

We need robots that can improvise, but it's not easy to teach them right from wrong

November 27 2017, by Feras Dayoub, Jonathan Roberts And Michael Milford



Credit: George Becker from Pexels

Industrial robots have existed since the 1960s, when [the first Unimate robotic arm](#) was installed at a General Motors plant in the United States.

Nearly six decades on, why don't we have capable robots in our homes, beyond a few [simple domestic gadgets](#)?

One of the reasons is that the rules and conventions that govern our everyday lives are not as perfect as the rules that govern the process of, say, [assembling a car](#).

Our everyday rules do not cover all possible scenarios. This makes them filled with inconsistencies that will render useless any [robot](#) that strictly follows them.

For robots to play a more involved role in our lives, such as personal caregivers or reliable [home cooks and chefs](#), they will have to move away from following the simple rule-based operating procedures used by current robots.

Let's improvise

One potential solution is to build robots in a way that when they face a tricky situation, they can improvise.

They would do this by using their past experiences from training, combined with some context of their current situation. This approach will most likely lead to them making very complex decisions.

But there is no guarantee that the robot will choose an action that follows the social norms. These actions may appear to us humans as cheating and sometimes rude or even unfair.

So how are we going to teach a robot to act in a courteous, ethical and honourable way? Is this even a reasonable expectation?

Can robots improvise?

Consider your washing machine. One morning you are late, and in a hurry you put your clothes into the machine and pick the wrong washing program. A machine that simply follows the programmed rules will happily run and potentially ruin your favourite clothes. How annoying.

For an artificial intelligence (AI) powered washing machine to avoid this situation, it would have to go through a complex decision process to improvise. The machine might use its cameras to detect the brands of the clothes you're putting in the machine and then look them up online to find out the best washing program.

But you were in a hurry and you mixed multiple types of fabric together. This means that one washing program will be fine for some clothes but completely ruin others. A simple AI might decide to pause the machine and message you to ask for your decision.

Ah, but because you were in a hurry, you have left your phone behind. And the AI knows you need that washing done today so again, it must improvise.

Decision time

The next level of decision is for the AI to decide which of your clothes is the most sentimental to you, and save it. This is a much more complex process.

The AI now has to go through years of your recorded history to detect significant moments in your life and check if you were wearing any of the clothes you mindlessly put in the machine. But what is a significant moment and how can an AI detect it?

The AI-powered [washing machine](#) has now decided which washing program to run. It is for your heavily dirty white cotton t-shirt from your first year in college. It is using hot water and a long rinse.

This improvised decision may have ruined some of the other clothing but it resulted in you having your favourite t-shirt ready for you to wear that evening to a reunion with old friends (the AI had checked your diary).

How likely is all of this?

Robotics and AI researchers around the world are already demonstrating robots that can learn and improvise.

The robotic marimba player [Shimon](#) has played in a jazz band since 2015 to packed music venues. It listens to the music of the human band members as they improvise, and joins in.

Improvising musical robots are one thing, but cheating or deceitful robots are an altogether scarier idea.

In 2010 [researchers at Georgia Institute of Technology](#) demonstrated how one robot could deceive another robot by exploiting its vulnerabilities. The experiments involved the robots playing hide-and-seek. One robot would pretend it was in one location while secretly hiding in another.

The benefits of enabling robots and AI to cheat are studied as part of research in Human-Robot Interaction (HRI). Some of the results reported so far reveal that a [cheating robot has more human-like attributes](#), that lead to a [more natural interaction](#) between it and the people with whom it interacts.

The philosophical debate

This topic of machine morality is deep and complex with no straightforward answers but it's a recipe for a debate that emerges from the philosophy of morality.

This debate began with the science fiction literature from the mid-20th century – specifically Isaac Asimov's Three Laws of Robotics introduced in his 1942 [short story Runaround](#). As robots and AI become more capable, this debate is becoming central to our lives.

Philosophical thinkers have already been asking questions about [the challenge of moral machines](#) and [machines and moral reasoning](#) with a message to the HRI community that we should:

"[...] be careful what one wishes for ... it may come true. Therefore, we should be very careful about what abilities we program into computers, and what responsibilities we assign to them. In the end, if the [machines](#) are coming, it is humans who are constructing them."

We are still asking questions today about the morality of machines, and whether we can [teach AI and robots to understand about right and wrong](#)

.

We have a long way to go before we can build sophisticated robots with a moral compass.

The ongoing debate on the issue will not only help us build better robots but it will also make us reflect on our own ethical practices, hopefully leading us towards a better humanity.

This article was originally published on [The Conversation](#). Read the [original article](#).

Provided by The Conversation

Citation: We need robots that can improvise, but it's not easy to teach them right from wrong (2017, November 27) retrieved 18 April 2024 from <https://phys.org/news/2017-11-robots-easy-wrong.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.