Our dreams' weirdness might be why we have them, argues new AI-inspired theory of dreaming

14 May 2021

Illustration represents the overfitted brain hypothesis of dreaming, which claims that the sparse and hallucinatory quality of dreams is not a bug, but a feature, since it helps prevent the brain from overfitting to its biased daily sources of learning Credit: Georgia Turner

The question of why we dream is a divisive topic within the scientific community: it's hard to prove concretely why dreams occur and the neuroscience field is saturated with hypotheses. Inspired by techniques used to train deep neural networks, Erik Hoel (@erikphoel), a research assistant professor of neuroscience at Tufts University, argues for a new theory of dreams: the overfitted brain hypothesis. The hypothesis,

described May 14 in a review in the journal *Patterns*, suggests that the strangeness of our dreams serves to help our brains better generalize our day-to-day experiences.

"There's obviously an incredible number of theories of why we dream," says Hoel. "But I wanted to bring to attention a theory of dreams that takes dreaming itself very seriously—that says the experience of dreams is why you're dreaming."

A common problem when it comes to training AI is that it becomes too familiar with the data it's trained on—it starts to assume that the training set is a perfect representation of anything it might encounter. Data scientists fix this by introducing some chaos into the data; in one such regularization method, called "dropout," some data is randomly ignored. Imagine if black boxes suddenly appeared on the internal screen of a self-driving car: the car that sees the random black boxes on the screen and focuses on overarching details of its surroundings, rather than the specifics of that particular driving experience, will likely better understand the general experience of driving.

"The original inspiration for deep neural networks was the brain," Hoel says. And while comparing the brain to technology is not new, he explains that using deep neural networks to describe the overfitted brain hypothesis was a natural connection. "If you look at the techniques that people use in regularization of deep learning, it's often the case that those techniques bear some striking similarities to dreams," he says.

With that in mind, his new theory suggests that dreams happen to make our understanding of the world less simplistic and more well-rounded—because our brains, like deep neural networks, also become too familiar with the

"training set" of our everyday lives. To counteract the familiarity, he suggests, the brain creates a weirded version of the world in dreams, the mind's version of dropout. "It is the very strangeness of dreams in their divergence from waking experience that gives them their biological function," he writes.

Hoel says that there's already evidence from neuroscience research to support the overfitted brain hypothesis. For example, it's been shown that the most reliable way to prompt dreams about something that happens in real life is to repetitively perform a novel task while you are awake. He argues that when you over-train on a novel task, the condition of overfitting is triggered, and your brain attempts to then generalize for this task by creating dreams.

But he believes that there's also research that could be done to determine whether this is really why we dream. He says that well-designed behavioral tests could differentiate between generalization and memorization and the effect of sleep deprivation on both.

Another area he's interested to explore is on the idea of "artificial dreams." He came up with overfitted brain hypothesis while thinking about the purpose of works of fiction like film or novels. Now, he hypothesizes that outside stimuli like novels or TV shows might act as dream "substitutions"—and that they could perhaps even be designed to help delay the cognitive effects of sleep deprivation by emphasizing their dream-like nature (for instance, by virtual reality technology).

While you can simply turn off learning in artificial neural networks, Hoel says, you can't do that with a brain. Brains are always learning new things—and that's where the overfitted brain hypothesis comes in to help. "Life is boring sometimes," he says. "Dreams are there to keep you from becoming too fitted to the model of the world."