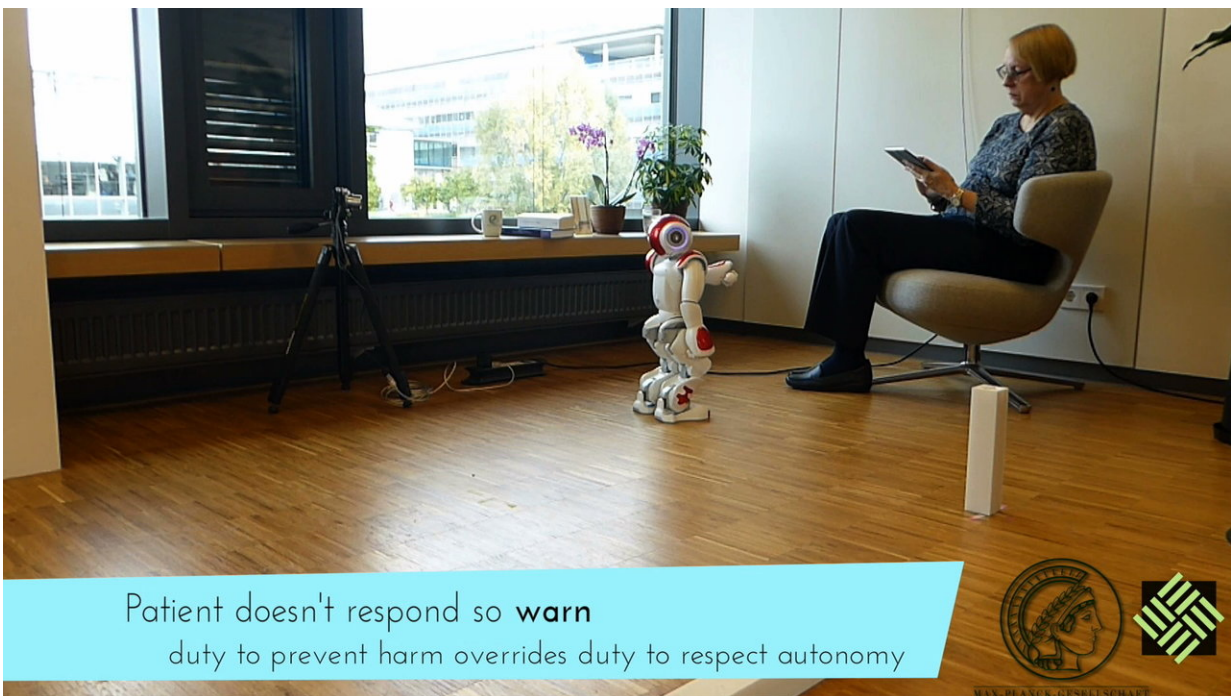


A principle-based paradigm to foster ethical behavior in autonomous machines

November 5 2018, by Ingrid Fadelli



A value-driven agent. Credit: Anderson, Anderson & Berenz.

A team of researchers at the University of Hartford, the University of Connecticut, and the Max Planck Institute for Intelligent Systems have recently proposed a case-supported, principle-based behavior paradigm toward ensuring the ethical behavior of autonomous machines. Their paper, published in [*Proceedings of the IEEE*](#), argues that ethically significant behavior of autonomous systems should be guided by explicit

ethical principles, determined via a consensus of ethicists.

"This year marks the 50th anniversary of the movie *2001: A Space Odyssey*," Michael Anderson, one of the researchers who carried out the study told TechXplore. "While reading *The Making of 2001: A Space Odyssey* at the turn of the century, it struck me that much of HAL's capability had lost its science fiction aura and was on the cusp of being realized. It also struck me that they had gotten the ethics so wrong: If HAL was around the corner, it was time to get the ethics right."

Anderson and his colleagues have been working on discovering ethical principles that can be embodied in autonomously functioning machines, as part of a project called [Machine Ethics](#). They believe that the behavior of artificial intelligence should be guided by ethical principles. These principles, determined by a consensus of ethicists, should not only help to ensure the [ethical behavior](#) of complex and dynamic systems, but also serve as a basis to justify this behavior.

The researchers developed an extensible, general case-supported and principle-based behavior paradigm called CPB. In this paradigm, an autonomous system decides its next action using a principle abstracted from cases for which ethicists have agreed upon the correct action to undertake.



Machine ethic website. Credit: Anderson, Anderson & Berenz.

"CPB uses machine learning to abstract a principle of ethical preference from ethically relevant features of particular cases of [ethical dilemmas](#), where the ethically preferable choice is clear," Anderson explained. "A system can use that principle to determine the ethically preferable action at any given moment."

The paradigm explicitly represents these principles so that they can also help to justify an autonomous system's actions. In other words, CPB can help to achieve a better understanding of a robot's behavior, offering logical explanations of why it chose to act in a particular way.

"When we were attempting to determine how the robot would know that it was in a situation where its ethics should kick in, we realized that it was always in such a state: ethics is not simply about not choosing incorrect actions, it is about choosing which of all its actions is ethically preferable in the given situation," Anderson said. "Given this, CPB is committed to determine ethically preferable actions whatever the situation, and does so whenever the situation changes."

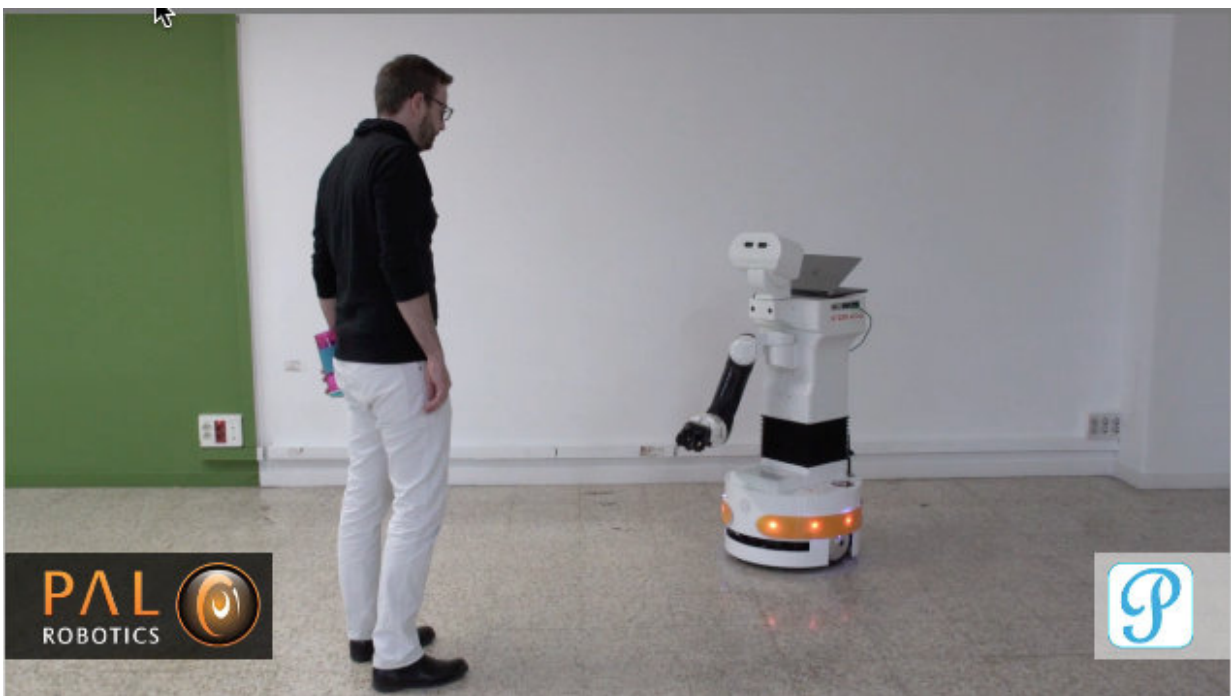


Nao medicine. Credit: Anderson, Anderson & Berenz.

The ultimate goal of the study was to determine ethically relevant features and duties in the field of eldercare, identifying principles that can guide the behavior of eldercare robots. However, the same paradigm can also be applied to robots operating in other domains. In the future, it could also provide ethical guidance for systems other than robots, such as virtual healthcare agents.

"We advocate a prima facie duty approach to ethics, where any duty might override another depending upon the situation," Anderson said. "In other words, there is no strict hierarchy as there is, say, in Asimov's laws. In the future, we believe that our feature/duty representation scheme will be useful in capturing the essence of all ethical dilemmas."

Vincent Berenz's contribution to the project, [Playful](#), recently published in [IEEE Robotics & Automation](#), allows reactive and interruptible behavior in robots. Such behavior is critical if the robot is to respond properly to constantly changing ethical concerns.



Vincent Berenz and TIAGo. Credit: Anderson, Anderson & Berenz.

"We believe that there is an expertise ethicists gain that comes with thinking long and hard about ethics," Anderson said. "We are not saying that they necessarily behave better themselves, being ethical is difficult for everyone, but that they have a better understanding of what should be done and are better able to defend these judgments than laypersons. Such careful thinking applied to simple cases can be helpful in making determinations on more difficult cases that exhibit the same ethically relevant features."

In the future, paradigms such as CPB could ensure that robots and intelligent machines make consistently ethical decisions, while also allowing humans to better understand their choices. The researchers are now working on broadening the scope of their work, in order to address a wider variety of ethical problems.

"We are planning on extending our research, adding more actions and considering more ethically relevant features to the end of developing an ethical principle that can be used to direct the [behavior](#) of all future eldercare robots," Anderson said. "We will be testing these new ideas on a PAL Robotics TIAGo [robot](#)."

More information: Michael Anderson et al. A Value-Driven Eldercare Robot: Virtual and Physical Instantiations of a Case-Supported Principle-Based Behavior Paradigm, *Proceedings of the IEEE* (2018). [DOI: 10.1109/JPROC.2018.2840045](https://doi.org/10.1109/JPROC.2018.2840045)

Vincent Berenz et al. The Playful Software Platform: Reactive

Programming for Orchestrating Robotic Behavior, *IEEE Robotics & Automation Magazine* (2018). [DOI: 10.1109/MRA.2018.2803168](https://doi.org/10.1109/MRA.2018.2803168)

www.machineethics.com/

playful.is.tuebingen.mpg.de/

© 2018 Science X Network

Citation: A principle-based paradigm to foster ethical behavior in autonomous machines (2018, November 5) retrieved 10 April 2024 from <https://techxplore.com/news/2018-10-principle-based-paradigm-foster-ethical-behavior.html>

<p>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.</p>
--