

MACHINE LEARNING

Intelligence without Sentience

Artificial intelligence is coming of age and challenging our belief that being smart and being conscious go hand-in-hand

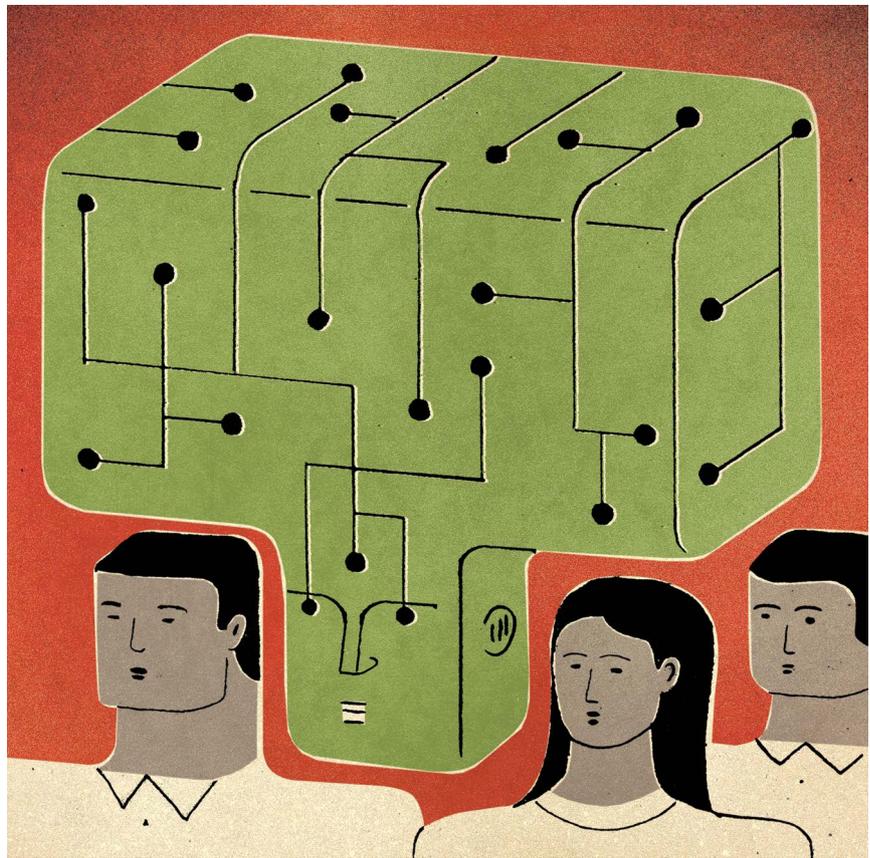
Throughout human history, intelligence and consciousness have been two closely allied concepts. If you have lots of the former, you are assumed, in some ill-defined way, to be more conscious than the dim-witted guy down the street. A smart gal would also be a very conscious one, somebody who could tell you in detail about her experiences (for that is what consciousness is, the ability to experience something, anything, whether it's a toothache, the sight of a canary-yellow house or searing anger). But this intimate relation may be unraveling.

Consider the latest advance from DeepMind, a small company in London co-founded in 2011 by Demis Hassabis, a British child chess prodigy, video game designer and computational neuroscientist. DeepMind was bought last year for hundreds of millions of dollars by Google. What its new code does is breathtaking: it teaches itself to play video games, often much better than human



BY CHRISTOF KOCH

Christof Koch is president and chief scientific officer at the Allen Institute for Brain Science in Seattle. He serves on *Scientific American Mind's* board of advisers.



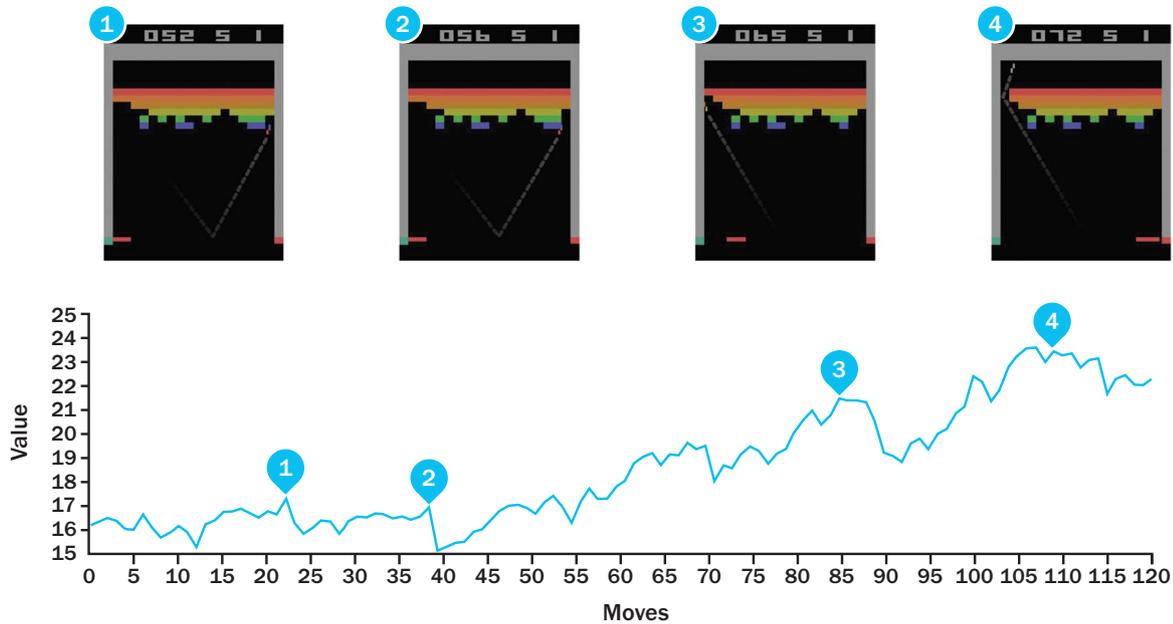
players. The technical breakthrough is described in a study published in February in *Nature*. (*Scientific American Mind* is part of Nature Publishing Group.)

To get a whiff of the excitement, go online and look for a YouTube video called *DeepMind Artificial Intelligence @ FDOT14*. It's a short excerpt, taken by smartphone, from Hassabis's talk at a 2014 tech conference, featuring a computer algorithm that learns to play the classic arcade game Breakout. The aim of the game, a variant of Pong, is for the player to break bricks aligned in rows on the top of the screen using a ball that bounces off the top and sidewalls. If the ball touches the bottom of the screen, the player loses one of three lives. To prevent that outcome, the player moves a paddle along the bottom to deflect the ball upward.

Co-created by Steve Wozniak of Apple fame, the game is primitive by today's standards yet compelling. Hassabis explained this onstage as he introduced the

audience to the algorithm. It started out knowing nothing and randomly fumbled the paddle, without much coordination, only occasionally hitting the ball. After an hour of training, playing over and over again, its performance improved, frequently returning the ball and breaking bricks. After two hours of training, it became better than most humans, returning balls fast and at steep angles.

The programmers let the algorithm continue to play on its own, and it kept on improving. After four hours of gaming, the algorithm discovered an innovative strategy to Breakout that boosted its performance way past that of any human. The algorithm accomplished this feat by learning to dig a tunnel through the wall on the side, allowing the ball to quickly destroy a large number of bricks from behind. Very clever. The achievement was so impressive that the assembled experts broke into spontaneous applause (a rare occurrence at scientific conferences).



DeepMind's artificial-intelligence algorithm learns to master Breakout, an arcade game that involves clearing away bricks using a ball and paddle. As it gains experience after many moves, the software finds new ways to raise its performance, including an innovative strategy (shown in panel 4) that digs through the wall to break a large number of bricks from behind.

To understand what's going on and why it's such a big deal, let's look under the hood. The algorithm incorporates three features, all gleaned from neurobiology: reinforcement learning, deep convolutional networks and selective memory replay.

A lasting legacy of behaviorism, the field that dominated the study of human and animal behavior in the first part of the 20th century, was the idea that organisms learn optimal behavior by relating the consequence of a particular action to a specific stimulus that preceded it. This stimulus is said to reinforce the behavior.

Consider my Bernese mountain dog, Ruby, as a puppy, when I had to house-break her. After giving Ruby water to drink at prescribed intervals, I immediately took her to a particular spot in the garden and waited—and waited. At some point, she would spontaneously pee, and I would lavishly praise her. If an indoor accident happened, I talked sternly to her. Dogs respond well to such positive and negative social signals. Over a month or two Ruby learned that an internal stimulus—a full bladder—followed by a behav-

ior—peeing in her special spot—predicted a reward and avoided punishment.

Reinforcement learning has been formalized and implemented in neural networks to teach computers how to play games. Gerald Tesauro of IBM used a particular version of reinforcement learning—temporal-difference learning—to design a network that played backgammon. The program analyzes the board and examines all the possible legal moves and responses of the opposing player to these moves. All the resulting board positions are fed into the program's heart, its value function.

The action that is chosen by the program is the one that leads to the board position with the highest score. After a turn, the network is slightly tweaked so that the program predicts what happens next a little bit better than what it predicted following its previous move. Starting from scratch, the program becomes better and better by trial and error. What makes reinforcement learning a challenge is that there is usually a substantial delay between any one particular move

and its eventual beneficial or detrimental outcome. Overcoming this handicap requires training, training and more training—beating human experts at backgammon required Tesauro's program to play 200,000 games against itself.

The second ingredient of DeepMind's success is called a deep convolutional network. It is based on a model of the brain circuitry found in the mammalian visual system by Torsten Wiesel and the late David H. Hubel, both then at Harvard University, in the late 1950s and early 1960s (work for which they would later be awarded a Nobel Prize). The model postulates a layer of processing elements, or units, that compute a weighted sum of an input. If the sum is sufficiently large, the model turns the unit's output on; otherwise, it remains off.

The visual system is thought by some theoreticians to be essentially nothing but a cascade of such processing layers—what is labeled a feed-forward network. Each layer receives input from a previous layer and passes on the output to the next level. The first layer is the retina that captures the rain of arriving photons. It accounts for variations in image brightness and passes these data on to the next processing stage. The last layer consists of a bunch of units that signal whether or not

some high-level feature, such as your grandmother or Jennifer Aniston, is present in that image.

Learning theorists developed mathematically sound methods to adjust the weights on these units—how influential one input should be relative to another one—to get such feed-forward networks to learn to perform specific recognition tasks. For instance, a network is exposed to tens of thousands of images from the Internet, each one labeled according to whether or not the photograph includes a cat. After every exposure, all weights are slightly adjusted. If the training is sufficiently long (again, the training is very computer-intensive) and the images are processed in deep enough networks—those with many layers of processing elements—the neural network generalizes and can accurately recognize a new photograph as containing a feline. The network has learned, in a supervised manner, to distinguish cat images from those of dogs, people, cars, and so on. The situation is not that dissimilar from a mother going through a picture book with her

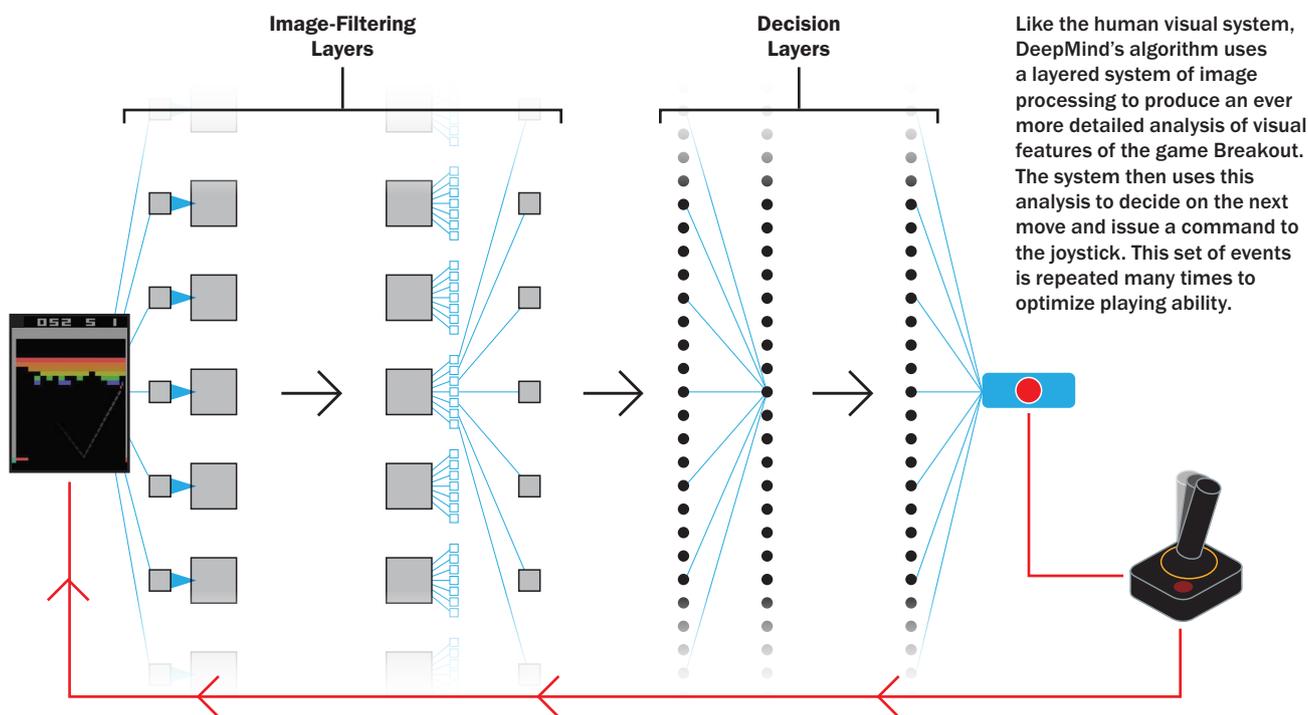
toddler while pointing out all the cats to the child. Deep convolutional networks are all the rage at Google, Facebook, Apple and other Silicon Valley companies seeking to automatically label images, translate speech to text, detect pedestrians in videos and find tumors in breast scans.

The supervised learning differs from reinforcement learning. In the former, every input image is paired with a label—one image contains a cat; another does not. In reinforcement learning, the consequence of any action in the game score unfolds in time—the actions may yield benefits (improved scores) but only many moves later.

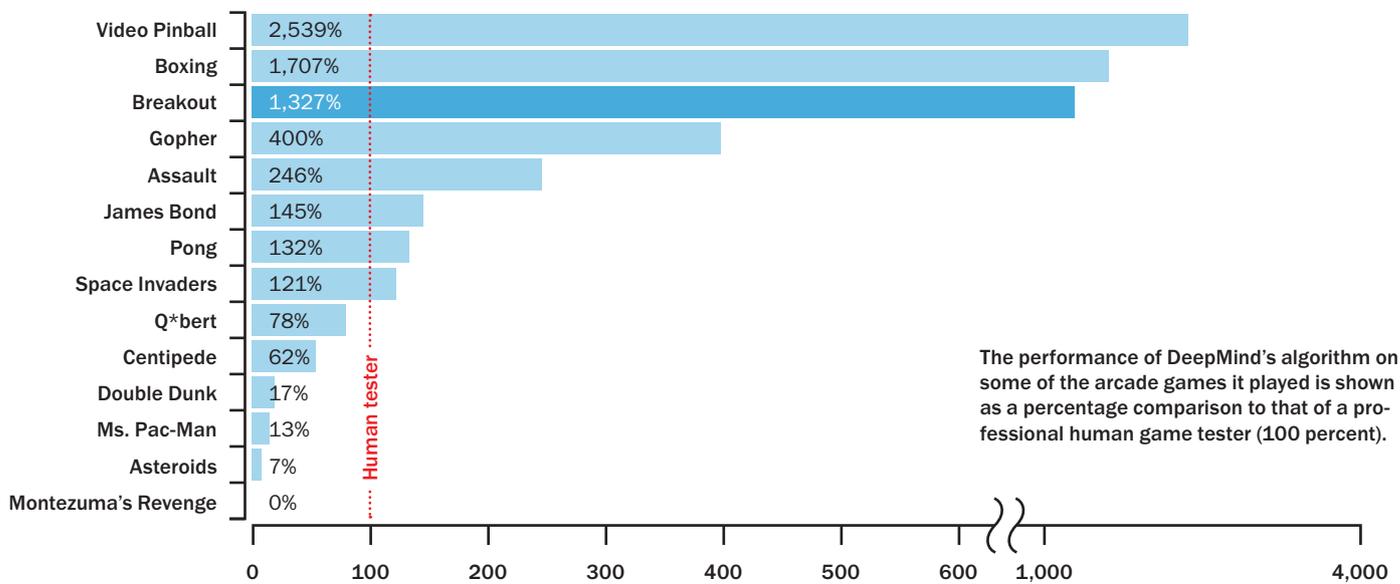
Hassabis and his large team (the *Nature* paper included 19 co-authors in all) used a variant of reinforcement learning called Q-learning to act as a supervisor for the deep-learning network. The input to the network consisted of a blurry version of the colored game screen, including the game score—the same as seen by a human player—as well as the screens associated with the last three moves. The output of the network was a command to the

joystick—to move in one of the eight cardinal directions, with or without activating the red “fire” button. Starting with a random setting of its weights, the proverbial blank slate, the algorithm figured out which actions would lead the all-important score to increase—when exactly the paddle was most likely to successfully intercept the ball on the bottom to break a brick on its upward trajectory. In this manner, the network learned and, through repetition, reinforced training of successful ways to play Breakout, outperforming a professional human game tester by a stunning 1,327 percent.

The third critical component of the algorithm was selective memory replay—similar to what is thought to occur in the hippocampus, a brain region associated with memory. In the hippocampus, activity patterns of nerve cells associated with a particular experience, such as running a maze, reoccur but at a faster pace on replay. That is, the algorithm would randomly recall a particular game episode, including its own actions, that it encountered earlier on from its memory bank



THIS PAGE AND OPPOSITE PAGE: SOURCE: "HUMAN-LEVEL CONTROL THROUGH DEEP REINFORCEMENT LEARNING," BY VOLODYMYR MNIH ET AL., IN NATURE, VOL. 518, FEBRUARY 26, 2015



The performance of DeepMind's algorithm on some of the arcade games it played is shown as a percentage comparison to that of a professional human game tester (100 percent).

and would retrain itself using this earlier experience, updating its evaluation function appropriately.

The folks at DeepMind were not content to let their algorithm learn just one game. They trained the same algorithm on 49 different Atari 2600 games, all of which were designed to engage generations of teenagers. They included Video Pinball, Star Gunner, Robot Tank, Road Runner, Pong, Space Invaders, Ms. Pac-Man, Alien and Montezuma's Revenge. The same algorithm, with the same settings, was used in all cases. Solely the output differed according to the specific needs of each game. The results blew all competing game-playing algorithms out of the water. What's more, the algorithm performed at 75 percent or better than the level achieved by a human professional game tester in 29 of these games, sometimes by a very large margin.

The algorithm did have its limitations. Its performance grew progressively worse as games demanded ever more long-term planning. For instance, the algorithm's performance in Ms. Pac-Man was pretty dismal because the game requires one to, say, choose which path in the maze to take to avoid being gobbled up by a ghost that is still 10 or more moves away in the future.

The program, however, heralds a new sophistication in AI. Deep Blue, the IBM

program that beat chess grandmaster Garry Kasparov in 1997, and Watson, the IBM system that bested Ken Jennings and Brad Rutter in the quiz show *Jeopardy*, were highly specialized collections of algorithms carefully handcrafted to their particular problem domain. The hallmark of the new generation of algorithms is that they learn, like people, from their triumphs and their failures. Starting with nothing but the raw pixels from the game screen, they eventually compete in side-scrolling shooters, boxing games and car-racing games. Of course, the worlds in which they operate are physically highly simplistic, obeying restrictive rules, and their actions are severely limited.

There is no hint of sentience in these algorithms. They have none of the behaviors we associate with consciousness. Existing theoretical models of consciousness would predict that deep convolutional networks are not conscious. They are zombies, acting in the world but doing so without any feeling, displaying a limited form of alien, cold intelligence: an algorithm "ruthlessly exploits the weakness in the system that it has found. This is all automatic," Hassabis said in his 2014 talk. Such algorithms, including those that control Google's self-driving cars or the ones that execute trades in the financial markets, demonstrate that for the first time in the planet's history, intel-

ligence can be completely dissociated from sentience, from consciousness.

They are smart in the sense that they can learn to adapt to new worlds, motivated by nothing but maximizing cumulative reward, as defined by the game score. I have no doubt that DeepMind designers are busy working on more sophisticated learning engines, teaching their algorithms to dominate first-person shooter games, such as *Doom* or *Halo*, or strategy games, such as *StarCraft*. These algorithms will become better and better at executing specific tasks in narrowly defined niches of the kind that abound in the modern world. They will neither create nor appreciate art, nor will they wonder at the beautiful sunset.

Whether this is a good thing for humankind in the long run remains to be seen. The reason we dominate the natural world is not because we are faster or stronger, let alone wiser, than other animals but because we are smarter. Perhaps these learning algorithms are the dark clouds on humanity's horizon. Perhaps they will be our final invention. **M**

FURTHER READING

■ **Human-Level Control through Deep Reinforcement Learning.** Volodymyr Mnih in *Nature*, Vol. 518, pages 529–533; February 26, 2015.