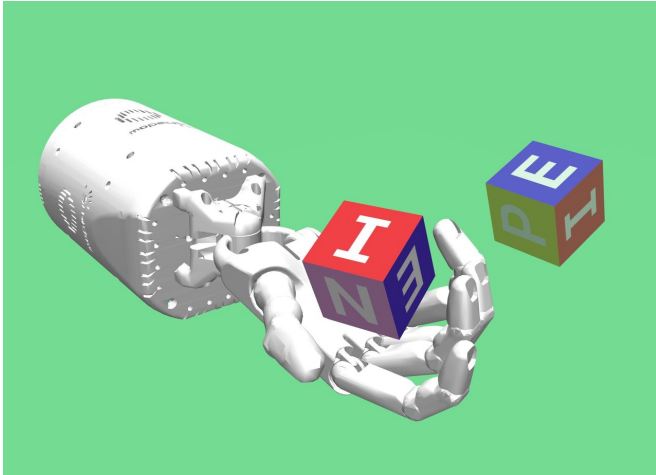


# Focus on a reinforcement learning algorithm that can learn from failure

2 March 2018, by Nancy Owano



Credit: OpenAI

Recent news from the OpenAI people is all about a bonus trio. They are releasing new Gym environments—a set of simulated robotics environments based on real robot platforms—including a Shadow hand and a Fetch research robot, said *IEEE Spectrum*.

In addition to that toolkit, they are releasing an open source version of Hindsight Experience Replay (HER). As its name suggests, it helps robots learn from hindsight, for goals-based robotic tasks.

Last but not least, they released a set of [requests](#) for robotics research. "If you're an ambitious sort," said Evan Ackerman in *IEEE Spectrum*, "OpenAI has also posted a set of requests for HER-related research."

"Although HER is a promising way towards learning complex goal-based tasks with sparse rewards like the robotics environments that we propose here, there is still a lot of room for improvement," they blogged. "Similar to our

recently published Requests for Research 2.0, we have a few ideas on ways to improve HER specifically, and reinforcement learning in general."

OpenAI is an AI research company. They publish at machine learning conferences and their blog posts communicate their research.

Elon Musk is a co-founder. It's sponsored by individuals and companies, and they aim to discover and enact "the path to safe [artificial general intelligence](#)."

An OpenAI video showing what they accomplished in the Gym environments portion was published Feb. 26.

They show the different tasks accomplished. A ShadowHand [robot](#) manipulates an object (shows a hand manipulating, including flexing fingers, a child's alphabet block, an egg-shaped object, and passing fingers through a small stick). They are also introducing a robot "nudge" robot mechanism that can slide a puck as well as grasp a small ball and lift it up

Specifically, these are the varied feats on show: ShadowHand has to reach with its thumb and a selected finger until they meet at a desired goal position above the palm. ShadowHand has to manipulate a block until it achieves a desired goal position and rotation. ShadowHand has to manipulate an egg until it achieves a desired goal position and rotation. ShadowHand has to manipulate a pen until it achieves a desired goal position and rotation.

All in all, "the latest environments simulate a Fetch robotic arm to [push](#) stuff around, and a ShadowHand to grip and manipulate things with robotic fingers," said Katyanna Quach in *The Register*.

The OpenAI HER offering is especially interesting;

training and reinforcement gets a rethink. HER allows an agent to learn from failures. As Ackerman wrote, HER "reframes failures as successes in order to help robots learn more like [humans](#)."

Jackie Snow in *MIT Technology Review* observed that "It does that by looking at how every attempt at one task could be applied to [others](#)."

Snow added, "HER doesn't give robots rewards for getting a step of a task right—it only hands them out if the entire thing is done properly."

Reframing failures as successes? Ackerman offered this explanation: "To understand how HER works, imagine that you're up to bat in a game of baseball. Your goal is to hit a home run. On the first pitch, you hit a ball that goes foul. ...you've also learned exactly how to hit a foul ball...With hindsight experience replay, you decide to learn from what you just did anyway, essentially by saying, 'You know, if I'd wanted to hit a foul ball, that would have been perfect!'"

How good is the HER implementation? "Our results show that HER can learn [successful](#) policies on most of the new robotics problems from only sparse rewards."

Kids playing blindfold games often tell the player, "You're getting warm, warmer." Key words in appreciating their research are sparse and dense rewards.

"Most reinforcement learning algorithms use 'dense rewards,' explained Ackerman, "where the robot gets cookies of different sizes depending on how close it gets to completing a task...Sparse rewards mean that the robot gets just one cookie only if it succeeds, and that's it: Easier to measure, easier to program, and easier to [implement](#)."

© 2018 Tech Xplore

APA citation: Focus on a reinforcement learning algorithm that can learn from failure (2018, March 2) retrieved 28 October 2020 from <https://techxplore.com/news/2018-03-focus-algorithm-failure.html>

*This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.*