

A hallucinating artificial intelligence might see something like this product of Google’s Deep Dream algorithm.

DEBORAH LEE SOLTESZ/FLICKR

Could artificial intelligence get depressed and have hallucinations?

**By**[**Matthew Hutson**](http://www.sciencemag.org/author/matthew-hutson)**Apr. 9, 2018 , 12:10 PM**

As artificial intelligence (AI) allows machines to become more like humans, will they experience similar psychological quirks such as hallucinations or depression? And might this be a good thing?

Last month, New York University in New York City hosted a symposium called [**Canonical Computations in Brains and Machines**](http://gias.nyu.edu/canonical-computation-brains-machines-2018/), where neuroscientists and AI experts discussed overlaps in the way humans and machines think. Zachary Mainen, a neuroscientist at the Champalimaud Centre for the Unknown, a neuroscience and cancer research institute in Lisbon, [**speculated**](https://www.youtube.com/watch?v=dMHhcf3jAPE) that we might expect an intelligent machine to suffer some of the same mental problems people do.

He spoke with *Science* after the symposium; this interview has been edited for brevity and clarity.

Top of Form

Bottom of Form

**Q: Why do you think AIs might get depressed and hallucinate?**

**A:** I’m drawing on the field of computational psychiatry, which assumes we can learn about a patient who’s depressed or hallucinating from studying AI algorithms like reinforcement learning. If you reverse the arrow, why wouldn’t an AI be subject to the sort of things that go wrong with patients?

**Q: Might the mechanism be the same as it is in humans?**

**A:** Depression and hallucinations appear to depend on a chemical in the brain called serotonin. It may be that serotonin is just a biological quirk. But if serotonin is helping solve a more general problem for intelligent systems, then machines might implement a similar function, and if serotonin goes wrong in humans, the equivalent in a machine could also go wrong.



 Zachary Mainen

TOR STENSOLA

**Q: How does serotonin help us think?**

**A:** Serotonin is a neuromodulator, a special kind of neurotransmitter that quickly broadcasts its message to a large fraction of the brain. For instance, the neuromodulator dopamine seems to act as a global reward signal. Something good just happened, and the whole brain needs to learn about this. Computational approaches to neuroscience see other neuromodulators as other sorts of “control knobs” similar to those used in AI. One important “knob” from AI is the learning rate.

**Q: When would you want to adjust the learning rate in an AI or the brain?**

**A:** There are situations in which the world suddenly changes and everything becomes a lot more uncertain and difficult than usual. For example, when you’ve gone to foreign city. In these situations, your old knowledge—your usual model of the world—is out of date and needs to be sidelined or reworked to adapt to the new situation.

In the lab we can create that sort of situation by teaching a mouse or a person a game with certain rules and then abruptly changing the rules. When that happens, serotonin neurons light up. People think of serotonin as related to happiness, but serotonin neurons appear to send a message that is not good or bad but more “oops” or surprise—my old expectations are not right any more.

**Q: If serotonin primarily signals uncertainty rather than happiness, how does this explain its importance to depression?**

**A:**In the lab, serotonin release has been implicated in brain plasticity [i.e. its ability to change]. It seems to be especially important in breaking or suppressing outdated beliefs. These results suggest to us that treating depression through pharmacology is not so much about improving mood, but rather as helping to cope with change.

Depression can be seen as getting stuck in a model of the world that needs to change. An example would be someone who suffers a severe injury and needs to think of themselves and their abilities in a new way. A person who fails to do so that might become depressed. Selective serotonin reuptake inhibitors [such as Prozac, and which are a common type of antidepressant] can facilitate brain plasticity. Psychedelics, like LSD, psilocybin, or DMT may be acting similarly but on a shorter time scale. In fact, psilocybin is currently being tested in clinical trials for depression.

**Q: Could AI have emotions, too, or even get depressed?**

**A:** Yes, I think robots would likely have something like emotions. Similar issues face a person or an AI, for example, when the environment changes radically. Humans or machines with low serotonin or its equivalent may fail to rewire themselves adequately, getting stuck in the rut that we call depression.

Posted in:

* [**Technology**](http://www.sciencemag.org/category/technology)

doi:10.1126/science.aat8056