irreducible intrinsic causal powers feel like from the inside.

IIT offers a principled, coherent, testable, and elegant account of the relationship between these two seemingly disparate domains of existence — the physical and the mental — grounded in extrinsic and intrinsic causal powers. Causal power of two different kinds is the only sort of stuff needed to explain everything in the universe. These powers constitute ultimate reality.

Further experimental work will be essential to validate, modify, or perhaps even reject these views. If history is any guide, future discoveries in

laboratories and clinics, or perhaps off-planet, will surprise us.

We have come to the end of our voyage. Illuminated by the light of our pole star — consciousness — the universe reveals itself to be an orderly place. It is far more enminded than modernity, blinded by its technological supremacy over the natural world, takes it to be. It is a view more in line with earlier traditions that respected and feared the natural world.

Experience is in unexpected places, including in all animals, large and small, and perhaps even in brute matter itself. But consciousness is not in digital computers running software, even when they speak in tongues. Ever-more powerful machines will trade in fake consciousness, which will, perhaps, fool most. But precisely because of the looming confrontation between natural, evolved and artificial, engineered intelligence, it is absolutely essential to assert the central role of feeling to a lived life.
