

in the distant Universe, and provide fuel for observations using the future James Webb Space Telescope. ■

**Rycharð Bouwens** is at the Leiden Observatory, Leiden University, Leiden NL-2333, the Netherlands. e-mail: bouwens@strw.leidenuniv.nl

1. Loeb, A. & Barkana, R. *Annu. Rev. Astron. Astrophys.* **39**, 19–66 (2001).

2. Abel, T., Bryan, G. L. & Norman, M. L. *Science* **295**, 93–98 (2002).
3. Hashimoto, T. *et al. Nature* **557**, 392–395 (2018).
4. Zheng, W. *et al. Nature* **489**, 406–408 (2012).
5. Zitrin, A. *et al. Astrophys. J. Lett.* **810**, 12 (2015).
6. Oesch, P. A. *et al. Astrophys. J.* **819**, 129 (2016).
7. Smit, R. *et al. Nature* **553**, 178–181 (2018).
8. Capak, P. L. *et al. Nature* **522**, 455–458 (2015).
9. Willott, C. J., Carilli, C. L., Wagg, J. & Wang, R. *Astrophys. J.* **807**, 180 (2015).
10. Inoue, A. K. *et al. Astrophys. J. Lett.* **780**, 18 (2014).
11. Inoue, A. K. *et al. Science* **352**, 1559–1562 (2016).
12. Bradač, M. *et al. Astrophys. J.* **785**, 108 (2014).

13. Hoag, A. *et al. Astrophys. J.* **854**, 39 (2018).
14. Roberts-Borsani, G. W. *et al. Astrophys. J.* **823**, 143 (2016).
15. Smit, R. *et al. Astrophys. J.* **784**, 58 (2014).
16. Laporte, N. *et al. Astron. Astrophys.* **562**, 8 (2014).
17. Planck Collaboration. *Astron. Astrophys.* **596**, 107 (2016).
18. Oesch, P. A., Bouwens, R. J., Illingworth, G. D., Labbé, I. & Stefanon, M. *Astrophys. J.* **855**, 105 (2018).
19. Bowman, J. D., Rogers, A. E. E., Monsalve, R. A., Mozdzen, T. J. & Mahesh, N. *Nature* **555**, 67–70 (2018).

## COMPUTATIONAL NEUROSCIENCE

# AI mimics brain codes for navigation

An artificial-intelligence technique called deep learning has now been used to model spatial navigation. The system develops a representation of space similar to that of the grid cells found in the mammalian brain. [SEE LETTER P.429](#)

FRANCESCO SAVELLI & JAMES J. KNIERIM

Deep learning is an approach to artificial intelligence that is inspired by the brain's neural networks. The technique is contributing to a plethora of technologies, from automated video analysis to language translation. One page 429, Banino *et al.*<sup>1</sup> use this framework to gain insights into real-life neuronal networks — in particular, how geometrically regular representations of space can facilitate flexible navigation strategies.

Deep-learning networks can be taught how to process inputs to achieve a particular output — for instance, learning to pick out a particular face in many photos of different people. The networks are 'deep' in that they are made up of sequential layers of repeated computational units. Each unit receives inputs from similar units in the previous layer and sends outputs to those in the next. Mathematically, such a network can be viewed as a high-dimensional function, which can be modulated by altering how the outputs of one layer are weighted in the next.

The network tunes the function during a training phase, which typically relies on a set of input–output examples. For instance, a deep-learning system might be shown a series of photos, and told which ones contain the face it aims to identify. Its weights are automatically tuned by optimization algorithms until it learns to make a correct identification. The network's deep organization gives it a prodigious ability to spot and take advantage of the most useful features and patterns that recur across the examples, and distinguish different faces. But one downside is that the final network tends to be a black box — the computational solutions derived during training often cannot be deciphered from the myriad

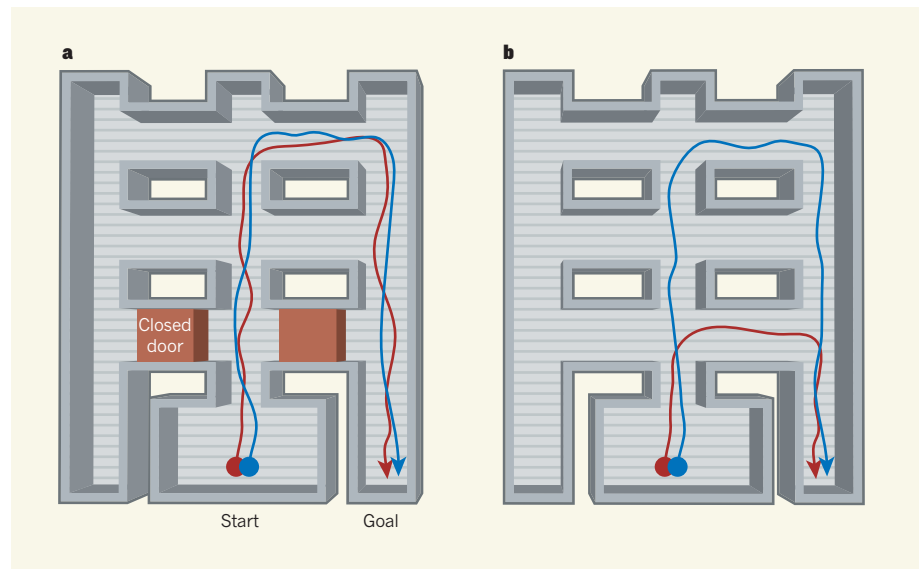
weights assigned throughout the layers.

Deep-learning networks can successfully perform perceptual tasks<sup>2</sup>, but there have been fewer studies of complex behavioural tasks such as navigation. A key aspect of real-life navigation is estimating one's position following each step, by calculating the displacement per step on the basis of orientation and distance travelled. This process is called

path integration and is thought by neuroscientists, cognitive scientists and roboticists to be crucial for generating a cognitive map of the environment<sup>3–5</sup>. There are several kinds of neuron associated with the brain's cognitive maps, including place cells, which fire when the organism occupies a particular position in the environment, and head-direction cells, which signal head orientation.

A third type of neuron, the grid cell, fires when the animal is at any of a set of points that form a hexagonal grid pattern across the environment. Grid cells are thought to endow the cognitive map with geometric properties that help in planning and following trajectories. These cells are found in the brain's hippocampal formation, a region that, in humans, is involved in spatial learning, autobiographical memories and knowledge of general facts about the world.

Banino and colleagues set out to generate path integration in a deep-learning network. Because path integration involves



**Figure 1 | An AI system learns to take shortcuts.** In the mammalian brain, place cells fire when an animal is at a particular position within an environment, head-direction cells fire when the head is in a particular orientation, and grid cells fire when the animal is at points that form a hexagonal grid across the environment. Banino and colleagues<sup>1</sup> trained an artificial-intelligence system called a deep-learning network to navigate, by providing it with simulations of rodent foraging patterns, including about the activity of place and head-direction cells. Some computational units in the network developed grid-cell-like firing patterns (not shown). **a**, While learning to navigate towards a goal, similar paths were taken by both a system using grid cells (red line indicates a sample path) and a system that used place and head-direction cells instead (blue line). **b**, But when shortcuts were introduced, for example by opening previously closed doors, only the system that used grid cells found the shorter routes, highlighting the ability of grid-cell-like activity to promote flexible navigation strategies. (Figure adapted from Extended Data Fig. 10 in ref. 1.)

remembering the output from the previous processing step and using it as input for the next, the authors used a network involving feedback loops. They trained the network using simulations of pathways taken by foraging rodents. The system received information about the simulated rodent's linear and angular velocity, and about the simulated activity of place and head-direction cells — the latter two acting as an 'oracle' for the current location and head direction of the rodent.

The authors found that patterns of activity resembling grid cells spontaneously emerged in computational units in an intermediate layer of the network during training, even though nothing in the network or the training protocol explicitly imposed this type of pattern. The emergence of grid-like units is an impressive example of deep learning doing what it does best: inventing an original, often unpredicted internal representation to help solve a task.

Grid-like units allow the network to keep track of position on the basis of path integration. Can they also help the system to learn to navigate efficiently from its current position to a goal location? To address this question, Banino *et al.* added a reinforcement-learning component, in which the network learned to assign values to specific actions taken at specific locations. Higher values were assigned to actions that brought the simulated rodent closer to the goal, acting as a reward. The grid-like representation markedly improved the ability of the network to solve goal-directed tasks, compared to control simulations in which the start and goal locations were encoded instead by place and head-direction cells. The trained network found smarter shortcuts when obstacles such as closed doors were removed (Fig. 1), and even extrapolated paths towards goals in a previously unexplored annex of a familiar environment. These results support the idea that grid cells enable the brain to perform vector calculations (calculations about the length and direction of a path) to assist path planning through compartmentalized<sup>6</sup> or previously unexplored<sup>7</sup> environments.

In the future, the authors' network could be used to explore the consequences of interactions between grid and place cells. In the current network, the simulated place layer does not change during training. However, in the brain, grid and place cells influence each other in ways that are not well understood. Although real-life place cells can remain spatially selective in the absence of grid-cell inputs<sup>8</sup>, these inputs seem important when an animal is far from external landmarks that can be used to define locations<sup>9–11</sup>. Under these conditions, place cells presumably rely on path integration and grid cells to maintain an accurate estimate of position. By developing the network such that the place-cell layer can be modulated by grid-like inputs, we could begin to unpack this relationship.

From a broader perspective, it is interesting that the network, starting from very general

computational assumptions that do not take into account specific biological mechanisms, found a solution to path integration that seems similar to the brain's. That the network converged on such a solution is compelling evidence that there is something special about grid cells' activity patterns that supports path integration. The black-box character of deep-learning systems, however, means that it might be hard to determine what that something is.

Likewise, the fact that the grid representation enhanced goal-directed performance is a compelling proof-of-concept of the role of grid cells in the brain. But the authors had to use correlational analyses, guided by qualitative intuitions, to indirectly infer that the network was making vector calculations. The inability to directly manipulate these calculations in the model makes it difficult to examine the computational principles, algorithms and encoding strategies that make grid-cell representations of space such an efficient solution for navigation. As such, the theoretician ends up in the same quandary as the experimentalist: trying to tease apart a poorly understood complex system to understand it. Making deep-learning systems more intelligible to human reasoning is an exciting challenge for the future. ■

**Francesco Savelli and James J. Knierim** are at the Zanvyl Krieger Mind/Brain Institute, Johns Hopkins University, Baltimore, Maryland 21218, USA. J.J.K. is also in the Solomon H. Snyder Department of Neuroscience, Johns Hopkins University. e-mails: fsavelli.research@gmail.com; jknierim@jhu.edu

1. Banino, A. *et al.* *Nature* **557**, 429–433 (2018).
2. LeCun, Y., Bengio, Y. & Hinton, G. *Nature* **521**, 436–444 (2015).
3. McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I. & Moser, M.-B. *Nature Rev. Neurosci.* **7**, 663–678 (2006).
4. Gallistel, C. R. *The Organization of Learning* (MIT Press, 1990).
5. Thrun, S., Burgard, W. & Fox, D. *Probabilistic Robotics* (MIT Press, 2005).
6. Carpenter, F., Manson, D., Jeffery, K., Burgess, N. & Barry, C. *Curr. Biol.* **25**, 1176–1182 (2015).
7. Savelli, F., Luck, J. D. & Knierim, J. J. *eLife* **6**, e21354 (2017).
8. Koenig, J., Linder, A. N., Leutgeb, J. K. & Leutgeb, S. *Science* **332**, 592–595 (2011).
9. Muessig, L., Hauser, J., Wills, T. J. & Cacucci, F. *Neuron* **86**, 1167–1173 (2015).
10. Wang, Y., Romani, S., Lustig, B., Leonardo, A. & Pastalkova, E. *Nature Neurosci.* **18**, 282–288 (2015).
11. Mallory, C. S., Hardcastle, K., Bant, J. S. & Giocomo, L. M. *Nature Neurosci.* **21**, 270–82 (2018).

This article was published online on 9 May 2018.

## OCEANOGRAPHY

# Climate change and oxygen in the ocean

**Computer simulations show that areas of the ocean that have low levels of dissolved oxygen will expand, but then shrink, in response to global warming — adding to an emerging picture of the finely balanced processes involved.**

LAURE RESPLANDY

**G**lobal warming has reduced the amount of dissolved oxygen in the ocean by 2% since 1960 (ref. 1). A major concern is that the rate of loss of dissolved oxygen has already increased by up to 20% in tropical waters, expanding the volume of regions called oxygen minimum zones (OMZs), where levels of dissolved oxygen are already very low<sup>2,3</sup>. The expansion of tropical OMZs threatens the survival of marine organisms that rely on dissolved oxygen for respiration, and affects the biogeochemical cycling of carbon and nitrogen, potentially amplifying global warming<sup>4</sup>. Writing in *Global Biogeochemical Cycles*, Fu *et al.*<sup>5</sup> suggest that, in the long term, tropical OMZs might shrink after their initial expansion, reversing their impact on warming.

Although, overall, tropical oceans have lost dissolved oxygen in the past 50 years, observations indicate strong regional and temporal variations. For example, OMZs have clearly

expanded since the 1970s in the equatorial Pacific and Atlantic oceans<sup>3,6</sup>, but a long-term reconstruction<sup>7</sup> of the OMZ close to the Californian and Mexican coasts suggests that it had been shrinking for a century before it started to expand in the 1990s. In the Indian Ocean, the northern part of the OMZ is shrinking, whereas the southern part is expanding<sup>8</sup>.

Three competing processes control the levels of dissolved oxygen in the ocean. The first is the transfer of atmospheric oxygen to the surface ocean (which, in turn, is tied to oxygen solubility, the capacity of the water to hold on to dissolved oxygen). The second is ocean circulation (ventilation), which carries the oxygen-rich surface waters to the ocean interior. And, finally, there is the biological respiration of dissolved oxygen, the process by which microorganisms consume organic matter produced at the surface as it sinks to the deep ocean. OMZs develop in tropical intermediate waters (those at depths of 200–1,000 metres) because the physical supply