New Scientist

WEEKLY November 7–13, 2020

FOCUS ON
CORONAVIRUS
Europe locks down again as second-wave cases surge
What went wrong in Germany?
How antibodies from survivors could help save lives

THE TRUE PURPOSE OF OUR DREAMS

An intriguing new theory about what goes on inside our heads at night, and why

FORGOTTEN CODES
The computer systems that underpin our lives are built on legacy software. What happens when it fails?

PLUS VIRUSES THAT MAKE ENERGY / **TRUTH SEEKERS REVIEWED** / **NUCLEAR FUSION BUT COOLER** / COMET WITH ICE LIKE CANDYFLOSS

Science and technology news **www.newscientist.com**

No3307 US$6.99 CAN$7.99

Dream power

The fictions we conjure while we sleep may do something far more profound than reinforcing learning, says neuroscientist **Erik Hoel**

IF ALIENS ever visited Earth, they might notice something strange. Nearly everyone, everywhere, spends a significant part of their day paying attention to things that aren't real. Humans often care fiercely about events that never happened, whether in TV shows, video games, novels, movies. Why care so much about fictions?

Perhaps, these aliens might hypothesise, humans are too stupid to distinguish between truth and falsehood. Or perhaps they pay attention to fake events for the same reason that they eat too much cheesecake: both are non-natural outcomes of evolved interests.

The aliens' confusion might deepen when they learned that humans fall asleep and dream. For dreams are also fictions. Dreaming takes time and energy, so presumably has an evolutionary purpose. The aliens might begin to wonder what they are missing about the importance of experiencing things that never happened.

As someone who grew up in my family's bookstore, and as a novelist, this question of the importance of fictions is especially dear to me. I think the imaginary aliens are in the same position as a scientist attempting to explain the evolved purpose of dreams – and if we can identify the biological reason for dreaming, we can ask if it applies to the artificial dreams we call fictions.

As a neuroscientist, I've been working on a hypothesis that draws on what we've learned about artificial neural networks to cast dreaming as a way to improve our performance in waking life, just not in the way we might think. If correct, it may also explain some of this strange human attraction to the unreal in our waking lives.

The study of dreams, also known as oneirology, suffered something of a false start in the first decades of the 20th century, when it was tainted by association with Sigmund Freud's ideas about psychosexual development. Freud argued that dreams are an expression of repressed desires resulting from traumatic experiences in early life.

HARDZIEJ STUDIO

These ideas have been discredited, but dream research never quite shook the association.

Luckily, over recent decades, neuroimaging and behavioural research have reinvigorated the field by giving us insight into the biological mechanisms underlying dreams. We now know that dreams are the result of localised firing of neurons that is probably induced by the brain's many feedback connections and not dependent on information from external stimuli. Dreaming represents a unique physiological state in which activity similar to that we see when we are awake is promoted while behaviour is essentially cut off by powerful chemical systems that induce paralysis.

Yet although we now know a good amount about the mechanisms of dreaming, we have little insight into its function. Some argue that we don't need to understand what dreams are for. Perhaps they are just a by-product of sleep, which may have evolved for some other reason, such as to clear the metabolic detritus generated by neuronal activity.

But this "null hypothesis" of dreams has been challenged by a slew of ideas about how dreams have an evolved purpose. After all, we spend hours every night dreaming in a distinct stage of sleep.

Making memories?

Generally, these dream hypotheses have trouble accounting for the distinct phenomenology of dreams: their unique, highly specific nature, which is what sets them apart from waking experience. Dreams are sparse, in that they mostly don't contain the vivid sensory detail of waking life. Dreams are hallucinatory, in that they contain warped concepts and perceptions that are biased or unrealistic. And dreams are narrative, in that they are fabulist versions of the kinds of events we might encounter in real life, just rendered strange.

Consider the leading hypothesis, which is that dreaming is somehow involved in the process of memory storage. This idea draws on the metaphor of the brain as a › computer: explicit memories are created and then stored, the way one encodes data on a hard drive. Neuroscience has long drawn on such metaphors, even from before it was called "neuroscience", when the metaphors were pneumatic pressures or mechanical clocks. But sometimes metaphors can lead their proposers astray. In the case of sleep and memory, it is well known that various improvements can occur after a good night's sleep, such as performance on some tasks, but it is less clear that acts of pure memorisation, like lists of numbers, are actually significantly improved.

What would it even mean to help store a memory over a night? The clearest hypothesis about memory storage and sleep is based on studies showing that memories, in the form of the specific neural sequences of firing that are seen while we are awake, are sometimes "replayed" during sleep in mammals. Perhaps dreams are just that: replays of memories.

While neurons that learn do seem to increase in their firing frequency during sleep, two facts suggest the idea falls short. The first is that replay has been more strongly associated with non-REM sleep than the REM stage, where the most intense narrative dreaming occurs. The second is that it is unclear whether memories are actually being replayed during so-called "replay." Indeed, careful studies have demonstrated that the brain more commonly produces never-before-seen patterns during these periods rather than previously seen waking patterns.

Behavioural evidence is also a problem for the idea that dreams are somehow replays of memories, or even just by-products of the integration of memories. If this were the case, we would expect to dream actual memories, yet dreaming specific previous memories is actually so rare that it is considered pathological, often a sign of post-traumatic stress disorder.

Yet there is no doubt that dreams do play a role in memory and learning. Consider how I learned to juggle. As an undergraduate, I took a class on memory and as part of my homework I was assigned to learn to juggle

"Dreams may serve as 'noise injections' to counteract the risk of learning too narrowly"

The purpose of dreaming remains one of the great mysteries of the brain

TANG MING TUNG/GETTY IMAGES

SCOTT MACBRIDE/GETTY IMAGES

in one night and then demonstrate it in front of the class. I practised all evening, tossing tennis balls helplessly, but eventually fell into bed, certain of embarrassment the next day. On waking, I immediately jumped out of bed, picked up the tennis balls, and found I could juggle perfectly. It was an incredible lesson. It seemed that something had happened in my sleep that had built on my waking experience.

Even so, I find it hard to accept that I had stored or replayed memories of my juggling during sleep. When I went to bed, I couldn't juggle. If I had replayed my failures, what would be the gain? Most importantly, it is doubtful I dreamed of precise juggling events. More probably, if I dreamed of juggling at all, it was of sparse and hallucinatory fragments.

This is backed up by studies that have had participants play games like *Tetris*, which they were novices at, and found that they reported *Tetris*-like dreams – imagine falling hallucinatory blocks – but no replays of specific *Tetris* games. It seems that the best way to get someone to actually dream

How can someone who went to sleep failing to learn to juggle wake up the next morning as a juggler?

about something is to have them learn a difficult and novel task, and then have them overtrain on it, as with playing *Tetris* for hours and hours.

Deep lessons

A new and growing trend in neuroscience might help explain why this is the case, and offer a clear explanation for why dreams possess their distinct phenomenology. This trend seeks to apply the lessons of deep learning and the study of artificial neural networks to the brain. These techniques are, after all, originally inspired by how the brain functions, and remain the only set of techniques by which machines can reach human-level cognitive performance on complex tasks.

From a deep-learning perspective, learning isn't like storing memories on a computer. Instead, it is about fine-tuning a huge, layered network of connections based on an inherently limited set of example data – the "training" data set. With every example that the system sees, the pattern and strength of the network's connections are tweaked until it can parse the training data set effectively, which would be things like classifying images, playing a game or driving a car.

The hope is that the performance generalises beyond the training data set to new, unseen data sets. But it doesn't always work so well because training data sets are often inherently biased in all sorts of impossible-to-notice ways. Often a network gets so fine-tuned to the specifics of the data set it is trained on that it fails to generalise to new ones.

This is called overfitting, and it is a ubiquitous problem in deep learning. A number of common techniques have been adopted to deal with this issue. Most involve exposing the network to some sort of stochasticity, introducing noise and randomness into the system.

One such strategy is "domain randomisation", wherein the inputs are warped in a highly biased way during learning, effectively inducing a hallucination in the network. This sort of thing has been found to be indispensable, for example, when the research company OpenAI trained a deep neural network to learn how to manipulate a robot hand to solve Rubik's cubes.

There is good reason to think the brain faces an identical challenge of overfitting. Animals' days are, after all, statistically pretty self-similar. Their "training set" is limited and highly biased. But still, an animal needs to generalise its abilities to new and unexpected circumstances, both in terms of physical movement and reaction, and cognition and understanding. It doesn't need to remember everything perfectly; it needs to generalise from the limited things it has seen and done.

This is the overfitted brain hypothesis (OBH): that animals, being so good at learning, are constantly in danger of fitting themselves too well to their daily lives and tasks.

I've recently been working on developing the OBH, exploring how dreams could be a way to beat back the tide of daily overfitting. Essentially, under the OBH, dreams are "noise injections" that serve the purpose not of enforcing what is learned when › awake, but rather counteracting the overfitting associated with that learning.

You can't do domain randomisation on an awake brain because most organisms are negotiating a high-wire act during daily life; they would certainly hurt themselves in myriad ways. However, you can use an offline period to do something similar by creating sparse and hallucinatory inputs, driven by top-down activity, that resemble the events and actions an animal might encounter, but that are corrupted and biased away from the drudgery of daily life.

According to the OBH then, dreams are exactly this: self-generated corrupted inputs. And the act of dreaming has the effect of improving generalisation and performance in waking life. This is how someone can go to sleep failing on their training task of juggling, and then wake up a juggler.

HARDZIEJ STUDIO

The advantage of this hypothesis is that it takes the phenomenology of dreams seriously, rather than as some sort of epiphenomenon or unexplained by-product of some other neural background process. Indeed, it is the strange phenomenology of dreams that makes them so effective at combating overfitting. While it may seem weird, experiencing events that are related to a task, but fundamentally different from it, can actually help performance. Dreaming of flying may help you keep your balance while running. And deep-learning practitioners should perhaps take a lesson from the brain and make their efforts to combat overfitting look as "dream-like" as possible for their networks.

Waking dreams

Of course, this is still very much a hypothesis – and an untested one at that. There is much work that needs to be done to assess what the behavioural benefits of dreams are and whether they match the sort of reductions in overfitting that we might expect in humans and other animals according to the OBH. Additionally, dream physiology – how synapses change during dreams and when dreaming occurs during sleep – are all still being investigated more generally.

But by viewing dreams through this new lens, we can at least move beyond computer and storage metaphors and begin to think of learning as a set of trade-offs, where memorisation competes with generalisation, and learning the specifics of something too well can be as bad as not learning at all.

If dreams have this functional purpose, and the OBH is true, then the artificial dreams we call fictions might satisfy some of that same fundamental drive. I spent 10 years writing my first novel, *The Revelations*, which is about consciousness and murder. I can give all the standard cultural reasons for why fictions are important, entertaining, revelatory – but the OBH implies there is something more. Maybe art is also pleasurable for humans because we are constantly being overfitted to reality.

In this view, the sparse, sometimes hallucinatory, corrupted unreality put forward by authors, film-makers, and those first early shamans around some campfire, all help to stop our minds becoming too fixed in their ways. They don't just expand the "training set" that humans have access to, but do so in ways that assist with generalisation and therefore cognition more broadly.

Perhaps the hypothetical aliens wouldn't be so puzzled by our obsessions with fictions once they figured this out. They wouldn't be shocked either that as human civilisation developed, daily life became more complex, and so it became easier for us to overfit to it – until eventually we humans began to spend more time with artificial dreams than we do with biological ones. Just like how the invention of cooking essentially allowed us to expand digestion beyond our stomachs, maybe the invention of fictions allowed us to get the benefits of dreams when we are awake. ■

Erik Hoel is a neuroscientist at Tufts University in Medford, Massachusetts. His debut novel, *The Revelations*, will be published in April 2021