

## ACTION RECOGNITION HAPPENS VERY QUICKLY IN THE HUMAN BRAIN

**Andrea Tacchetti<sup>1</sup> and Leyla Isik<sup>2,3\*</sup>**

<sup>1</sup> DeepMind Technologies Limited, London, United Kingdom

<sup>2</sup> Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, United States

<sup>3</sup> Department of Cognitive Science, Johns Hopkins University, Baltimore, MD, United States

### YOUNG REVIEWERS:



DUKE  
NEUROSCI-  
ENCE CAMP  
2018

AGES: 9-11

“What is that person doing?” Without even thinking about it, your brain is processing what you see and using it to recognize other people’s actions, like running and walking. Recognizing actions comes easily to humans, but it is an extremely challenging problem for computers to solve. In particular, it is difficult for computers to recognize the same actions from different viewpoints, like front or side views, or against different backgrounds. As a result, scientists and engineers have not been able to develop computer algorithms that can recognize actions as well as humans can. Our research goal is to better understand how the human brain recognizes actions so we can replicate those functions in computer systems.

Using MEG, a device that allows us to read brain signals, and machine learning, a branch of computer science, we were able to decode the signals that occur in the human brain when it recognizes an action.

Because MEG detects very rapid changes in brain signals, it allows us to understand how action information arises over time. We found that the brain recognizes actions very quickly, within 200 milliseconds (ms, 1/1,000 of a second). Further, the brain immediately ignores changes in viewpoint, like front view to side view, and still recognizes an action just as quickly. By understanding the order in which the brain breaks down this complex problem, we have learned many important things about the brain computations involved in human action recognition.

## THE PROBLEM

Recognizing other people's actions is critical to our everyday lives. We can easily tell when someone is waving at us or running away from something. This helps us decide what we should do next. Recognizing actions is also becoming increasingly important for computer systems to recognize actions as scientists develop artificial intelligence (AI) that must interact with humans, like assistive robots and self-driving cars.

Remarkably, humans can recognize actions even when there are complex **transformations**: image changes that alter the way a scene looks, but do not meaningfully change its contents. In Figure 1, for example, it is easy to recognize the actions in each image, despite the many different changes: different actors (children, a dog, adults), different

### TRANSFORMATIONS

Image changes that alter the way a scene looks, like changes in viewpoint, but do not meaningfully change its contents.

#### Figure 1

Actions take many forms. (A, B) People can easily recognize different actions, like running, regardless of who is performing them, the scenery in the background, or viewpoint. (C) Still frames from videos used in our experiment of people performing different actions from different viewpoints. The challenge for computers is to tell apart visually similar actions—like the same person walking vs. running, and also recognizing the same action when it is performed by different people and seen from a different viewpoint.

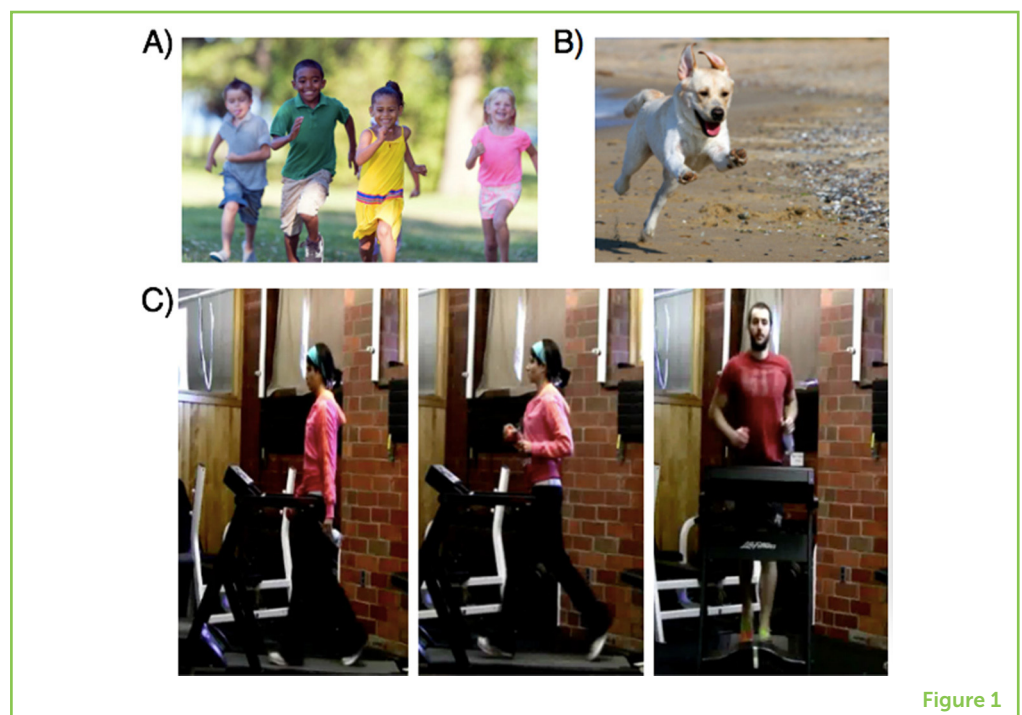


Figure 1

camera locations (in front or on the side of the actor), and different backgrounds (a park, the beach, a room with a brick wall).

In particular, take the three pictures in the bottom row of Figure 1. If you compare them, the first two pictures look much more similar to each other than the second two pictures. However, despite their visual similarity, you can easily tell that the first two images depict different actions (running vs. walking). We call this “**discrimination**.” You can also tell that the second two pictures contain the same action (both running), despite how different the pictures look. We call this “**generalization**.” Being able to both discriminate and generalize is difficult for even state-of-the-art computer algorithms. An **algorithm** is a set of rules or calculations that a computer follows to solve problems. Discrimination and generalization are carried out by separate processes in the brain. How are these different steps broken down by the brain to recognize actions? Answering this question will help us identify the ideal way to implement these steps in computer algorithms and develop better AI.

In this study, we asked when the human brain recognizes different actions, and when it does so in a way that generalizes across different transformations. Does the brain first identify actions and then link the same action across different views? Or are discrimination and generalization solved at the same time? By knowing what the brain computes when, we can better understand its underlying processes.

## WHAT WE DID AND HOW WE DID IT

### Action Recognition Video Dataset

For our study, we filmed videos of different people doing five common, everyday actions: running, walking, jumping, eating, and drinking. We filmed the videos from two different views: forward facing and side facing (see Figure 1, bottom row). This allowed us to ask if the brain signals can discriminate between the different actions, and generalize across the different views. We kept other elements of the videos, like the lighting and background, constant. We next needed a method to measure the brain’s response to these different videos.

### “Mind Reading” with a Technique Called Magnetoencephalography (MEG)

The human brain is constantly processing lots of information. One major challenge in studying the human brain is reading brain signals

#### DISCRIMINATION

The ability to tell different things apart (like recognizing running vs. walking in the case of actions).

#### GENERALIZATION

The ability to recognize the same thing despite visual changes or transformations.

#### ALGORITHM

A set of rules or calculations that a computer follows to solve a problem.

## Figure 2

(A) MEG measures brain activity by detecting the change in magnetic field produced by millions of brain cells (neurons) firing at the same time. These brain cells firing creates an electric current, which leads to a change in magnetic field. (B) A person sits in the MEG machine, which contains sensors in the helmet around the person's head. (C) Drawing of a MEG helmet, which has 306 sensors evenly distributed across the head. Each sensor measures the surrounding change in magnetic field.

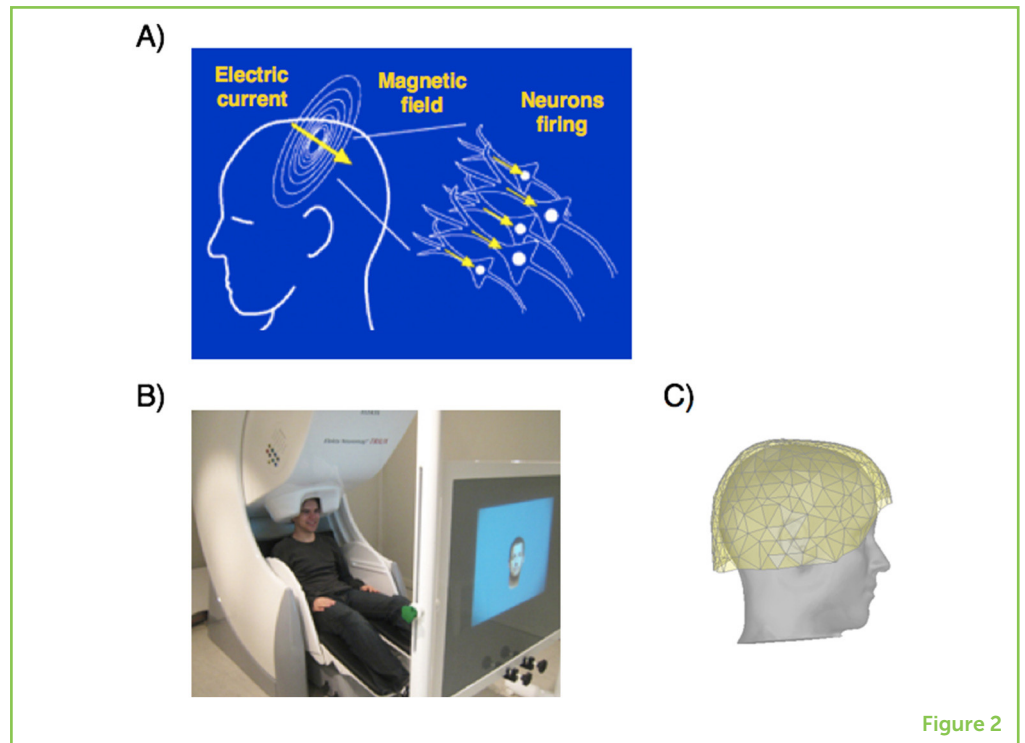


Figure 2

in a non-invasive manner: without having to cut open a person's skull and record from inside of their brain.

In our study, we used a tool called magnetoencephalography or MEG (Figure 2). MEG detects changes in magnetic fields and that is why it is called "magneto" encephalography. When millions of brain cells are activated, or "fire" together, they create a change in the magnetic field around the head. These are very weak changes that are much weaker than the earth's magnetic field, but we can still detect them using MEG, because it is very sensitive.

The MEG has a helmet with 306 sensors that measure the brain's weak magnetic activity. Unlike other methods of measuring brain activity, MEG allows us to collect lots of data very quickly: we can get a new MEG measurement 1,000 times a second. The tradeoff is that we cannot tell exactly which areas of the brain the signals are coming from. As a result, we use signals from the entire brain. These signals are very complex and we need clever ways to analyze them.

We used **machine learning** to detect patterns in our MEG data (see Box 1). We did this by using MEG on many people while they watched the videos of the actions we mentioned earlier, so that the computer algorithm could learn what the pattern of brain activity looked like in response to each action. This is called **training**. Then, once the algorithm had learned, we presented a new pattern that the algorithm had

## MACHINE LEARNING

A branch of computer science whose goal is to get computers to learn complex tasks based on data.

## TRAINING

The machine learning procedure used to find a classifier with one set of data.

**Box 1**

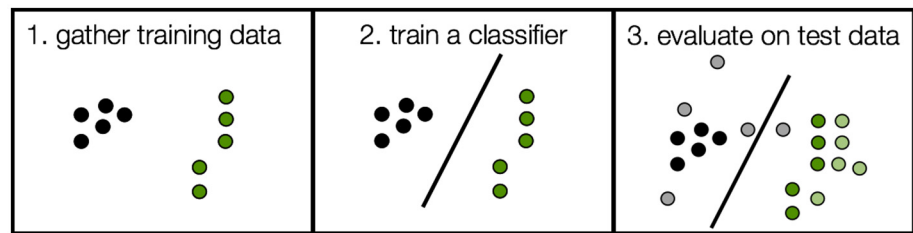
What is machine learning?

**CLASSIFIER**

A machine learning algorithm used to solve tasks by splitting data into different categories.

**TESTING**

The machine learning procedure to evaluate your classifier with a new set of data (not used in testing).



**Machine learning** is a branch of computer science whose goal is to get computers to learn complex tasks based on data. For example, we could give the computer green dots and black dots—these data are called **training data**. We can then ask the computer to find the line that best separates them. This line is called a **classifier**. We can then use it to “classify” new data points as green or black. These new data points are called **test data** (the colors correspond to the training data color, but are illustrated in a lighter shade), and the computer has never seen them before. In the above picture, they are the lighter shaded dots in panel 3. We can see how accurate the classifier is by how many of the test data points it classifies correctly. In our case, that would be 9 out of 10 or 90%. If the classifier were to make random guesses, it would be at chance. In this case chance is 50% (one out of two), because there are two different classes.

In the experiment described below, instead of green vs. black, the classifier is trained to distinguish the five actions from each other. Chance in this experiment is 20% (one out of five). We repeat this training and test procedure at each time point to get a measure of classification accuracy over time. We use the classification accuracy as a measure of whether the brain signals can tell apart the different categories (green vs. black, or running vs. walking).

Box 1

not seen before, called **testing**, and asked our algorithm, “What video was the person watching?” In other words, which pattern in our training data is this test data most similar to? This is sometimes called “neural decoding.” It allows us to read out which action someone was viewing from the person’s brain activity. We use the readout accuracy as a measure of whether the brain signals can tell apart different actions.

We repeat this procedure each time we get a new MEG reading (once every ms). This allows to see how information in the brain changes over time.

**WHAT WE FOUND****The Human Brain Rapidly Recognizes Actions**

Subjects sat in the MEG and watched the videos. We used our machine learning system to predict, at each time point, which action the subjects viewed—running, walking, jumping, eating, or drinking. If the algorithm’s classification accuracy is above chance, the brain signals we recorded contain information that can discriminate between actions.

To start, we looked at videos of all the same view (for example, front facing only). Before a video starts at time zero, our prediction is about 20% (or 1 out of 5), the same as chance, because there is nothing on the screen (Figure 3, blue curve) and the classifier is randomly guessing. Once the video starts, we see a dramatic rise in the classifier's accuracy, indicating that there is action information present in the brain signals. The classifier's accuracy peaks at around 40%. For our purposes, we are interested in when the action information is first present in the MEG signals, so we look for the first time when decoding is significantly above chance (based on statistical tests). The classifier's accuracy reaches this level beginning at 200 ms after the video starts. This is extremely fast, and after only six frames of the video.

### The Brain Recognizes Actions Across Different Views

In the previous experiment, our machine learning algorithm was trained and tested on data from subjects viewing videos of actions from one view, such as from the front only. We ultimately want to understand when the brain can generalize across the different viewpoints. To do this, we trained our algorithm on training data from the subjects watching the videos from one view and tested the algorithm on the second view.

Just like the earlier experiment, before the video starts at time zero, the classifier is randomly guessing (accuracy is at chance, 1 out of 5 or 20%). Once the video starts, we can again predict what action the subject was watching beginning 200 ms after the video begins (Figure 3, red curve). This means that the brain signals can generalize across the different views. Not only does the brain quickly recognize different actions, but it does so in a way that can ignore changes in view. Interestingly, these two opposing goals, discrimination and generalization, happen at the same time. We know this because the red and blue curves are on top of each other showing that as soon as the brain recognizes actions,

#### Figure 3

Reading action information from the human brain. This plot shows the classifier's accuracy (the percent of trials correctly classified) at reading out action from our subjects' brain activity (y-axis) over time (x-axis). Time 0 indicates when the video starts. Beginning 200 ms after the video starts (black arrow), we can tell which action someone was viewing at an accuracy significantly above chance. This is true if we are looking at actions all in the same view, like from the front (blue) and even across different views (red).

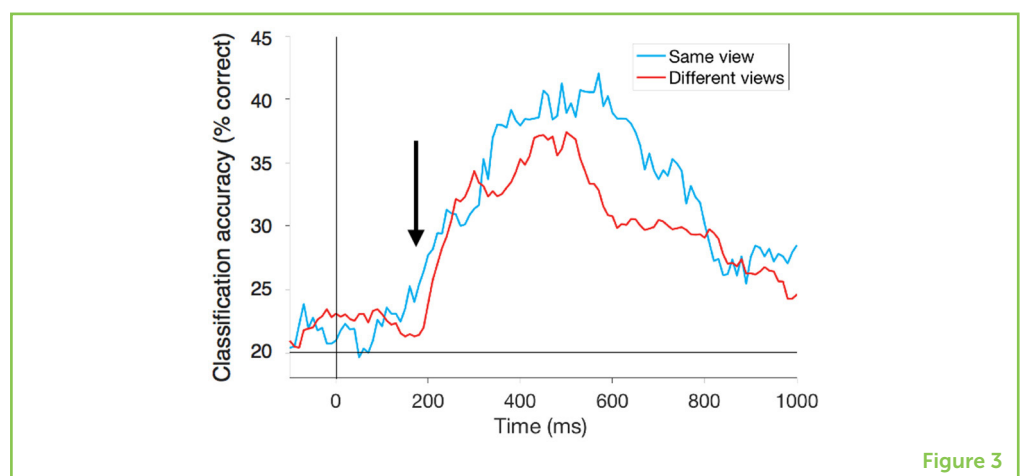


Figure 3

it can ignore changes in viewpoint. This finding suggests that both discrimination and generalization happen in one step in the brain, and therefore it may make sense to implement them at the same stage of a computer algorithm.

## **STUDYING HOW THE HUMAN BRAIN RECOGNIZES ACTION CAN HELP TO CREATE MORE EFFECTIVE ARTIFICIAL INTELLIGENCE**

Action recognition is a crucial aspect of humans' daily lives, and a skill that is necessary for AI systems to safely and meaningfully interact with humans. In our work, we identified the brain signals that discriminate between different actions, and showed that these signals arise very quickly, after seeing only 200 ms of video. Importantly, these brain signals must be flexible enough to work under different conditions. We showed that these brain signals immediately generalize across different viewpoints, also at only 200 ms.

This timing information can be used to help computer algorithms to recognize actions in a better, more human-like way. The speed of action recognition that we saw, for example, tells us that the brain's algorithm is relatively simple and most likely processes the visual input only once (meaning there is no need to remember the video to recognize the action or process it multiple times). The fact that the brain can recognize the same actions from different viewpoints just as quickly tells us that computer systems can discard viewpoint information in the early stages of processing.

Understanding how the brain performs action recognition is a first step toward a complete understanding of human intelligence and can guide the development of better AI systems.

## **ORIGINAL SOURCE ARTICLE**

Isik, L., Tacchetti, A., and Poggio, T. 2018. A fast, invariant representation for human action in the visual system. *J. Neurophysiol.* 119:631–40. doi: 10.1152/jn.00642.2017

**SUBMITTED:** 02 July 2018; **ACCEPTED:** 16 January 2019;

**PUBLISHED ONLINE:** 08 February 2019.

**EDITED BY:** Michael A. Yassa, University of California, Irvine, United States

**CITATION:** Tacchetti A and Isik L (2019) Fast action recognition in the human brain. *Front. Young Minds* 7:9. doi: 10.3389/frym.2019.00009

**CONFLICT OF INTEREST STATEMENT:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**COPYRIGHT** © 2019 Tacchetti and Isik. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

## YOUNG REVIEWERS

### DUKE NEUROSCIENCE CAMP 2018, AGES: 9-11

Duke Neuroscience Camp is a part of Duke Youth Programs. We accept applications from high school students from all over the world. Anyone can join, and no prior experiences are required. The only thing needed is an interest in neuroscience! This year we frequently discussed minds and machines, and as a group we enjoyed applying what we learned about brain-machine interfaces and human-robot interaction when reviewing Drs. Tacchetti and Islik's article.



## AUTHORS

### ANDREA TACCHETTI

I want to build a world where computers and robots help us take care of ourselves and of each other. Computers already help doctors figure out why we are sick; soon they will drive our cars and help us take care of the elderly. I work for DeepMind in London where I build computer systems that are capable of understanding what humans and other robots are doing, and what they will be doing next. If I am successful, autonomous systems will operate seamlessly in our complex social environment.



### LEYLA ISIK

I am an Assistant Professor of Cognitive Science at Johns Hopkins University. I am fascinated by the link between human and artificial intelligence (AI), and in particular understanding the problems that are easiest for humans (like seeing and moving around in the world), but the most difficult for AI. The goal of my research is to understand how you recognize information about people based on what you see. Outside of research, I love to run, cook, and spend time outdoors. \*lisik@mit.edu

