

THE CONSCIOUSNESS WAGER

In 1998, two top thinkers made a bet about whether the hardest problem in neuroscience would be cracked by 2023. With five years to go, Per Snaprud tracks down the pair to find out how close we've got

TWENTY years ago this week, two young men sat in a smoky bar in Bremen, northern Germany. Neuroscientist Christof Koch and philosopher David Chalmers had spent the day lecturing at a conference about consciousness, and they still had more to say. After a few drinks, Koch suggested a wager. He bet a case of fine wine that within the next 25 years someone would discover a specific signature of consciousness in the brain. Chalmers said it wouldn't happen, and bet against.

It was a bit of fun, but also an audacious gamble. Consciousness is truly mysterious. It is the essence of you – the redness of red, the feeling of being in love, the sensation of pain and all the rest of your subjective experiences, conjured up somehow by your brain. Back then, its illusive nature meant that many believed it wasn't even a valid subject for scientific investigation.

Today, consciousness is a hot research area, and Koch and Chalmers are two of its most influential figures. Koch is head of the Allen Institute for Brain Science in Seattle. Chalmers is a professor at New York University and famous for coining the phrase the “hard problem” to distinguish the difficulty of understanding consciousness from that of grasping other mental phenomena. Much

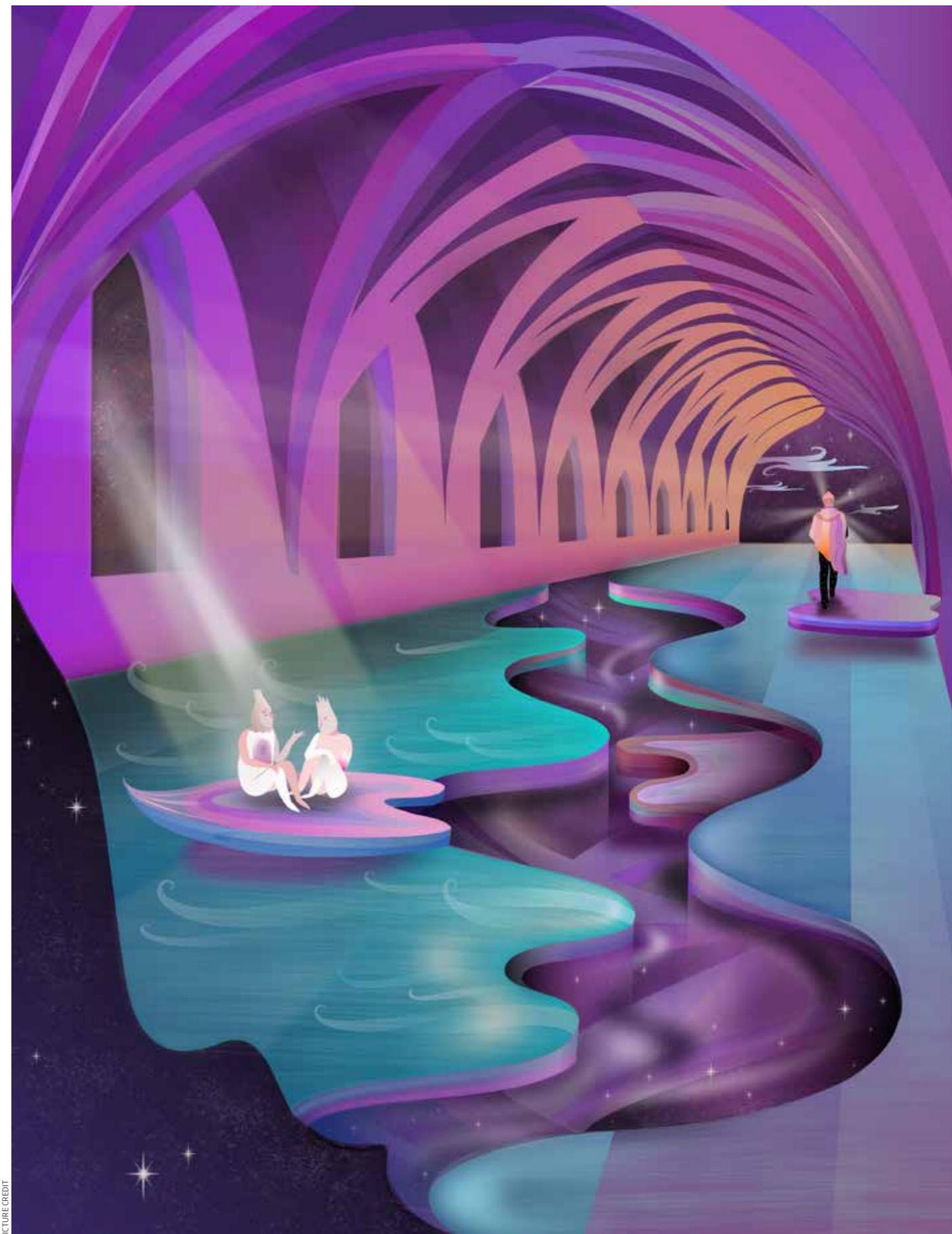
progress has been made, but how close are we to solving the mystery? To find out, I decided to ask Chalmers and Koch how their bet was going. But there was a problem – they had mislaid the terms of the wager. Luckily, I too was in Bremen 20 years ago and was able to come to their rescue.

The consciousness bet has its roots in Koch's research. In the mid-1980s, as a young scientist, he began collaborating with Francis Crick who had co-discovered the double helix structure of DNA. Both men were frustrated that science had so little to say about consciousness. Indeed, the *International Dictionary of Psychology* described it thus: “a fascinating but elusive phenomenon; it is impossible to specify what it is, what it does, or why it evolved. Nothing worth reading has been written about it.” The pair believed this was partly because researchers lacked a practical approach to the problem. In his work on DNA, Crick had reduced the mystery of biological heritability to a few intrinsic properties of a small set of molecules. He and Koch thought consciousness might be explained using a similar approach. Leaving aside the tricky issue of what causes consciousness, they wanted to find a minimal physical signature in the brain sufficient for a specific subjective experience. Thus began the

search for the “neural correlates of consciousness”.

This approach, which allows incremental progress and appeals to researchers no matter what their philosophical stance, has been central to the study of consciousness ever since. Indeed, neural correlates of consciousness – the subject of Koch's wager – was the topic under discussion at the Bremen conference. There, Koch argued that gamma waves might be involved, based on research linking awareness to this kind of brain activity, where neurons fire at frequencies around 40 hertz. Conference delegates also heard that pyramidal cells in the brain's outer layer or cortex might play a key role (see “Where's consciousness?”, overleaf).

There was no shortage of ideas. However, the early ones all proved too simplistic. Take the notion, which Koch and Crick favoured for a while, that a sheet-like structure beneath the cortex called the claustrum is crucial for consciousness. There was reason to be optimistic: a case study in 2014 showed that electrically stimulating this structure in a woman's brain caused her to stare blankly ahead, seemingly unconscious, until the stimulation stopped. But another study described someone who remained fully conscious after his claustrum was destroyed ►



PICTURE CREDIT

by encephalitis, undermining that idea.

Undeterred, those searching for the neural correlates of consciousness have come up with more sophisticated ideas. But are we really any closer to cracking the problem? Last year, I had the perfect opportunity to find out when I met up with Chalmers at a consciousness conference in Budapest, Hungary.

I asked him how his bet with Koch was shaping up. Looking a bit dejected, he told me that they had met three months earlier in New York, and the subject of the wager came up. “Sadly, neither of us could remember the exact terms,” he said. It was then that I realised I might be able to help. Although I wasn’t with them at the bar when it all happened, the following day I had interviewed Chalmers, who mentioned the bet he had made just hours earlier.

Back home, I started looking for notes from our long-ago meeting. Eventually, I found a cassette stowed away in a box on top of a shelf in my study. The faded label read: “David Chalmers interview”. After more searching, I found a tape recorder and pressed play.

Chalmers is describing how he left the conference banquet after midnight and continued to a bar with Koch and a few others.

“We had a good time. It got light very early and we walked back around 6 o’clock.” Even though he still hasn’t slept, his voice is remarkably alert. Distant memories re-emerge. We are in an outdoor restaurant under a hazy sky. Chalmers wears a black leather jacket and has shoulder-length hair.

Fast-forwarding through the tape I find what I’m looking for. Towards the end, Chalmers mentions the bet and specifies what kind of signature of consciousness it refers to: “a small set of neurons characterised by a small list of intrinsic properties. I think we said less than 10.” Intrinsic properties could be, say, a neuron’s pattern of electrical firing, or genes regulating the production of various neurotransmitters. “And it will be clear in 25 years,” he says. Bingo! I emailed a clip from the recording to Chalmers. He immediately replied: “Thanks – this is great!” and forwarded the message to Koch.

These days, the prevailing ideas of consciousness have more to do with the properties of networks of neurons than those of specific cells. One proposal that has been particularly influential is called global workspace theory. It suggests that information coming from the outside world competes for

attention in the cortex and a structure in the centre of the brain called the thalamus. If the signal it generates outcompetes other information, it is broadcast across the brain. Only then do you consciously register it. A popular version of this theory is that special “workspace neurons” in the cortex, primarily in the front of the brain, broadcast information through their long-range connections.

Measuring experience

Other network-based ideas suggest that consciousness is the result of information being combined so that it is more than the sum of its parts. One that has grabbed much attention is integrated information theory (IIT). It is the brainchild of neuroscientist Giulio Tononi at the University of Wisconsin, who believes that the amount of a consciousness in any system – which he calls phi – can be measured. Very approximately, phi will be high in a system of specialised modules that can interact rapidly and effectively. This is true for large parts of the brain’s cortex. In contrast, phi is low in the cerebellum, which contains 69 billion of the 86 billion nerve cells that make up the human brain, but is composed of modules that work largely independently of each other.

How do these two ideas stack up? Global workspace theory seems to fit with a lot of findings about the brain. However, it does not convince some, including philosopher Ned Block at New York University, who questions whether it explains subjective experience itself or just indicates when information is available for reasoning, speech and bodily control. IIT also reflects some observations about the brain. A stroke or tumour, for example, may destroy the cerebellum without significantly affecting consciousness, whereas similar damage to the cortex usually disrupts subjective experience, and can even cause a coma. The theory is quite controversial – not least because it posits that something inanimate like a grid of certain logic gates may have an extremely high degree of consciousness – but it also has some high-profile supporters. One of them is Koch.

This is something that Chalmers pointed out in his first message: “I’m thinking that with your current advocacy of IIT, this bet is looking pretty good for me.” Koch replied the same day in an upbeat tone, defending his allegiance to the idea. “There has been a lot of progress in the intervening years concerning the neuronal correlates of consciousness (NCC). The latest, best estimate for the NCC is a

hot-zone in the posterior cortex but, rather surprisingly and contrary to Global Workspace Theories, not in the front of the cortex.”

He attached two papers he had recently co-authored to support this last point. They link the front of the cortex with monitoring and reporting, but not subjective experience. He also pointed out that the back of the cortex has much higher phi than the front, due to the connectivity of its neurons, and is therefore more closely coupled to consciousness according to IIT. “The intrinsic properties that we spoke about in Bremen would be the intrinsic connectivity of neurons in [the] posterior cortex,” he wrote. So, the posterior hot zone is Koch’s horse in the race.

Chalmers replied a few hours later, highlighting the details of the bet. It is about finding a link between a subjective experience and a small number of nerve cells with a handful of intrinsic properties. Chalmers didn’t think that something associated with phi should count as intrinsic: “I’m thinking that phi is a network property of a group of neurons, so not an intrinsic property of specific neurons,” he wrote. Another question was whether Koch’s hot zone constituted a small group of neurons. “Overall, I agree that IIT is cool and shares something of the bold spirit of what you proposed then, but the specific idea seems different.”

The reply arrived the following day: “You are right – it is unlikely that there is a special

“A consciousness-o-meter would work by zapping the brain with magnetic pulses”

‘magical’ property inside some particular neurons,” Koch wrote. But he then proceeded to describe how intrinsic factors, such as genes, might shape connectivity in a way that makes it difficult to distinguish between intrinsic and extrinsic properties.

Although intriguing, these emails left many questions unanswered about the bet, so I arranged a Skype interview with Chalmers and Koch. I began by asking them how they see their chances of winning.

“Let’s go over what we said in Bremen,” says Koch, and gets into an exposition on the number of neurons that could be reasonably described as a “small set”. Chalmers looks amused. He then leans forward: “I think fairly crucial to Christof’s original view is that it would be a special *kind* of neuron, with some special properties. And that is what to me



EVERETT COLLECTION/INCALPANY

looks somewhat unlikely,” he says. “Well, that we simply don’t know, Dave. This is what our institute does, and we just got \$120 million to do more of it,” says Koch, referring to the Allen Institute’s efforts to map and characterise different cell types in the brain. “OK,” says Chalmers. “But ‘we don’t know’ isn’t good enough for you. You need to actually discover this by 2023 for the purposes of this bet. And that’s looking unlikely.” Koch stares into the distance for a moment, nods and then smiles: “I agree, it’s unlikely because the networks are so complex.” But he hasn’t given up all hope. “A lot can happen in five years!”

Whoever wins, Chalmers and Koch are united in their belief that this is an important quest. Would success mean that we had cracked the mystery of consciousness? “Well, there is a difference between finding a correlate and finding an explanation,” says Chalmers. Nevertheless, he hopes that neural correlates of consciousness will get us closer to answering the question of why it exists. Koch agrees: “The only thing I’d like to add is that looking for NCCs also allows us to answer a host of practical, clinical and legal questions.”

One such is how to measure the level of consciousness in brain-damaged people who lack the ability to communicate. Koch has recently launched a project that he hopes will solve this within a decade by creating a “consciousness-o-meter” to detect consciousness for instance by zapping the brain with magnetic pulses.

That would be impressive, but it is still far from solving the hard problem. I ask Koch if he ever feels that consciousness is too great a mystery to be solved by the human mind.

In the TV show *Humans*, finding the consciousness switch is easy – it’s under the chin

“No!” he answers. Chalmers is more circumspect, however. He suggests consciousness may even be something fundamental that doesn’t have an explanation. “But we won’t know until we try. And even if there are some things that remain mysteries, there is going to be a whole lot of other stuff that we *can* understand.”

Artificial intelligence could even solve the riddle for us one day, he says. It sounds like wild speculation, but he is serious. “Absolutely!” AI might eventually evolve consciousness, he says, but perhaps that’s not even essential. “If it turns out that there is some completely rational reason that there is consciousness in the universe, maybe even an unconscious system could figure that out.”

On 20 June 2023, the bet will be settled. Koch wants the occasion marked by a workshop and an official announcement of the winner. He is an optimist. But even if he were to win, finding the NCC is just the first step towards the big goal: a fundamental theory of consciousness. Chalmers hopes it will be as definitive and widely accepted as current theories in physics – but he believes it will take far longer than five years. “It’s going to be 100, 200 years. So let’s re-evaluate then.” ■

Per Snaprud is an editor at Forskning & Framsteg in Stockholm, Sweden. His new book, *Return of Consciousness*, is published by Natur & Kultur

Where’s consciousness?

Over the past two decades, there have been numerous candidates for the minimal structures or events in the brain required for a subjective experience

THREE OLD IDEAS

Gamma waves
Neurons firing in synchrony at 40 hertz

Pyramidal cells
in the neocortex

The **claustrum** on-off switch

TWO LEADING IDEAS

Global workspace theory
Specialised modules send messages into a vast network where they compete for dominance. The winner is broadcast globally, and thus enters consciousness

Integrated information theory
Phi, a measure of consciousness, is higher in systems with many interconnected modules, such as a “hot zone” at the back of the brain