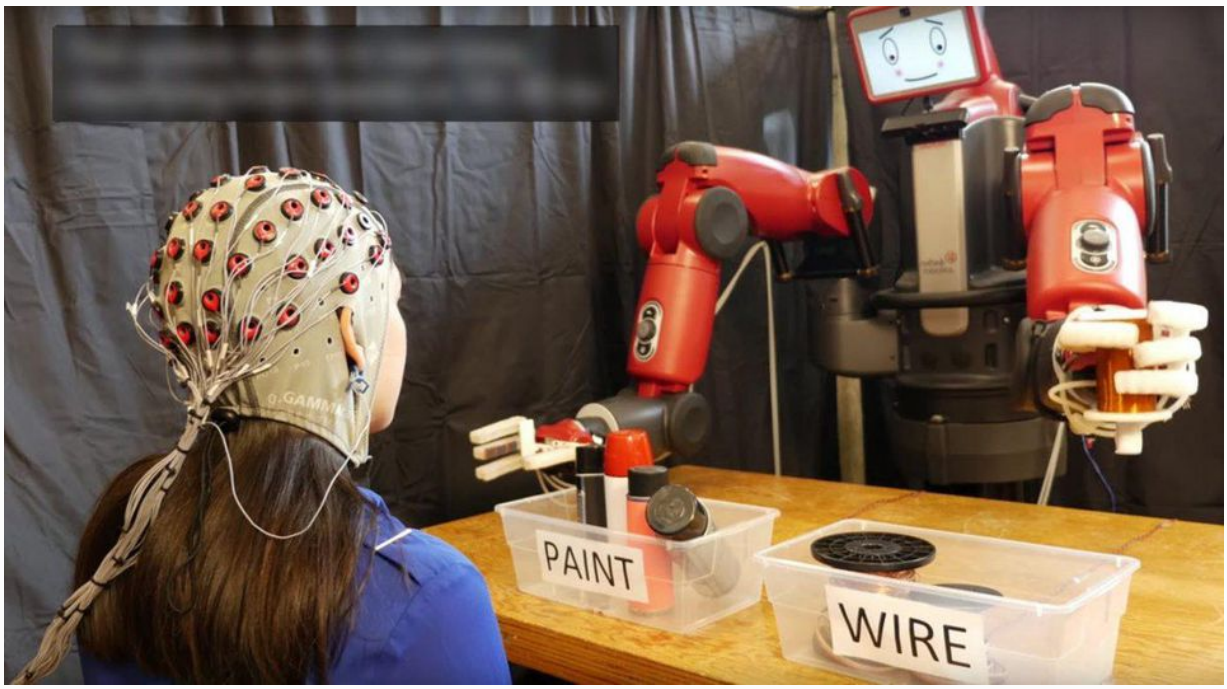
EDITORS' PICK | Mar 6, 2017, 10:00am EST

Robot Makes Mistake, Human Notices And Does Nothing, Robot Responds And Corrects Mistake

Kevin Murnane Former Contributor
Consumer Tech

This article is more than 3 years old.



Which box does this go in? CREDIT: MIT CSAIL/YOUTUBE

Last week I wrote about how science infuses science fiction in the Syfy network series *The Expanse*. Now comes a story about how the researchers at MIT's CSAIL (Computer Science and Artificial Intelligence Laboratory) are turning science fiction into science.

The researchers created a system that allows a robot to correct its mistakes in real time when a human observer notices that a mistake is being made.

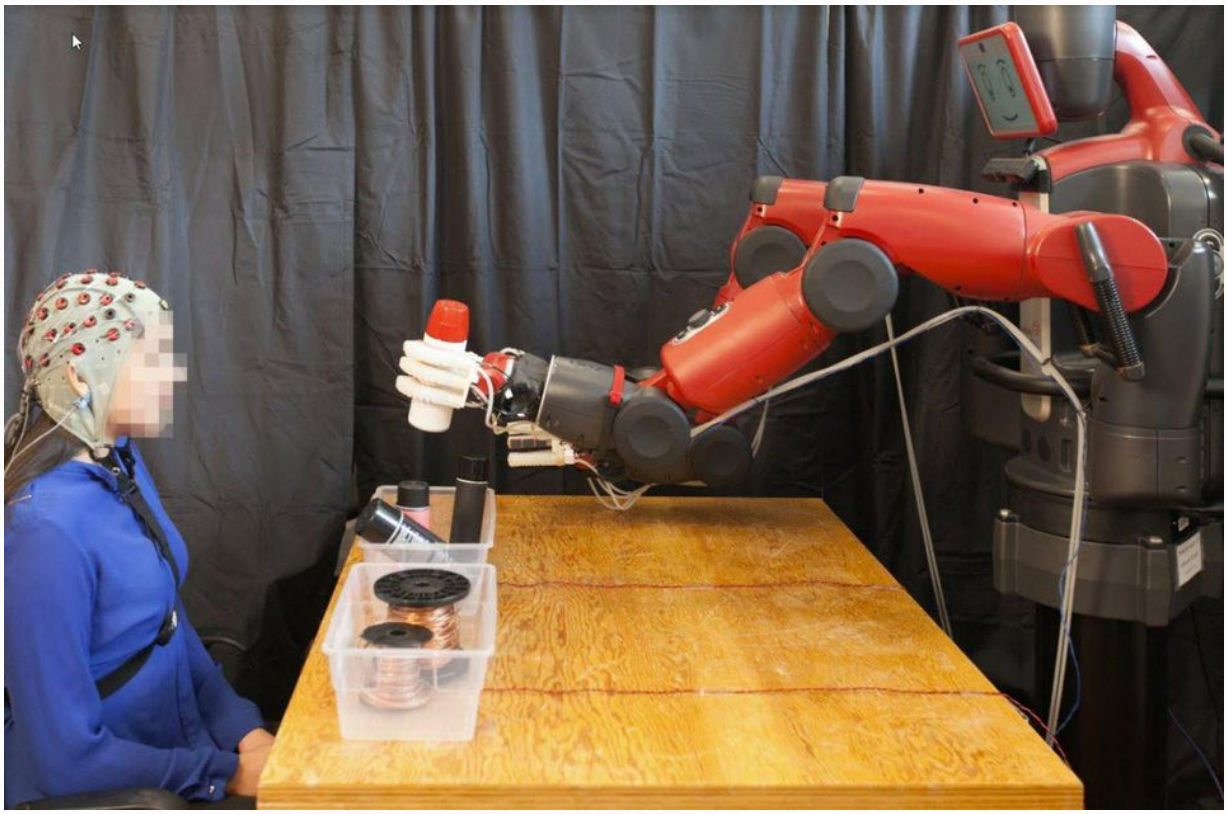
The observer sat across a table from a Baxter robot made by Rethink Robotics. Situated on the table between the two were two LED-equipped targets, one to the right and the other to the left. Baxter's task was to reach for a target when its LED lit up. The robot's arm rested midway between the two targets and when the LED came on it started to reach toward one or the other. If Baxter reached for the wrong one and the EEG signal from the observer included an ErrP, the robot received a message that switched it to reaching for the other target.

Electrode cap CREDIT: JASON DORFMAN/MIT CSAIL

The classifier is a potential weak link in the system. If it fails to identify the ErrP signal, the robot will not correct its mistake. Did the researchers build in a backup to protect against an initial classifier error? No, they didn't have to because human observers come equipped with a fallback routine.

When humans see that that a previously noticed error continues, a second ErrP - called a secondary error – occurs. If the classifier missed the first ErrP signal, it still has a chance to pick up the second which will send a message to the robot to correct its mistake.

The 12 human observers in the study didn't receive any training and only one of them had previous experience with EEG. However, different people produce ErrP signals that are similar but not identical and the classifier needed training to recognize each observer's unique ErrP. In order to train the classifier, each observer went through 200 trials in which the robot reached for one target or the other. Ten of the trials were used for testing after the other 190 were used for training.



The observer and the robot CREDIT: JASON DORFMAN/MIT CSAIL

Accuracy was measured in terms of how frequently the robot reached for the correct target. If the EEG-based error correction system failed completely, random chance would result in Baxter making the correct choice on 50% of the trials. Based on the first ErrP signal alone, the robot was accurate 64.88% of the time. Robot mistakes were successfully corrected in real time when a human observer noticed a mistake was happening.

The initial ErrP signal is more difficult to detect and varies more widely across individuals than the second ErrP. In cases where the initial ErrP

signal was not identified correctly and the second ErrP came into play, the robot was accurate 78.49% of the time.

When you stop and think about it, these results are remarkable. Keep in mind that this is basically a proof-of-concept study. The robot's task was very simple, and neither the technology used in the study nor the training of the classifier program benefited from extensive refinement after years of focused testing. Nevertheless, almost 80% of the time a robot was able to correct a mistake based on nothing more than a human observing that a mistake had been made.

Clearly, a lot of work needs to be done but imagine the possibilities. You say "Alexa, play *The Expanse*" and she cranks up a trance playlist. As soon as you hear the first few notes and realize she screwed it up, Alexa corrects and turns on the TV.

Not speaking. CREDIT: TUNISU/PIXABAY

How cool would that be? How about this one. You say "Alexa, play that fantasy movie with Charlize Theron" and *Mad Max: Fury Road* appears on your screen which is the film you wanted to see. Your partner thinks it's a mistake because she wanted to watch *Snow White and the Huntsman,* the robotic system auto-corrects and now her preference appears on the screen.

You think that's a mistake and *Mad Max* comes back; she thinks it's another mistake and here comes the *Huntsman*. The screen wears itself out switching back and forth while you and your partner engage in a silent battle of wills.

The robot notices that a human problem has occurred and an error-correction message appears on the screen. Do the humans auto-correct?

Follow me on Twitter or LinkedIn. Check out my website.

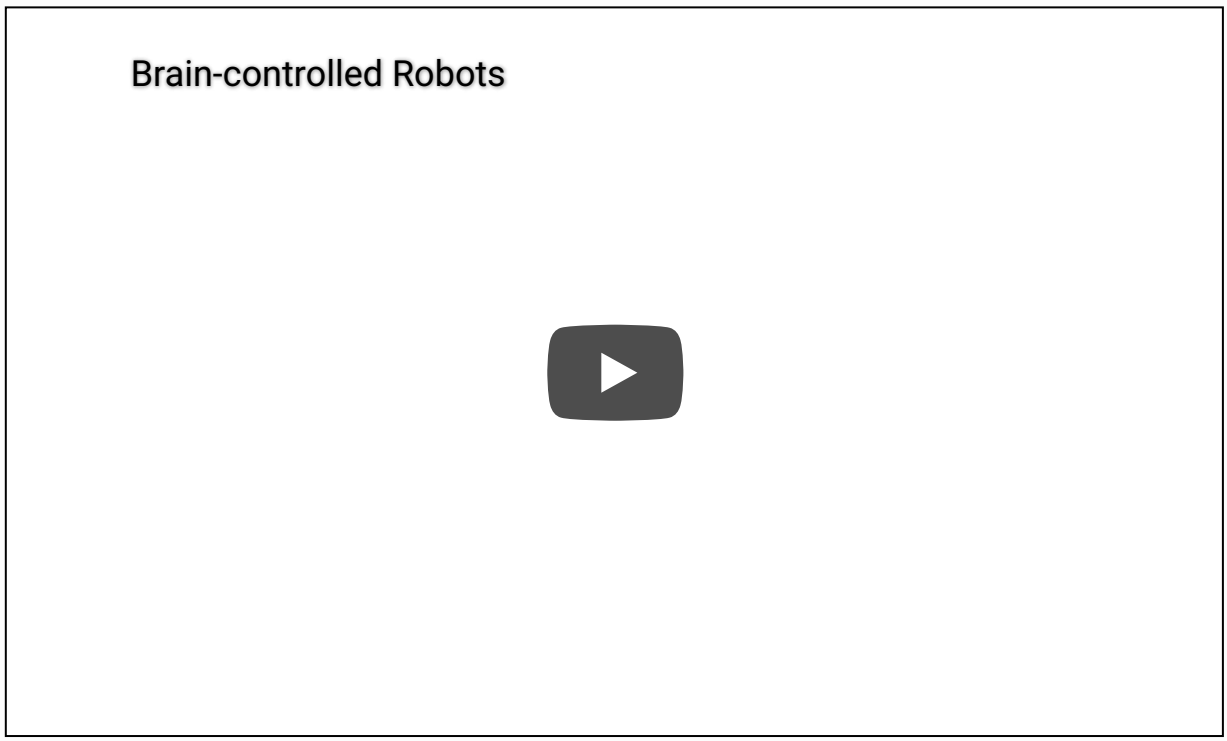
Kevin Murnane

Follow

I'm a cognitive scientist, retired professor, musician, gamer, and avid cyclist with a B.A in History, an M.S. in History and Philosophy of Science, and a Ph.D. in... **Read More**

The observer just sits there and watches; she doesn't physically interact with the robot or anything else. If the observer recognizes that a mistake is occurring, the robot changes course and does the right thing.

How in the world did the researchers at CSAIL achieve this? It's ingenious, really. EEG (electroencephalogram) data are gathered from the observer who wears a cap of electrodes that measure patterns of electrical activity in the brain. One distinctive pattern, called an ErrP (Error-related Potential) signal occurs whenever a person recognizes that either she or someone else is making a mistake. The EEG data are analyzed in real time by a classifier program that is trained to recognize ErrP signals. When an ErrP is identified, a message is sent to the robot telling it to change its course of action.



The researchers tested their system as follows. (The study described in the research paper is different from the one depicted in the video.)