



See Inside

Is Consciousness Universal?

Panpsychism, the ancient doctrine that consciousness is universal, offers some lessons in how to think about subjective experience today

By Christof Koch | Wednesday, January 1, 2014 | 36 comments

For every inside there is an outside, and for every outside there is an inside; though they are different, they go together.

—Alan Watts, *Man, Nature, and the Nature of Man*, 1991

I grew up in a devout and practicing Roman Catholic family with Purzel, a fearless and high-energy dachshund. He, as with all the other, much larger dogs that subsequently accompanied me through life, showed plenty of affection, curiosity, playfulness, aggression, anger, shame and fear. Yet my church teaches that whereas animals, as God's creatures, ought to be treated well, they do not possess an immortal soul. Only humans do. Even as a child, to me this belief felt intuitively wrong. These gorgeous creatures had feelings, just like I did. Why deny them? Why would God resurrect people but not dogs? This core Christian belief in human exceptionalism did not make any sense to me. Whatever consciousness and mind are and no matter how they relate to the brain and the rest of the body, I felt that the same principle must hold for people and dogs and, by extension, for other animals as well.



Pin it

Image: GALLERY STOCK (dog)

It was only later, at university, that I became acquainted with Buddhism and its emphasis on the universal nature of mind. Indeed, when I spent a week with His Holiness the Dalai Lama earlier in 2013 [see “The Brain of Buddha,” *Consciousness Redux*; *Scientific American Mind*, July/August 2013], I noted how often he talked about the need to reduce the suffering of “all living beings” and not just “all people.” My readings in philosophy brought me to panpsychism, the view that mind (*psyche*) is found everywhere (*pan*). Panpsychism is one of the oldest of all philosophical doctrines extant and was put forth by the ancient Greeks, in particular Thales of Miletus and Plato. Philosopher Baruch Spinoza and mathematician and universal genius Gottfried Wilhelm Leibniz, who laid down the intellectual foundations for the Age of Enlightenment, argued for panpsychism, as did philosopher Arthur Schopenhauer, father of American psychology William James, and Jesuit paleontologist Teilhard de Chardin. It declined in popularity with the rise of positivism in the 20th century.

As a natural scientist, I find a version of panpsychism modified for the 21st century to be the single most elegant and parsimonious explanation for the universe I find myself in. There are three broad reasons why panpsychism is appealing to the modern mind.

We Are All Nature's Children

The past two centuries of scientific progress have made it difficult to sustain a belief in human exceptionalism.

Consider my Bernese mountain dog, Ruby, when she yelps, whines, gnaws at her paw, limps and then comes to me, seeking aid: I infer that she is in pain because under similar conditions I behave in similar ways (sans gnawing). Physiological measures of pain confirm this inference—injured dogs, just like people, experience an elevated heart rate and blood pressure and release stress hormones into

their bloodstream. I'm not saying that a dog's pain is exactly like human pain, but dogs—as well as other animals—not only react to noxious stimuli but also consciously experience pain.

All species—bees, octopuses, ravens, crows, magpies, parrots, tuna, mice, whales, dogs, cats and monkeys—are capable of sophisticated, learned, nonstereotyped behaviors that would be associated with consciousness if a human were to carry out such actions. Precursors of behaviors thought to be unique to people are found in many species. For instance, bees are capable of recognizing specific faces from photographs, can communicate the location and quality of food sources to their sisters via the waggle dance, and can navigate complex mazes with the help of cues they store in short-term memory (for instance, “after arriving at a fork, take the exit marked by the color at the entrance”). Bees can fly several kilometers and return to their hive, a remarkable navigational performance. And a scent blown into the hive can trigger a return to the site where the bees previously encountered this odor. This type of associative memory was famously described by Marcel Proust in *À la Recherche du Temps Perdu*. Other animals can recognize themselves, know when their conspecifics observe them, and can lie and cheat.

Some people point to language and the associated benefits as being the unique defining feature of consciousness. Conveniently, this viewpoint rules out all but one species, *Homo sapiens* (which has an ineradicable desire to come out on top), as having sentience. Yet there is little reason to deny consciousness to animals, preverbal infants [see “The Conscious Infant,” *Consciousness Redux*; *Scientific American Mind*, September/October 2013] or patients with severe aphasia, all of whom are mute.

None other than Charles Darwin, in the last book he published, in the year preceding his death, set out to learn how far earthworms “acted consciously and how much mental power they displayed.” Studying their feeding and sexual behaviors for several decades—Darwin was after all a naturalist with uncanny powers of observation—he concluded that there was no absolute threshold between lower and higher animals, including humans, that assigned higher mental powers to one but not to the other.

The nervous systems of all these creatures are highly complex. Their constitutive proteins, genes, synapses, cells and neuronal circuits are as sophisticated, variegated and specialized as anything seen in the human brain. It is difficult to find anything exceptional about the human brain. Even its size is not so special, because elephants, dolphins and whales have bigger brains. Only an expert neuroanatomist, armed with a microscope, can tell a grain-size piece of cortex of a mouse from that of a monkey or a human. Biologists emphasize this structural and behavioral continuity by distinguishing between *nonhuman* and *human* animals. We are all nature's children.

Given the lack of a clear and compelling Rubicon separating simple from complex animals and simple from complex behaviors, the belief that only humans are capable of experiencing anything consciously seems preposterous. A much more reasonable assumption is that until proved otherwise, many, if not all, multicellular organisms experience pain and pleasure and can see and hear the sights and sounds of life. For brains that are smaller and less complex, the creatures' conscious experience is very likely to be less nuanced, less differentiated and more elemental. Even a worm has perhaps the vaguest sense of being alive. Of course, each species has its own unique sensorium, matched to its ecological niche. Not every creature has ears to hear and eyes to see. Yet all are capable of having at least some subjective feelings.

The Austere Appeal of Panpsychism

Taken literally, panpsychism is the belief that everything is “enminded.” All of it. Whether it is a brain, a tree, a rock or an electron. Everything that is physical also possesses an interior mental aspect. One is objective—accessible to everybody—and the other phenomenal—accessible only to the subject. That is the sense of the quotation by British-born Buddhist scholar Alan Watts with which I began this essay.

I will defend a narrowed, more nuanced view: namely that any complex system, as defined below, has the basic attributes of mind and has a minimal amount of consciousness in the sense that it feels like *something* to be that system. If the system falls apart, consciousness ceases to be; it doesn't feel like anything to be a broken system. And the more complex the system, the larger the repertoire of conscious states it can experience.

My subjective experience (and yours, too, presumably), the Cartesian “I think, therefore I am,” is an undeniable certainty, one strong

enough to hold the weight of philosophy. But from whence does this experience come? Materialists invoke something they call emergentism to explain how consciousness can be absent in simple nervous systems and emerge as their complexity increases. Consider the wetness of water, its ability to maintain contact with surfaces. It is a consequence of intermolecular interactions, notably hydrogen bonding among nearby water molecules. One or two molecules of H₂O are not wet, but put gazillions together at the right temperature and pressure, and wetness emerges. Or see how the laws of heredity emerge from the molecular properties of DNA, RNA and proteins. By the same process, mind is supposed to arise out of sufficiently complex brains.

Yet the mental is too radically different for it to arise gradually from the physical. This emergence of subjective feelings from physical stuff appears inconceivable and is at odds with a basic precept of physical thinking, the Ur-conservation law—*ex nihilo nihil fit*. So if there is nothing there in the first place, adding a little bit more won't make something. If a small brain won't be able to feel pain, why should a large brain be able to feel the god-awfulness of a throbbing toothache? Why should adding some neurons give rise to this ineffable feeling? The phenomenal hails from a kingdom other than the physical and is subject to different laws. I see no way for the divide between unconscious and conscious states to be bridged by bigger brains or more complex neurons.

A more principled solution is to assume that consciousness is a basic feature of certain types of so-called complex systems (defined in some universal, mathematical manner). And that complex systems have sensation, whereas simple systems have none. This reasoning is analogous to the arguments made by savants studying electrical charge in the 18th century. Charge is not an emergent property of living things, as originally thought when electricity was discovered in the twitching muscles of frogs. There are no uncharged particles that in the aggregate produce an electrical charge. Elementary particles either have some charge, or they have none. Thus, an electron has one negative charge, a proton has one positive charge and a photon, the carrier of light, has zero charge. As far as chemistry and biology are concerned, charge is an intrinsic property of these particles. Electrical charge does not emerge from noncharged matter. It is the same, goes the logic, with consciousness. Consciousness comes with organized chunks of matter. It is immanent in the organization of the system. It is a property of complex entities and cannot be further reduced to the action of more elementary properties. We have reached the ground floor of reductionism.

Yet, as traditionally conceived, panpsychism suffers from two major flaws. One is known as the problem of aggregates. Philosopher John Searle of the University of California, Berkeley, expressed it recently: “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.” Indeed, if consciousness is everywhere, why should it not animate the iPhone, the Internet or the United States of America? Furthermore, panpsychism does not explain why a healthy brain is conscious, whereas the same brain, placed inside a blender and reduced to goo, would not be. That is, it does not explain how aggregates combine to produce specific conscious experience.

Integrated Panpsychism

These century-old arguments bring me to the conceptual framework of the integrated information theory (IIT) of psychiatrist and neuroscientist Giulio Tononi of the University of Wisconsin–Madison. It postulates that conscious experience is a fundamental aspect of reality and is identical to a particular type of information—integrated information. Consciousness depends on a physical substrate but is not reducible to it. That is, my experience of seeing an aquamarine blue is inexorably linked to my brain but is different from my brain.

Any system that possesses some nonzero amount of integrated information experiences something. Let me repeat: any system that has even one bit of integrated information has a very minute conscious experience.

IIT makes two principled assumptions. First, conscious states are highly differentiated; they are informationally very rich. You can be conscious of an uncountable number of things. Think of all the frames from all the movies that you have ever seen or that have ever been filmed or that will be filmed! Each frame, each view, is a specific conscious percept.

Second, each such experience is highly integrated. You cannot force yourself to see the world in black and white; its color is an integrated part of your view. Whatever information you are conscious of is wholly and completely presented to your mind; it cannot be subdivided. Underlying this unity of consciousness is a multitude of causal interactions among the relevant parts of your brain. If parts of the brain become fragmented and balkanized, as occurs in deep sleep or in anesthesia, consciousness fades.

To be conscious, then, you need to be a single, integrated entity with a large repertoire of highly differentiated states. Even if the hard disk on my laptop exceeds in capacity my lifetime memories, none of its information is integrated. The family photos on my Mac are not linked to one another. The computer does not know that the boy in those pictures is my son as he matures from a toddler to an awkward teenager and then a graceful adult. To my computer, all information is equally meaningless, just a vast, random tapestry of 0s and 1s. Yet I derive meaning from these images because my memories are heavily cross-linked. And the more interconnected, the more meaningful they become.

These ideas can be precisely expressed in the language of mathematics using notions from information theory such as entropy. Given a particular brain, with its neurons in a particular state—these neurons are firing while those ones are quiet—one can precisely compute the extent to which this network is integrated. From this calculation, the theory derives a single number, Φ (pronounced “fi”) [see “A Theory of Consciousness,” *Consciousness Redux*; *Scientific American Mind*, July/August 2009]. Measured in bits, Φ denotes the size of the conscious repertoire associated with the network of causally interacting parts being in one particular state. Think of Φ as the synergy of the system. The more integrated the system is, the more synergy it has and the more conscious it is. If individual brain regions are too isolated from one another or are interconnected at random, Φ will be low. If the organism has many neurons and is richly endowed with synaptic connections, Φ will be high. Basically, Φ captures the quantity of consciousness. The quality of any one experience—the way in which red feels different from blue and a color is perceived differently from a tone—is conveyed by the informational geometry associated with Φ . The theory assigns to any one brain state a shape, a crystal, in a fantastically high-dimensional qualia space. This crystal is the system viewed from within. It is the voice in the head, the light inside the skull. It is everything you will ever know of the world. It is your only reality. It is the quiddity of experience. The dream of the lotus eater, the mindfulness of the meditating monk and the agony of the cancer patient all feel the way they do because of the shape of the distinct crystals in a space of a trillion dimensions—truly a beatific vision. The water of integrated information is turned into the wine of experience.

Integrated information makes very specific predictions about which brain circuits are involved in consciousness and which ones are peripheral players (even though they might contain many more neurons, their anatomical wiring differs). The theory has most recently been used to build a consciousness meter to assess, in a quantitative manner, the extent to which anesthetized subjects or severely brain-injured patients, such as Terri Schiavo, who died in Florida in 2005, are truly not conscious or do have some conscious experiences but are unable to signal their pain and discomfort to their loved ones [see “A Consciousness Meter,” *Consciousness Redux*; *Scientific American Mind*, March/April 2013].

IIT addresses the problem of aggregates by postulating that only “local maxima” of integrated information exist (over elements and spatial and temporal scales): my consciousness, your consciousness, but nothing in between. That is, every person living in the U.S. is, self by self, conscious, but there is no superordinate consciousness of the U.S. population as a whole.

Unlike classical panpsychism, not all physical objects have a Φ ; that is different from zero. Only integrated systems do. A bunch of disconnected neurons in a dish, a heap of sand, a galaxy of stars or a black hole—none of them are integrated. They have no consciousness. They do not have mental properties.

Last, IIT does not discriminate between squishy brains inside skulls and silicon circuits encased in titanium. Provided that the causal relations among the circuit elements, transistors and other logic gates give rise to integrated information, the system will feel like something. Consider humankind's largest and most complex artifact, the Internet. It consists of billions of computers linked together using optical fibers and copper cables that rapidly instantiate specific connections using ultrafast communication protocols. Each of these processors in turn is made out of a few billion transistors. Taken as a whole, the Internet has perhaps 10^{19} transistors, about the number of synapses in the brains of 10,000 people. Thus, its sheer number of components exceeds that of any one human brain. Whether or not the Internet today feels like something to itself is completely speculative. Still, it is certainly conceivable.

When I talk and write about panpsychism, I often encounter blank stares of incomprehension. Such a belief violates people's strongly held intuition that sentience is something only humans and a few closely related species possess. Yet our intuition also fails when we are first told as kids that a whale is not a fish but a mammal or that people on the other side of the planet do not fall off because they are upside down. Panpsychism is an elegant explanation for the most basic of all brute facts I encounter every morning on awakening: there

is subjective experience. Tononi's theory offers a scientific, constructive, predictive and mathematically precise form of panpsychism for the 21st century. It is a gigantic step in the final resolution of the ancient mind-body problem.

Further Reading

Panpsychism in the West. David Skrbina. MIT Press, 2005.

Consciousness: Confessions of a Romantic Reductionist. Christof Koch. MIT Press, 2012.

Integrated Information Theory of Consciousness: An Updated Account. Giulio Tononi in *Archives Italiennes de Biologie*, Vol. 150, No. 4, pages 293–329; December 2012.

YES! Send me a free issue of Scientific American with no obligation to continue the subscription. If I like it, I will be billed for the one-year subscription.



Email Address

Name