

# Generic, situation-aware guidelines to help robots co-exist successfully alongside humans

June 30 2017

---



Credit: University of Hertfordshire

Artificial intelligence experts from the University of Hertfordshire, Dr Christoph Salge and Professor Daniel Polani, have designed a concept which could lead to a new set of generic, situation-aware guidelines to help robots work and co-exist successfully alongside humans.

Empowerment, which has been developed over the course of twelve

years, is discussed in the latest edition of the journal *Frontiers in Robotics and AI* today, as a potential replacement for Asimov's celebrated Three Laws of Robotics – the most famous set of guidelines to govern robotic behaviour to date.

The paper shows how Empowerment has the potential to equip a [robot](#) with guidelines or motivations that cause it to a) protect itself and keep itself functioning b) do the same for a human partner c) stick around and follow the human's lead. In the future this principle could be implemented on a range of robots that interact closely with humans in challenging environments, such as elder care robots, hospital robots, self-driving cars or exploration robots.

## **Empowering robots to change their environment**

Motivated by the term from sociology and psychology, [empowerment](#) stands for the opposite of helplessness; it is the ability to change one's environment and to be aware of that possibility. Over the past twelve years, leading University of Hertfordshire researchers have been developing ways to translate this social concept into a quantifiable and operational mathematical/technical language, endowing robots with a drive towards being empowered.

The principle of empowerment states that an agent should attempt to keep its options open, and will try to move to states in its world where it has the most options it can reliably attain. Since 2005, when it was first introduced, researchers have generalized the empowerment principle and applied it to various scenarios. The resulting behaviours are surprisingly "natural" in many cases, and typically only require the robot to know the dynamics of the world, but no specialized Artificial Intelligence behaviour coded for the particular scenario.

Empowerment has also already begun to be adopted by pioneers in

[artificial intelligence](#), such as Google DeepMind.

## **Need for ethical standards and guidelines for robots**

Dr Christoph Salge, Research Fellow at the University of Hertfordshire said, "There is currently a lot of debate on ethics and safety in robotics, including a recent call for ethical standards or guidelines for robots. In particular there is a need for robots to be guided by some form of generic, higher instruction level if they are expected to deal with increasingly novel and complex situations in the future - acting as servants, companions and co-workers.

"In the challenging scenarios of the future, we will not be able to rely on a clearly defined functionality that requires robots to be safely separated from humans, or the scenarios to be simplistic or very well defined in advance."

"Imbuing a robot with these kinds of motivation is difficult, because robots have problems understanding human language and specific behaviour rules can fail when applied to differing contexts. For example, some robots will have automatisms that stop moving whenever they encounter resistance, as a typical safety feature to avoid damaging themselves or injuring a human. But there might be a situation where a robot actually should move to provide a safer space – for instance, to move something away from the human, to get out of the human's escape route, or to actively block the human from stepping into a dangerous trajectory."

"From the outset, formalising this kind of behaviour in a generic and proactive way poses a difficult challenge. We believe that our approach can offer a solution."

Daniel Polani, Professor of Artificial Intelligence at the University of

Hertfordshire, added: "As we toyed with the idea of using empowerment in more complex situations, we realized that several of the original goals of the Three Laws of Robotics by Asimov might be addressable in the context of empowerment.

"While much of the public discourse is about how it is difficult or impossible to rein in robots' behaviour, and most certainly in keeping robots - in the most naive sense – 'ethical', in the paper we discuss possibilities to map such requirements into the formal and operational language of empowerment."

**More information:** Christoph Salge et al. Empowerment As Replacement for the Three Laws of Robotics, *Frontiers in Robotics and AI* (2017). [DOI: 10.3389/frobt.2017.00025](https://doi.org/10.3389/frobt.2017.00025)

Provided by University of Hertfordshire

Citation: Generic, situation-aware guidelines to help robots co-exist successfully alongside humans (2017, June 30) retrieved 2 May 2024 from <https://phys.org/news/2017-06-situation-aware-guidelines-robots-co-exist-successfully.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.