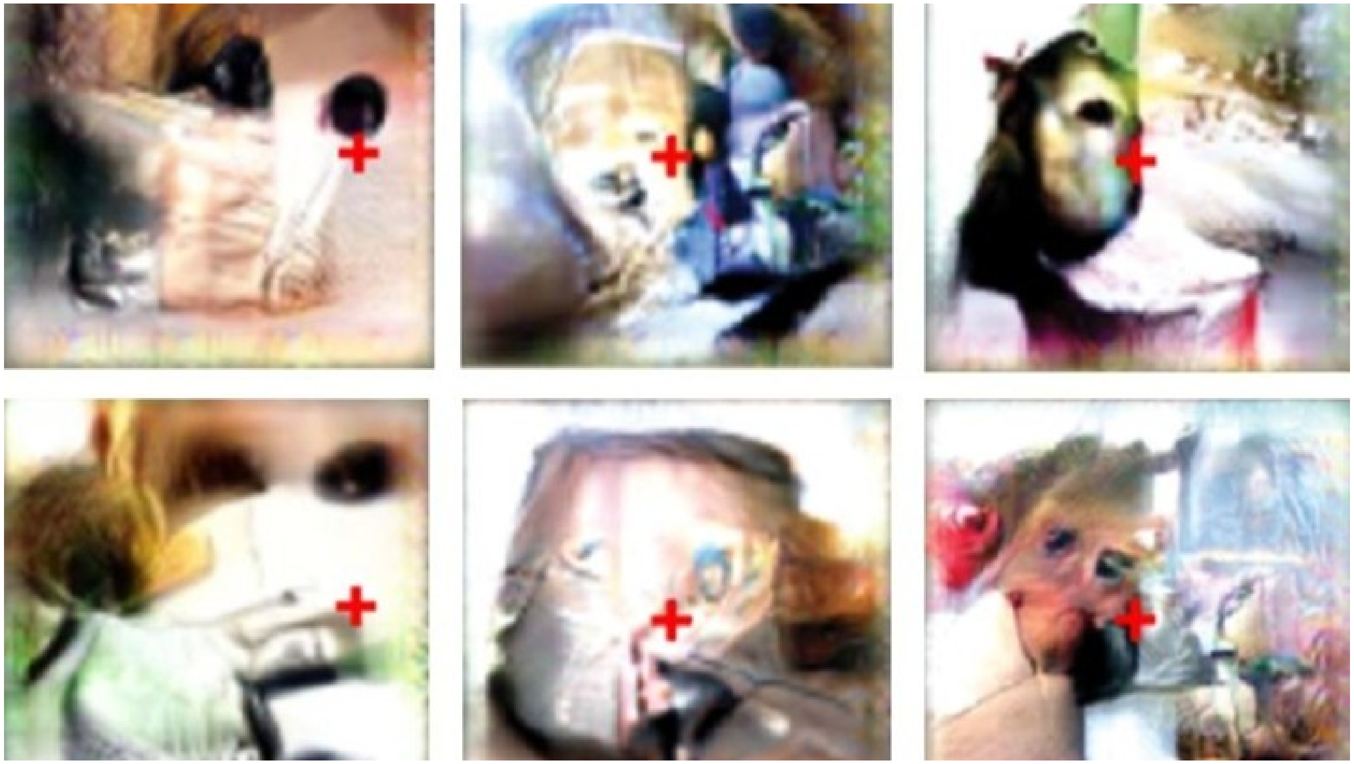
SCIENCE

AI Evolved These Creepy Images to Please a Monkey's Brain

What happens when an algorithm can ask neurons what they want to see?

ED YONG 2:00 PM ET



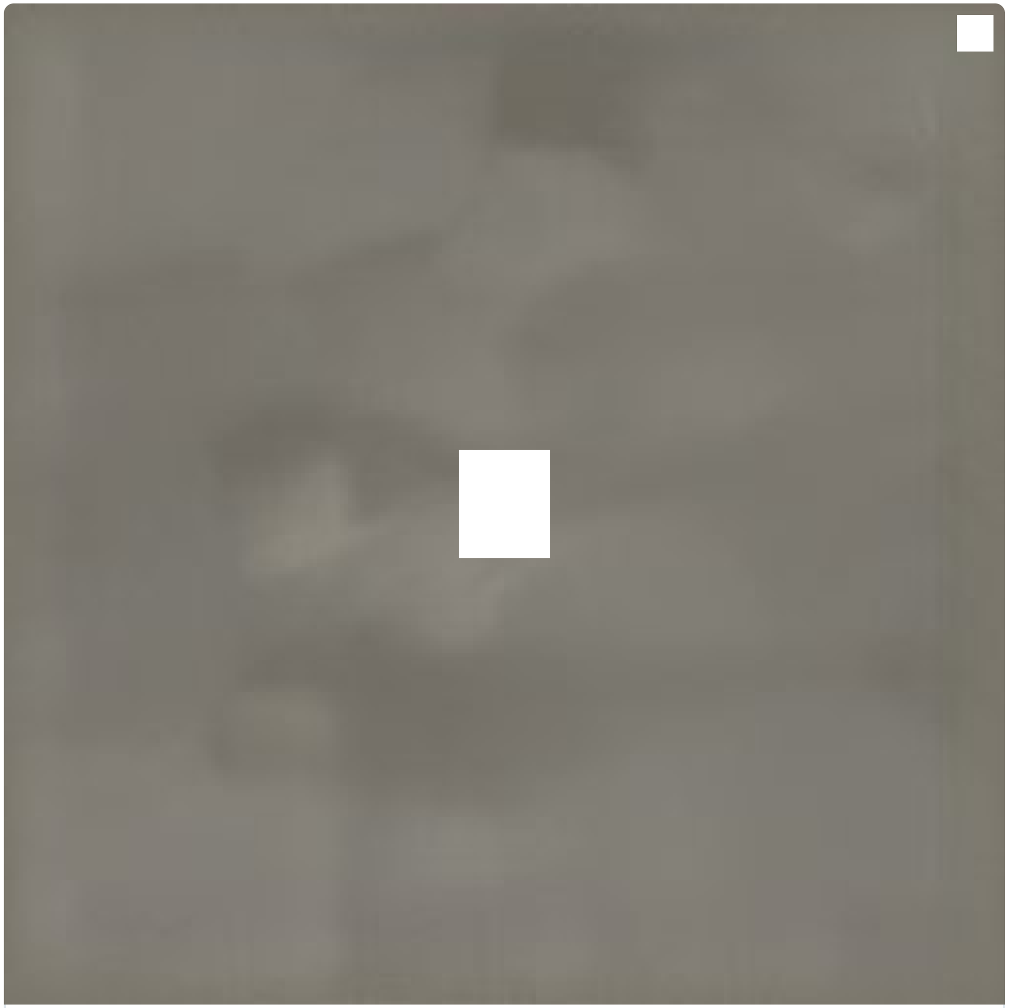
These images, produced by an artificial-intelligence algorithm called XDREAM, can stimulate particular neurons far better than any natural picture. (COURTESY OF CARLOS R. PONCE, ET AL. / HARVARD MEDICAL SCHOOL)

In April 2018, a monkey named Ringo sat in a Harvard lab, sipping juice, while strange images flickered in front of his eyes. The pictures were created by an artificial-intelligence algorithm called XDREAM, which gradually tweaked them to stimulate one particular neuron in Ringo's brain, in a region that's supposedly specialized for recognizing faces. As the images evolved, the neuron fired away, and the team behind XDREAM watched from a nearby room.

At first, the pictures were gray and formless. But as time passed, "from this haze, something started staring back at us," says the neuroscientist Carlos Ponce. Two black dots with a black line beneath them, all against a pale oval. A face, albeit an abstracted one. Soon a red patch appeared next to it, which reminded the watching researchers of the red collar worn by a monkey who lives in the cage opposite Ringo's. "We all looked at it and said, 'Oh, that's Anthony,'" says Margaret Livingstone, a neuroscientist at Harvard Medical School.

"And then, a few days later, we evolved Diane," she adds. Diane is one of the monkeys' caretakers, who feeds them while wearing blue scrubs and a white face mask. And when the team hooked XDREAM up to another of the monkey's visual neurons, it produced a distorted image of a face in a white mask.

XDREAM's images look like glitchy Kandinsky paintings viewed during a bad trip. You really wouldn't want to hang them on your wall. But each one is close to the ideal stimulus for a particular neuron. And collectively, they tell us something interesting about how our brain makes sense of the world, and how much we still don't understand about that process. “If cells are dreaming, [these images] are what the cells are dreaming about,” says Ponce. “It exposes the visual vocabulary of the brain, in a way that's unbiased by our anthropomorphic perspective.”



Peter Schade
@peterfschade

New pre-print on #bioRxiv 'Evolving super stimuli for real neurons using deep generative networks'. With @HombreCerebro, @WillXiao1, @Till_S_Hartmann, @gkreiman, and @mlivingstonehms biorxiv.org/content/early/…

96 6:03 PM - Jan 18, 2019

39 people are talking about this

The first hints of that vocabulary emerged in 1962, when Torsten Wiesel and David Hubel showed that specific neurons in the brain's visual centers are tuned to specific stimuli—lights moving in particular directions, or lines aligned in particular ways. Since then, other neuroscientists have identified neurons that respond to colors, curvatures, faces, hands, and outdoor scenes. But here's the catch: Those scientists always *chose* which kinds of shape to test, and their intuition might not reflect the *actual* stimuli to which the neurons are attuned. "Just because a cell responds to a specific category of image doesn't mean you really understand what it wants," says Livingstone.

So why not ask the neurons what they want to see?

[Read: The human remembering machine]

That was the idea behind XDREAM, an algorithm dreamed up by a Harvard student named Will Xiao. Sets of those gray, formless images, 40 in all, were shown to watching monkeys, and the algorithm tweaked and shuffled those that provoked the strongest responses in chosen neurons to create a new generation of pics. Xiao had previously trained XDREAM using 1.4 million real-world photos so that it would generate synthetic images with the properties of natural ones. Over 250 such generations, the synthetic images became more and more effective, until they were exciting their target neurons far more intensely than any natural image. “It was exciting to finally let a cell tell us what it’s encoding instead of having to guess,” says Ponce, who is now at Washington University in St. Louis.

There’s a risk that XDREAM could become a glorified Rorschach test, in which researchers see what *they* want to see. Is that red splotch really Anthony’s collar? Is the white one really Diane’s masked face? To check, the team used another algorithm to confirm that the synthetic images they saw as face-like really do look more like actual faces than other natural photos. They also showed that the neurons that prompt XDREAM to create face-like motifs themselves respond best to photos of true faces.

I mention to Ponce that XDREAM’s images are really unsettling, as if they’ve been plucked from some deep recess of the uncanny valley. “Yes!” he laughs. “They are!” He thinks they’re so good at stimulating monkey visual neurons that they’re also tickling our cells in a way that makes us feel uncomfortable. If one could use XDREAM on human neurons, “would we find similar images or different, and what would we think of them?” he asks. “At the moment, that’s not something anyone can do. But it makes me wonder.”

Livingstone also wonders whether XDREAM’s disquieting output hints at why so many mythical creatures are exaggerated versions of familiar things. Visual neurons, it seems, like exaggeration: In previous studies, her team showed that face-selective cells will respond more strongly to caricatures than to actual faces. “I think that gargoyles and leprechauns, these archetypes that people imagine ... there’s a basis in our brains for them,” she says.

Beyond being weird, the most striking thing about XDREAM’s images is that they’re mostly unrecognizable. The team probed 46 neurons across six monkeys, and a few face-like motifs aside, most of the resulting images were messes of color, texture, and shape, which didn’t fit into obvious buckets. “It is striking that cells that were thought to code for simple objects or object parts may in fact code for much more complex visual stimuli,” says Leyla Isik, a neuroscientist at Johns Hopkins University. “Some may find it unsatisfying that the generated images cannot be described easily in terms of semantic categories. This ‘limitation,’ however, may just be a reality of the complex nature of the primate visual cortex.”

Through these experiments, researchers are learning more not just about the brain itself, but also about how to simulate it. Many neuroscientists have developed artificial neural networks that can analyze images and recognize objects, ostensibly by doing something close to what the brain’s actual visual centers do. But how close?

To find out, Pouya Bashivan at MIT used one such neural network to create images that should, theoretically, stimulate an actual brain in particular ways. His colleagues, Kohitij Kar and James DiCarlo, then showed these synthetic images to monkeys to see whether they worked as predicted.

The results were encouraging, if mixed. The neural network succeeded in fashioning images that would stimulate specific neurons more strongly than natural photos. But it wasn't as good at another task: exciting one neuron while suppressing all its neighbors. This varied scorecard suggests that the network isn't yet capturing everything there is to capture about the visual system.

[Read: The AI-art gold rush is here]

Still, it's capturing *something*. Bashivan's team focused on a region that supposedly responds to simple curves, but the images that his network churned out included grids, lattices, and cinnamon-roll swirls. Much like XDREAM's hallucinogenic not-quite-faces, these complex images suggest that our understanding of how the brain sees the world is too basic. "If we only go by the intuitions of human researchers, we might get it wrong," says Bashivan. "We'll do better if we have models that contain all the knowledge in the field."

"As biologists, many of us are still skeptical that current neural networks are similar enough to the brain to model it reliably," Ponce says. But like Bashivan, he thinks that such models are the way forward, and studies such as these will help improve them. "Both approaches are about understanding a black box: the brain," he says. "Both methods are necessary."

We want to hear what you think about this article. Submit a letter to the editor or write to letters@theatlantic.com.