



Sie sind hier: Startseite > Presse & Öffentlichkeit > Presseportal > **Pressemeldung**

Pressemeldung

Nr. 145 / 2017

07. Juli 2017

Machines will soon be able to imitate human moral behavior - Research results from the Institute of Cognitive Science at Osnabrueck University

Self-driving cars are the first generation of robots that share the same everyday habitat with us. It is therefore necessary to develop rules and expectations for autonomous systems that define how these systems should behave in critical situations. The Institute of Cognitive Science at Osnabrueck University has now published a study in *Frontiers in Behavioral Neuroscience* (<http://journal.frontiersin.org/article/10.3389/fnbeh.2017.00122/full>) that highlights the feasibility of incorporating human moral decisions into machines, and suggests that autonomous vehicles will soon be able to deal with moral dilemmas in road traffic.

On the political side, the debate about the feasibility of modeling human moral decisions is being led by an initiative of the German Federal Department of Transport and Digital Infrastructure (Bundesministerium für Transport und Digitale Infrastruktur; BMVI), which formulated 20 ethical principles for self-driving cars. The study from Osnabrueck now brings empirical data to this debate.

„To be able to define rules and guidelines, a two-step process is needed. First, the moral decisions of humans in critical situations have to be analyzed and understood. In the second step, this behavior needs to be described statistically, in order to derive rules which can then be used by machines“, explains Prof. Dr. Gordon Pipa, one of the leading scientists in the study.

To put both steps into practice, the authors made use of virtual reality to observe the behavior of participants in simulated traffic situations. To this end, the participants drove down a road in a typical suburban setting on a foggy day. In the course of the experiment, the participants were confronted with unavoidable dilemma situations, in which humans, animals and/ or inanimate objects were blocking their way. Ethical considerations had to be made since the participants could always only spare one of two obstacles, but had to sacrifice the other. The observed decisions were later statistically analyzed and translated into rules. The results suggest that in such moral dilemma situations, our moral behavior can be explained by rather simple models based on values of life which are assigned to each human, animal and object.

Leon Suetfeld, first author of the study, puts it like this: „Human moral behavior can be explained and predicted with impressive precision by comparing the values of life that are associated with each human, animal and inanimate object. This shows that human moral decisions can in principle be explained by rules, and these rules can be adopted by machines.“

These new insights from Osnabrueck contradict the 8th principle of the BMVI report, which makes the assumption that moral decisions cannot be modelled.

How can this fundamental difference be explained? Algorithms can either be based on categorical rules or on statistical models that put multiple factors in relation. Laws, as an example, are based on categorical rules. In contrast, human behavior and modern artificially intelligent (AI) systems incorporate statistical probabilities into their assessments. This incorporation of statistical probabilities allows both humans and AI systems to adapt to and evaluate new situations that they have never encountered before. In their work, Suetfeld and colleagues used such a methodology to describe the data. „The rules don't have to be formulated in an abstract manner by a human sitting at their desk, but can be derived and learnt from human behavior directly. This raises the question of whether we should make use of these learnt and conceptualized rules in machines as well.“, says Suetfeld.

„Now that we have a way of implementing moral decision making for machines, two moral dilemmas remain“ says Prof. Dr. Peter Koenig, another co-author of this publication — „First of all, we need to decide on the influence of moral values on the guidelines for machine behavior. Secondly, we need to discuss whether or not machines should (always) behave like humans.“ „Using Virtual Reality to Assess Ethical Decisions in Road Traffic Scenarios: Applicability of Value-of-Life-Based Models and Influences of Time Pressure“ sind erschienen in „*Frontiers in Behavioral Neuroscience*“ (<http://journal.frontiersin.org/article/10.3389/fnbeh.2017.00122/full>)

More information for the media:

Prof. Dr. Gordon Pipa, Osnabrueck University,
Institute of Cognitive Science,
Wachsbleiche 27, 49090 Osnabrueck,
Tel: +49 541 969 2277,
E-Mail: gordon.pipa@uni-osnabrueck.de

Leon Suetfeld, Osnabrueck University,
Institute of Cognitive Science,
Wachsbleiche 27, 49090 Osnabrueck,
Tel: +49 541 969 7091,
E-Mail: lsuetfel@uni-osnabrueck.de



© Universität Osnabrück

Participants were seated at the wheel of a virtual car driving towards a set of obstacles in a suburban setting. A collision was unavoidable, and participants were only given the choice, which of the two obstacles they would spare, and which one they would sacrifice. Copyright: Osnabrueck University