The Emperor's New Mind

What’s in it for me? Travel to the end of the universe to understand the depths of the human mind.

Do computers have minds? AI fans have been claiming that they do – or they soon will – since at least 1989; that’s when mathematical physicist Roger Penrose first presented his argument against the idea that intelligent computers can be conscious. These blinks venture through the most important ideas in mathematics, computation, physics, psychology, and philosophy to build a fascinating argument for the complexity of the human mind. From Turing machines to relativity theory to split-brain experiments, these blinks make a compelling case for the enduring miracle of our universe and the mystery of our own consciousness. Along the way you’ll learn

why math is real; why time is an illusion; and why quantum physics underwrites consciousness.

Whether computers can have minds is a question of whether the human mind is computable.

Back in 1950, famous British computer scientist Alan Turing proposed a test to measure computer intelligence. Put simply, a machine passes the test if a human interacting with it can’t tell that he’s interacting with a machine. For example, a human interrogator could be chatting with a digital computer over text, trying to determine whether he’s talking to a computer or another human. Some computers can indeed imitate human conversation well enough to pass such a test. But does that mean that they’ve learned to “think” in the same way we do? Here’s the key message: Whether computers can have minds is a question of whether the human mind is computable. To proponents of a viewpoint known as strong AI, a computer behaving in a human, intelligent way is evidence that it possesses real, human intelligence. According to this view, even a thermostat possesses some kind of “mind” – albeit a very simple one. The author, however, thinks that our minds are fundamentally non-computable. To understand the depth of his argument, we’ll have to travel to the edge of the universe and back again. But first, let’s have a look at what “computability” actually means. If a problem is computable, it means that it can be solved through an effective computational program, using an algorithm. An algorithm is a sequence of step-by-step instructions that tell a computer what to do. Pioneering computer scientist Alan Turing was the first to devise a hypothetical model for running such algorithms. He imagined a scanner-like device running over an infinite strip of tape with squares inscribed with 0’s and 1’s. The device’s “state” changes with each number it scans. And the device also has the power to change the numbers on the tape. Which action it takes – whether it moves left or right, deletes or changes a number – depends on the initial number on the square as well as the state of the device. Turing showed that even complex algorithmic problems could be solved by this machine. Even though the Turing machine is a mathematical idealization, it gives us a useful measure of computability. Any operation that can be run by a Turing machine is algorithmic. In fact, all of our modern computers are essentially

Turing machines. But even Turing recognized that some problems can't be solved algorithmically. As it turns out, even some mathematical operations are not actually computable. In the next blink, we'll try to understand why.

The belief that math exists as an external reality stems from mathematical discoveries.

Is math just a numbers game that humans invented to keep busy? Many philosophers – and even some mathematicians – think so. But the author belongs to another school of thought, the Platonist view. Platonists believe that mathematics is firmly rooted in reality. One argument in their favor is that most mathematical ideas come about more like discoveries than inventions. The key message here is: The belief that math exists as an external reality stems from mathematical discoveries. While real numbers are the numbers we use in everyday life for balancing our bank accounts, measuring distances, or keeping time, more complicated mathematical operations go beyond this system. At one point, mathematicians realized that it would be useful to extract the square root of negative numbers – something not permitted with real numbers. So they created the imaginary number i. i is the square root of -1. From this imaginary number, a whole new system emerged: complex numbers. A complex number has the form a + ib, where a and b are real numbers and i is the imaginary number. Complex numbers have led to important and even beautiful discoveries like the Mandelbrot set, named after mathematician Benoit Mandelbrot. The Mandelbrot set is a group of complex numbers in which a specially defined sequence of mathematical functions is bounded. This means that if you map these functions on a graph, they never venture beyond a fixed boundary. However, when approaching this boundary, the shape of the functions gets infinitely more elaborate, revealing ever-finer recursive detail. Imagine a flower that, when you zoom in on it, is composed of ever-smaller flowers. Mathematicians had no idea about the magical properties of complex numbers until Mandelbrot discovered them. It wasn't like he invented that set of numbers; these properties were just there, waiting to be found – a powerful argument for mathematical Platonism. Another piece of evidence for the idea that math is rooted in reality comes from the brilliant logician Kurt Gödel. In the 1930s, Gödel showed that every logical system relies on some statements that can't be proven or disproven using the rules of the system. That means that even mathematical systems rely on some fundamental assumptions that have to be taken for granted. To the author, Gödel's incompleteness theorem shows that there's some "God-given" truth to math that we can't capture through logic alone. This could explain why purely logical systems like algorithms fail to capture all of reality – and even all of math. An example? No algorithm to date can plot the Mandelbrot set in its infinite detail.

The classical theories of physics do a marvelous job of explaining the world.

Long before the advent of modern science, the ancient Greeks came up with some pretty good theories about the geometry of physical objects. But it wasn't until Galileo started thinking about gravity and energy in the seventeenth century that we really began to understand the principles that govern our world. A little later, Isaac Newton worked Galileo's ideas into three essential laws of motion. The first one says that an object will remain at rest or continue in motion unless it's acted on by an external force. The second says that an object's change in motion is proportional to the external force that acts on it. And the third law of motion says that the forces two objects exert on each other are always equal. The key message is this: The classical theories of physics do a marvelous job of explaining the world. Isaac Newton's 1687 masterpiece Philosophiae Naturalis Principia Mathematica cemented the idea that, with a few basic mathematical principles, we could predict how things in the real world will behave. From this basic scheme, all other major theories of classical physics developed. In the nineteenth century, James Clerk Maxwell presented a set of equations that underlie the behavior of electric and magnetic fields, as well as light. These Maxwell equations were hugely important in the development of modern technologies such as radio, electric motors, and wireless communication. Moreover, it was Maxwell's assertion that the speed of light is fixed that led Einstein to develop his theory of special relativity. Einstein showed that if the speed of light is fixed, then our measures of space and time must actually be relative. Or, to put it the other way around, our experience of distance and time is relative to where in the universe we find ourselves, and how fast we're moving. To illustrate this, imagine twin brothers. One twin boards a spaceship to travel to a distant star at near-light speed, while the other remains on Earth. According to relativity theory, the space-traveling twin would still be youthful upon his return to Earth, while his brother would already be an old man. Einstein later expanded this into his theory of general relativity, which, on top of acceleration, also takes into account the effects of gravity on space-time. All of these theories have greatly contributed to our understanding of the universe – but, as we'll see in the next blink, they lead to a pretty rigid worldview.

Classical physics suggests a deterministic universe.

We can safely categorize the tried-and-tested theories of classical physics as superb. A superb theory of physics explains a lot, and does so accurately. It also has a certain beauty and simplicity. Take, for example, Newton's three laws of motion, which elegantly explain how objects behave here on Earth, and predict the motions of distant stars with relative accuracy – an accuracy upon which Einstein's relativity theories improved. Moreover, both Newton's and Einstein's ideas have been verified by our observations over and over again. Most contemporary theories of physics have not quite reached the superb level. Some of them are tentative at best. Others prevail because they're useful for our understanding of the world, even if they can't be verified – such as the big bang theory. The key message here is: Classical physics suggests a deterministic universe. In due time, physicists might find counterevidence to many of the new theories, or come up with more elegant explanations for why things are the way they are. But the classical theories seem to be here to stay – and they present us with a pretty clear view of the world. Let's briefly recap what classical physics has taught us. For one, it gave us the concept of spacetime, which is the multidimensional arena in which all of physics plays out. In this arena, physical objects behave according to precise mathematical laws. The objects include particles, but also fields, such as electromagnetic or gravitational fields. Classical physics teaches us that if we know the mass, position, and velocity of any physical object at any given time, we can determine its mass, position, and velocity for all later times. It seems that anything that happens in the future is completely fixed by the past. This idea lies at the core of the philosophy of determinism. All superb theories of classical physics lead us to a deterministic worldview. The implications for the human mind are quite depressing. How can we have free will if everything that happens is already fixed by simple physical reactions? It's also easy to assume that if our brains behave according to a few simple, physical principles, it shouldn't be a problem to simulate this behavior with wires and electrodes. But it's important to note that even if the human mind is completely deterministic, that still doesn't mean it's computable. For example, one can imagine a completely deterministic world that is still complex enough to be functionally non-computable. And luckily, determinism isn't a done deal. Because since the 1920s, another area of physics has been rocking the foundations of our classical worldview.

Quantum mechanics is characterized by uncertainty, indeterminism, and mystery – and it completely changed our worldview.

For a while, it seemed that the classical theories of physics, such as Newton's laws of motion, had the power to explain the whole universe. But when physicists began observing the behavior of molecules, atoms, and subatomic particles, they were shocked. These small particles didn't behave in a "classical" way at all. Take, for instance, protons, photons, and electrons, which often change their position and motion in completely counterintuitive ways. Sometimes they even seem to exist in two places at once! This random movement underlies many properties of physical materials, as well as physical processes such as freezing and boiling. And so, around 1925, physicists had to come up with a new set of theories to explain the strange behavior of the particles that make up our world. Here's the key message: Quantum mechanics is characterized by uncertainty, indeterminism, and mystery – and it completely changed our worldview. One of the most famous experiments in quantum physics is the so-called double-slit experiment. In this experiment, quantum particles called photons are fired through a wall with two narrow slits and onto a screen behind. Even though they're considered particles, the photons behave like waves when they pass through the slits. Not only do they deflect in random directions, but they also cancel themselves out in some areas, while enhancing each other in other areas. Because of this, they hit the screen in a characteristic striped interference pattern. The strange thing is that the wave-like behavior doesn't seem to be an effect of multiple photons working together. Instead, each individual photon seems to behave like a wave all on its own. When only one slit is opened, the photon travels through like a tennis ball you'd throw through a hole in the wall. But when both slits are open, the photon seems to pass through both slits, and then interfere with itself, as indicated by the interference pattern. Even stranger is that as soon as scientists try to monitor the paths more closely, the photon suddenly behaves "normally," passing either through one slit or the other. The implications of the double-slit experiment are head-scratching. In the quantum world, different alternatives seem to coexist simultaneously, and particles can be in two places at once. And on top of that, our measurement seems to affect how the particles behave! It seems that the determinism that classical physics has led us to doesn't hold at the micro-level. As we'll see, this leaves us with many confusing questions – but also some amazing possibilities.

We still don’t understand how quantum physics and classical physics work together.

To show again why quantum theory is so confounding, let’s consider the paradox of Schrödinger’s cat. This is a thought experiment proposed by Erwin Schrödinger to Albert Einstein in 1935, which we’ll alter here slightly for our purposes. Imagine a box with a cat inside, constructed so that no outside influence can pass through. There’s a device in the box that can be triggered by a single quantum event – let’s say by a photon hitting a photocell. When the device is triggered, it smashes a bottle with cyanide that kills the cat. So how can we know whether the cat is dead? The key message is this: We still don’t understand how quantum physics and classical physics work together. Recall that in the quantum world, different alternatives can coexist, only manifesting into a clear outcome when we observe them. This means that as long as the box remains unopened, the photon has both triggered and not triggered the device. Consequently, the cat is both dead and alive. With this thought experiment, Schrödinger wanted to show that the indeterminism of quantum mechanics should not apply to big physical objects like cats. At the quantum level, it’s quite possible that different alternatives coexist in a web of quantum superpositions. But in the “real” world, at the macro level, one alternative always wins out: cats are either dead or alive. Scientists have come up with various mathematical methods to define uncertain quantum states. The two most important of these are R and U. R is the vector that describes the quantum position of a single quantum particle – this is where the indeterminism comes into play, because particles can essentially be anywhere at once. U stands for Unitary transformation, and describes how a quantum system changes with time by relating the state of the system to the energy in it. Scientists assign number weightings to the different possible alternatives in a quantum system, and this gives us a probabilistic picture of how the system will behave. There’s still lots of debate about how R and U work together. The author thinks that solving this mystery could bring us a lot closer to understanding our universe, our minds, and the flow of time. For example, he believes that R is fundamentally time-asymmetric, meaning the calculation only works in one direction of time. This is important, because all classical theories of physics are time-symmetric. In principle, there’s nothing in them that prevents us from using them in the reverse direction of time. By understanding R, the author believes, we might finally be able to solve the mystery of time.

Our brain’s design is much more complex than that of a computer.

What does this all have to do with the human mind? Well, it means that it’s very likely that our minds don’t operate in the strict, deterministic way that the classical worldview suggests. Indeed, the author believes that quantum mechanics plays a crucial role in the way we think. Let’s take a closer look at how our brain works. The human brain is an organ of incredibly complex design. On the most basic level, there’s a large inner region of white matter that sorts and relays signals, and a thin outer surface of gray matter that processes them. The outer surface is called the cerebral cortex, and it’s where higher cognition and complex computation take place. It’s much thicker in humans than in other animals. The key message here is: Our brain's design is much more complex than that of a computer. Different parts of the cortex take on different tasks. For example, the visual cortex at the back of the brain processes what we see. Other senses also have designated areas in the brain. Our sensory organs pick up signals from the outside world, and nerve cells relay them to the cortex for processing. Finally, all the information is combined in the frontal lobes, the part of our brains that makes plans and executes them. For example, the frontal lobes can send out signals to our muscles to get us moving. The brain passes along signals through specialized nerve cells called neurons. When a strong-enough signal reaches a neuron, it becomes electrically charged. This electrical charge runs along the length of the neuron until it reaches the synapse, a small gap connecting the first neuron to the next one. The first neuron then releases certain chemicals into the synapse that either excite or inhibit the next neuron. On an abstract level, this isn't so different from a digital computer: there's an input signal, some information processing by small, connected units, and finally an output signal. And neurons transmit signals through an all-or-nothing principle: they either fire or they don't. This is similar to electrodes in a digital computer: there's either an electrical pulse in a wire or there isn't. So, in principle, it would be possible to build a computer out of neurons. But what about the other way around? This is where it becomes tricky. Neurons can have hundreds of thousands of synaptic connections, which are connected in a much more random and redundant way than most electronic circuits. And these synaptic connections seem to be constantly changing. This is what allows for our brain's plasticity – our brain can change within seconds, based on our actions and experiences. Ultimately and mysteriously, these myriad changing connections give rise to a single consciousness. How can this be?

Quantum physics may play an important role in human consciousness.

We know a great deal about the structure and function of the brain. What we don't know is how all these different, complicated, parallel processes give rise to our consciousness. This is where quantum physics could come into play. There's at least one spot where quantum effects are directly involved: the retina. Under the right conditions, a single photon hitting our retina is enough to trigger a nerve signal. It takes about seven photons for us to become aware of them. This means that there are at least a couple of nerve cells – namely those in the retina – that can be triggered by quantum events. Couldn't this be true for other neurons in our brain? Here's the key message: Quantum physics may play an important role in human consciousness. If the brain is indeed composed of neurons that can be triggered by single-quantum events, what does that mean for the way it operates? Well, for one, it means that there's much more indeterminism, uncertainty, and mystery at work than proponents of strong AI would like to believe. Perhaps the many parallel activities of our brain can be linked to the parallel alternatives that coexist in the quantum universe – alternatives that only resolve when the quantum state of a particle is observed. Perhaps this act of observation is what we perceive as consciousness. Conscious thinking, at least, seems to be all about resolving alternatives that previously ran in parallel. This seems to happen in a decidedly non-algorithmic fashion – perhaps with the indeterministic nature of quantum particles playing an important role. This would explain why mathematicians have a sense for truth in mathematics even for ideas that they can't prove algorithmically.

Many famous mathematicians have described how great insights came to them in a flash with a feeling of complete certainty. The author believes that in such moments, the mind comes in contact with some external Platonic reality. Like the author, many mathematicians also describe their thought processes as non-verbal, and instead as visual, geometric, or completely abstract. While quantum computers might be able to run many processes in parallel, there’s no evidence that they can achieve the kind of “oneness” that human consciousness exhibits. Conscious intelligence allows us to take into account our thoughts, senses, and past experiences to make informed judgments about new and unexpected situations. In this way, consciousness gives rise to a type of intelligence that computers will probably never achieve.

Final summary

The key message in these blinks: Fans of AI have long claimed that we can program computers to think like us. But there’s much more mystery at work in our brains than we can currently account for. Classical physics has given us a deterministic picture of the world. But more recently, quantum physics has shown that the subatomic level of the world is characterized by indeterminism and uncertainty. Such quantum processes could play a crucial role in the formation of our consciousness. Until we fully understand those processes, it’s unlikely that we’ll be able to program computers to possess human intelligence. Got feedback? We’d love to hear what you think about our content! Just drop an email to with The Emperor's New Mind as the subject line, and share your thoughts!