

# Machines on mission possible

Affordances are ways in which an animal or a robot can interact with the environment. The concept, borrowed from psychology, inspires a fresh take on the design of robots that will be able to hold their own in everyday tasks and unpredictable situations.

Jeremy Hsu

The choreography of robotic arms working on an automobile assembly line or wheeled robots moving around tall stacks of boxes inside warehouses may make the robot labour force of the future seem unstoppable. But these commercial robots performing fairly straightforward routines in predictable environments remain a far cry from robots that can physically interact in complex ways with new and changing environments.

“I think we understand industrial robotics very well and have turned that into a huge commercial success,” says Oliver Brock, a professor in the Robotics and Biology Laboratory at the Technical University of Berlin in Germany. “The next step of robotics — that is, robots entering everyday environments shared with humans — requires something rather different.”

Unlike humans, robots have rigid and narrow ways of interacting with the world that often break down even in basic scenarios. Notoriously, these limitations were on full view in the 2015 DARPA Robotics Challenge, where a range of robots failed at seemingly mundane tasks such as climbing a ladder, turning a valve or opening a door. A growing number of robotics researchers are now exploring the idea that to make robots capable of navigating such tasks, they need to be able to appreciate the possible interactions that exist between their bodies and the physical environment.

Surprisingly, a theory from psychology developed in the 1970s by James J. Gibson — controversial at the time — is providing inspiration for this line of thought. Gibson theorized that animals can have ‘direct perception’ of their environment, meaning that the visual patterns and other sensory stimuli they perceive give them direct information about the physical properties of different objects and what they can or cannot do with them. An additional mental processing step would not be required. For instance, when walking on an uneven terrain, an animal such as a human can directly perceive where steps should go, rather than making a mental calculation of full steps.

Building on that view, Gibson developed the concept of ‘affordance’ over many years, culminating in his final book, published in 1979, *The Ecological Approach to Visual Perception*. Affordances describe the action possibilities that an environment can offer (or afford) an animal. They depend on both the physical properties of objects in the environment and the physical capabilities of an individual animal. For example, a cup would offer very different interactive possibilities to a human as it would to a dog.

The power of affordance theory comes from suggesting that animals learn to grasp the action possibilities afforded by physical objects without needing to either create a detailed three-dimensional mental model of the world or perform logical reasoning about rules-based behaviour. This raises the intriguing possibility that robots could be designed with a similarly fast-acting sense of affordance perception. But do we even understand how humans manage it?

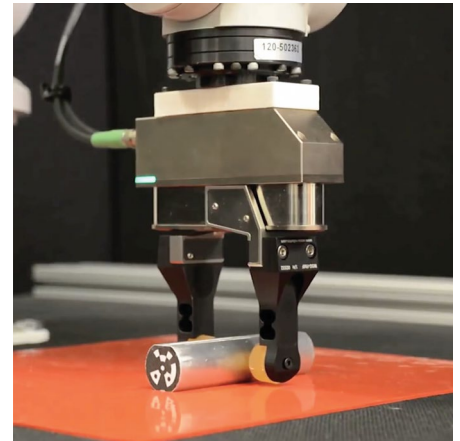
## Doorways, cliffs and slopes

Many of Gibson’s peers remained sceptical of affordance theory in the years following his death in 1979.

“Nobody wanted to touch it with a ten-foot pole,” says William Warren, professor of cognitive, linguistic and psychological sciences at Brown University in Providence, Rhode Island. “It seemed like it was mystical: maybe we could perceive the size and shape of an object, but perceiving the functional purpose of an object seemed extreme.”

In the early 1980s, Warren helped pioneer psychology experiments for studying how humans perceive affordance boundaries. First, he measured the physical dimensions of objects such as stairs and doorways. Next, he made measurements of human biomechanical capabilities. Those measurements allowed him to predict the affordance boundaries of whether or not people could climb a given set of stairs or walk through a certain doorway.

Then he put his predictions to the test and found that people indeed perceived those affordance boundaries. “People would



**Fig. 1 |** A robot grips a rod while pushing against a tabletop, using the environment to adjust its grasp. Credit: Nikhil Chavan-Dafle/Alberto Rodriguez

make judgments that mapped perfectly onto the limit,” Warren says.

Another early boost to affordance theory came from the psychologist Eleanor Gibson, colleague and wife of James Gibson. She took it upon herself to study the human learning mechanisms behind affordance perception, something that her husband never seemed interested in. Before James Gibson formally defined affordance, Eleanor Gibson conducted her famous ‘visual cliff’ experiment<sup>1</sup>, which showed how babies do not initially possess an fear of heights at younger ages. Instead, babies only gradually learn to become cautious when crawling near the edge of ledges. The study foreshadowed the modern cognitive science view of perception as a process of discovery.

Just as human babies learn to appreciate affordances through a lengthy learning process, some labs have pursued this approach to painstakingly teaching robots how to interact intelligently with different objects in their environments through physical exploration and interactions.

“If you want true intelligence in a robot, seeing how babies solve problems is a really smart way to do it,” says Karen Adolph, professor of psychology and neural science

in the Action Laboratory at New York University.

Having begun her academic career with Eleanor Gibson as her advisor, Adolph has since established a “wonderful lab with all sorts of crazy apparatuses” designed to test how human babies discover the physical limitations and possibilities of their bodies. In one series of studies, her group found that babies learn to walk by taking omnidirectional steps rather than simply walking in a straight line. That learning process allows babies to develop a highly tuned affordance perception for safely and efficiently getting around during their exploration of the world.

The learning process for walking requires a lot of data collection. In just one hour of free play, babies can walk an estimated 2,400 steps, covering the length of 7.7 American football fields<sup>2</sup>. But the payoff comes when experienced walking infants can quickly judge safe from risky ground on slopes and near drop-offs or gaps. Their affordance perception even enables them to improvise by scooting down steep slopes or making use of handrails to cross bridges.

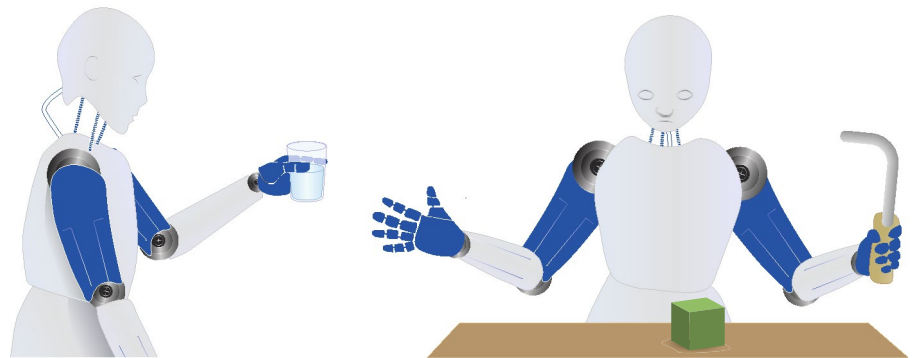
“There is very strong evidence that the thing they’re learning is the relationship between body and environment,” Adolph says.

### Pick your training data

The question is how robots could learn to perceive and act on affordances, for instance in the task of grasping and handling a wide range of everyday objects (Fig. 1). For decades, researchers have followed a standard planning-and-control model in teaching robots how to use grippers or manipulators, says Alberto Rodriguez, a professor of mechanical engineering and a robotics researcher at MIT in Cambridge, Massachusetts. First, the robot needed to be able to see or otherwise sense objects in their environment, so that it could identify the objects it might want to grab and estimate their relative position or state. That would allow the robot to plan its actions in reaching out and hopefully grabbing the object it wants to hold.

But that standard approach has trouble dealing with complicated environments filled with uncertainty. “It relies on the assumption that each step in the process provides good enough information to the next step so the robot can make intelligent decisions,” Rodriguez says.

The Amazon Robotics Challenge (also called Amazon Picking Challenge) has shown the limitations of this classic approach. During the first year of the challenge in 2015, robotic competitors received advance information about all possible objects to pick from a tote bag and



**Affordances are perceived directly by animals (and potentially by robots) as action possibilities in the environment.** The cup of liquid affords grasping and manipulation. The hand-held tool affords different kinds of behaviour upon the cube. The ladder with steps rising a certain height affords climbing. In all of these cases, there is a dynamic interaction between specific environmental features and the particular capacities of the organism.

store away so that they could train their object recognition software ahead of time. In the 2016 challenge, the difficulty ramped up as objects could be presented in more complicated arrangements with different objects lying on top of each other. In 2017, half the objects were shown only 45 minutes before the competition so competitors had to quickly learn how to pick things out that they had never seen before<sup>3</sup>. In these challenging conditions, the standard planning-and-control model typically breaks down.

But luckily robotics researchers had a new artificial intelligence tool to help robots handle new objects: deep learning.

“Some of the key ideas that have been explored in the machine learning community as it relates to robotic manipulation came to save the day,” Rodriguez says. “Because even if they don’t necessarily speak in the language of affordance, that is basically what ends up happening.” In deep learning, neural network algorithms are trained to recognize

subtle patterns in many thousands or millions of examples in huge datasets. Rather than always having robots physically fumble through real-world training sessions, researchers are building computer models that categorize affordances by looking at many different human-labelled examples in images or videos.

“Deep learning is aligned with affordances as it’s about fairly straightforward input/output mapping without the need for having rich, internal models of the world”, says Peter Battaglia, a research scientist at DeepMind, the London-based AI company owned by Google’s parent company, Alphabet. He suggested that “Gibson would be very happy with this trend in deep learning”.

The combination of deep learning and ‘big data’ could help AI incorporate affordances in a number of ways. Working in the cognitive and brain science lab of professor Michael Tarr at Carnegie Mellon University in Pittsburgh, Pennsylvania, PhD student Aria Wang has been training deep learning models on affordance examples that are hand-labelled by humans. The models can then attempt to identify and categorize affordances in new images based on the intermediate visual features of affordances.

“For now, we have a working model that can tell different affordances apart based on images and objects,” Wang says. “If there is any room, we can improve the model by getting more human-labelled data.”

A combined Princeton–MIT team led by Rodriguez won first place in the ‘stowing’ task of the 2017 Amazon Robotics Challenge by training a robot beforehand to recognize the best affordance-related ways to grasp any general object — a robust strategy that bypassed the need for additional training to recognize and pick up specific objects present in the competition. The researchers trained deep learning models on many different product images to identify the most graspable points for a robot using either a suction cup or a two-fingered gripper to grab things.

“Rather than focusing on the object, you’re focusing on the action itself,” Rodriguez says. “This is where the concept of affordance comes into play: what things in the environment have the right features for my action to be useful?”

There are limits to training computer vision models to recognize affordances. For example, supervised deep learning methods are only as good as the training data examples that humans have labelled by hand, Tarr points out. The computer models won’t necessarily be better than humans at categorizing affordances, and they probably will not yield detailed

insights about the exact mechanisms behind affordance perception. But using deep learning to discover underlying visual features of affordances, as perceived by humans, could potentially benefit both robots and other machines that rely on AI to navigate the world.

Along those lines, Tarr’s lab, in collaboration with others, have begun investigating the possibility of using human brain activity from people looking at affordance-related images to help train the computer models. The researchers collected functional magnetic resonance imaging (fMRI) data from human subjects who were asked to observe over 5,000 images from three different image datasets. The resulting **BOLD5000 dataset** is available online and eventually the plan is to crowdsource across different labs to scale up the dataset to 50,000 images.

### Discovering the world

Still, some robotics researchers remain uncertain how far computer vision models can go in training robots to intelligently interact with the world. Such scepticism stems in part from appreciating the limitations of generalizing lessons learned from one training dataset to a completely new situation. A robot may need to become adept at interacting with an object from a constantly changing perspective under different lighting conditions and possibly with the robot’s own arms or hands blocking part of the view. That kind of learning could benefit from acts of ‘continual discovery’ that involve direct physical interactions with objects rather than just looking at two-dimensional images, says Angelo Cangelosi, professor of machine learning and robotics at the University of Manchester in the UK.

The problems with over-reliance on computer vision have also become clear in autonomous vehicles, says Serge Thill, a cognitive scientist and associate professor of artificial intelligence at Radboud University in Nijmegen, the Netherlands. Companies such as Google’s Waymo, Tesla and Uber have generally trained their self-driving cars on both driving simulations and huge image datasets taken from the real world. But the brittleness of deep learning systems means that autonomous vehicles can run into trouble even if a few image pixels are changed — something that can lead to tragedy if the vehicles misclassify a person crossing the street at night as being some other object.

“I think we’ll find that the more resilient approaches involve robots discovering possibilities and interactions with the environment,” Thill says. “Deep learning will work nicely for two or three years and then we will hit the glass ceiling.”

Many AI researchers tend to underestimate the challenges of transferring the results of simulation training to physical robots, whereas “roboticists understand this through blood and sweat,” says Battaglia at DeepMind. He also readily acknowledges the limitations in how deep learning can help AI and robots develop affordances that go beyond each specific task at hand. “You don’t find that when something can pick up blocks in a simulation, it can also pick up forks and fly paper airplanes,” Battaglia says.

Still, even a simulation that does not perfectly capture all the details and complexities of real-world physical interactions can provide guidance for robotic actions, Battaglia notes. “My personal belief is the underlying backbone we’re gaining with simulated systems should be useful in real robotic applications.”

### Can machines match biology?

Whether robots learn affordances by analysing millions of examples or through actual physical exploration, they will almost certainly run up against a wall at some point, says Michael Turvey, an emeritus professor in psychology at the University of Connecticut in Storrs, Connecticut, and an early pioneer in developing Gibson’s affordance theory. He views machine-based computation as necessarily involving intermediary steps and representations based on mathematics and algorithms — something very different from the biological capabilities of direct perception.

“Building robotic artefacts is one of our current genius moves,” Turvey says. “At issue is [whether we are] in contact now with the right theory of perception and action, and the answer my colleagues would give is no.”

One of the biggest stumbling blocks comes from a lingering mystery: what information do humans and animals use in perceiving affordances? It’s more complicated than just eyeballing the height of a chair or step, or estimating the width of a doorway, because many affordances can involve ever-evolving relationships between each person’s body and changing environments. “We don’t really have a handle on how people learn those dynamic properties and perceive affordances,” says William Warren, the cognitive scientist at Brown University.

Biological brains also remain years ahead of AI and robots in their ability to grapple with affordances and complex physical interactions. Human babies may require some time to figure out how to interact with their world through trial and error, but AI agents and robots face a far steeper learning curve, even with access to potentially millions of simulation trials and dataset

examples. That's in large part because humans and animals have millions of years of built-in biological advantages over machines in understanding how to interact with the world.

"We've kind of forgotten we carry in our DNA a lot of prior knowledge from our ancestors and other organisms," says Wojciech Zaremba, head of robotics research and co-founder at OpenAI, a non-profit research company in San Francisco, California. "[Humans have] many, many years of solving a variety of tasks, plus hundreds of millions of years of solving other tasks in the jungle."

Many researchers doubt that robots can catch up anytime soon with humans in terms of intelligent interactions with their environments, given that even the most multi-layered deep neural networks pale in comparison with the human brain's trillions of connections. But Zaremba takes a somewhat more optimistic view based on the idea that computer simulations can close the gap between humans and machines.

The OpenAI team has shown how to transfer grasping and manipulation skills from simulations to a physical five-fingered robot hand<sup>4</sup>. Researchers trained the system on many simulated runs involving randomized conditions for the main task of orienting a colourful block to face a certain way. This 'domain randomization' strategy

is similar to the notion that "after you drive enough cars, you can drive any car," Zaremba says.

But the experience of the robotic hand with a diverse set of simulation 'universes' allowed it to do more than just maintain consistent, good-enough performance in real-world testing. It could also choose optimal grasping and manipulation solutions for any given situation for the block orientation task. Just as impressively, the robotic hand's simulation-trained behaviour began mimicking natural human hand behaviours — all without direct guidance from the human researchers.

Beyond mastering individual skills for general use, another big step for robots would be the capability of incorporating long-term action planning and foresight into each individual action, says Rodriguez, the robotics researcher at MIT. For example, asking a robot to use a wrench or screwdriver would require the robot to not only select the right grasp for picking up a tool, but also know how to adjust the grasp with the tool in hand to perform a specific task. Even telling the robot to fetch a book from another room would require appreciation of long-term implications for each action.

How to imbue robots and other artificial agents with such knowledge or capabilities that allow for direct perception in different environments remains an open research question. "Affordances are like a shortcut

where the knowledge is embedded into the system," says Paul Schrater, a researcher in psychology and computer science at the University of Minnesota in Saint Paul, Minnesota. "If we want to allow for creative agents that solve problems, we need to do a much better job of principled embedding. Affordances are the deep solution to the problem of how do we take knowledge and embed it in a way that controllers can use it effectively."

Karen Adolph from NYU echoes this view. "I agree with Eleanor Gibson: affordance is the most central concept for learning and perception and development and action — it pulls it all together. It's a really powerful concept that can be super informative for AI." □

### Jeremy Hsu

Freelance science writer New York, NY, USA.

e-mail: [jmichael.hsu@gmail.com](mailto:jmichael.hsu@gmail.com)

<https://doi.org/10.1038/s42256-019-0034-3>

### References

1. Gibson, E. J. & Walk, R. D. *Sci. Am.* **202**, 67–71 (1960).
2. Adolph, K. E. et al. *Psychol. Sci.* **23**, 1387–1394 (2012).
3. Leitner, J. *Nat. Mach. Intell.* <https://doi.org/10.1038/s42256-019-0031-6> (2019).
4. OpenAI et al. Preprint at <https://arxiv.org/abs/1808.00177> (2019).

### Competing interests

The author declares no competing interests.