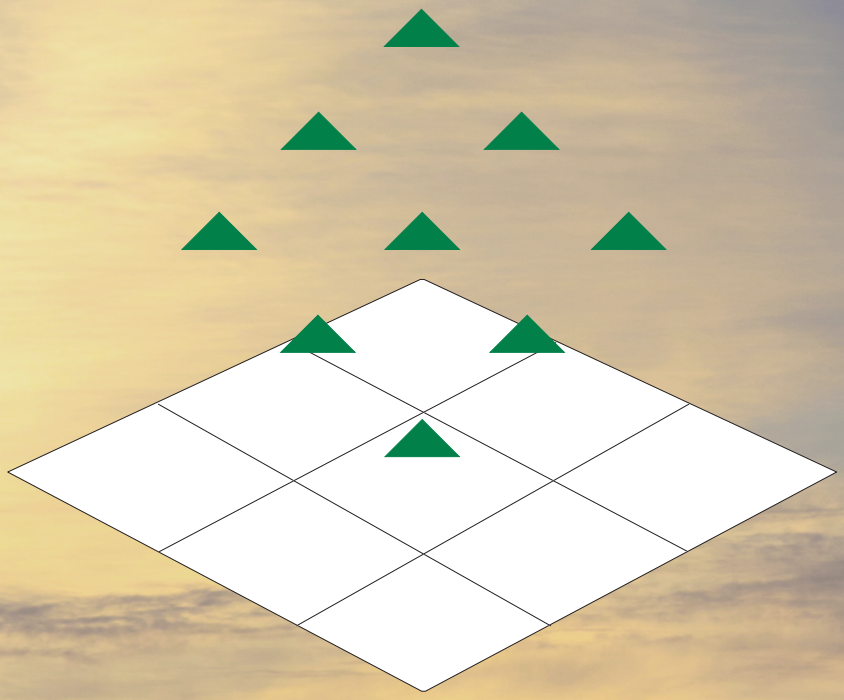

The 2016 AAAI Spring Symposium Series

Technical Reports

The 2016 AAAI Spring Symposium Series

Technical Reports
SS-16-01 – SS-16-07

(Collected in One Volume)

AAAI Press
Palo Alto, California

The Association for the Advancement of Artificial Intelligence
2275 East Bayshore Road, Suite 160
Palo Alto, CA 94303 USA

ISBN 978-1-57735-754-4

Manufactured in the United States of America

Introduction

The AAAI Spring Symposium Series is an annual set of meetings run in parallel at a common site. It is designed to bring colleagues together in an intimate forum while at the same time providing a significant gathering point for the AI community. The two and one half day format of the series allows participants to devote considerably more time to feedback and discussion than typical one-day workshops. It is an ideal venue for bringing together new communities in emerging fields.

The symposia are intended to encourage presentation of speculative work and work in progress, as well as completed work. Ample time is scheduled for discussion. Novel programming, including the use of target problems, open-format panels, working groups, or breakout sessions, is encouraged. AAAI Technical Reports are prepared, and distributed to the participants. Most participants of the symposia were selected on the basis of statements of interest or abstracts submitted to the symposia chairs; some open registration is allowed. All symposia are limited in size, and participants are expected to attend a single symposium.

The Association for the Advancement of Artificial Intelligence, in cooperation with Stanford University's Department of Computer Science, is pleased to present the 2016 Spring Symposium Series, held Monday through Wednesday, March 21-23, 2016 on the campus of Stanford University. The seven symposia are:

SS-16-01: AI and the Mitigation of Human Error: Anomalies, Team Metrics and Thermodynamics

SS-16-02: Challenges and Opportunities in Multiagent Learning for the Real World

SS-16-03: Enabling Computing Research in Socially Intelligent Human-Robot Interaction: A Community-Driven Modular Research Platform

SS-16-04: Ethical and Moral Considerations in Non-Human Agents

SS-16-05: Intelligent Systems for Supporting Distributed Human Teamwork

SS-16-06: Observational Studies through Social Media and Other Human-Generated Content

SS-16-07: Well-Being Computing: AI Meets Health and Happiness Science

Organizers and Committees

SS-16-01: AI and the Mitigation of Human Error: Anomalies, Team Metrics and Thermodynamics

Cochairs

Ranjeev Mittu (Naval Research Laboratory)
Gavin Taylor (US Naval Academy)
Don Sofge (Naval Research Laboratory)
W.F. Lawless (Paine College)

Organizing Committee

Ranjeev Mittu (Naval Research Laboratory)
Gavin Taylor (US Naval Academy)
Don Sofge, Naval Research Laboratory; Navy Center for Applied Research in Artificial Intelligence; don.sofge@nrl.navy.mil
William F. Lawless, Paine College, Departments of Math and Psychology; wlawless@paine.edu

Invited Keynote Speakers

Julie Adams, Vanderbilt University, Associate Professor of Computer Science and Computer Engineering, Electrical Engineering and Computer Science Department, julie.a.adams@vanderbilt.edu

James Llinas, SUNY at Buffalo, llinas@buffalo.edu

Stephen Russell, Chief, Battlefield Information Processing Branch, US Army Research Lab, MD; stephen.m.russell8.civ@mail.mil

Martin Voshell, Charles River Analytics, mvoshell@cra.com

SS-16-02: Challenges and Opportunities in Multiagent Learning for the Real World

Organizing Committee

Christopher Amato (University of New Hampshire, USA)
Frans Oliehoek (University of Amsterdam, NL and University of Liverpool, UK) Miao Liu (MIT, USA)
Karl Tuyls (University of Liverpool, UK)
Peter Stone (University of Texas at Austin, USA)
Jonathan How (MIT, USA)

Program Committee

Daniel Claes (University of Liverpool, UK)
Sam Devlin (University of York, UK)
Enrique Munoz De Cote (Institute of Astrophysics, Optics and Electronics, MX)
Matthijs Spaan (Delft University of Technology, NL)

SS-16-03: Enabling Computing Research in Socially Intelligent Human-Robot Interaction: A Community-Driven Modular Research Platform

Organizing Committee

Maja Mataric (University of Southern California)
Mark Yim (University of Pennsylvania)
Ross Mead (University of Southern California)

Program Committee

Braden McDorman (University of Southern California)
Simon Kim (University of Pennsylvania)
Andrew Specian (University of Pennsylvania)
Yue Chen (University of Pennsylvania)
Junghyo Lee (University of Pennsylvania)

SS-16-04: Ethical and Moral Considerations in Non-Human Agents

Cochairs

Bipin Indurkhya (Jagiellonian University, Kracow, Poland)
Georgi Stojanov (The American University of Paris, France)

Organizing Committee

Joanna Bryson (University of Bath, UK; Princeton University, USA)
Tom Lenaerts (Universite Libre de Bruxelles, Belgium)
Tony Veale (University College Dublin, Ireland)

SS-16-05: Intelligent Systems for Supporting Distributed Human Teamwork

Cochairs

Ofra Amir (Harvard University)
Krzysztof Z. Gajos (Harvard University)
Barbara J. Grosz (Harvard University)
Gary Olson (University of California, Irvine)
Judith Olson (University of California, Irvine)

Program Committee

Ya'akov (Kobi) Gal (Ben-Gurion University)
Jonathan Grudin (Microsoft Research)
Robert Kraut (Carnegie Mellon University)
Peter Stone (University of Texas, Austin)

SS-16-06: Observational Studies through Social Media and Other Human-Generated Content

Organizing Committee

Elad Yom-Tov (Microsoft Research)
Munmun De Choudhury (Georgia Tech)
Emre Kıcıman (Microsoft Research)

Program Committee

Omar Alonso (Bing)
Ceren Budak (University of Michigan)
Aron Culotta (Illinois Institute of Technology)
Carlos Castillo (Data Mining, Eurecat)
Ingemar Cox (University College London)
Brian Keegan (Harvard Business School)
Luis Luque (Norut)
Andrey Rzhetsky (University of Chicago)
Johan Ugander (Stanford University)
Hao Wang (IBM)
Ingmar Weber (Qatar Computing Research Institute)

SS-16-07: Well-Being Computing: AI Meets Health and Happiness Science

Organizing Committee

Takashi Kido, Cochair (RIKEN GENESIS, Japan)
Keiki Takadama, Cochair (The University of Electro-Communications, Japan)

Quantifying Our Health and Cognitive Performance Committee

Melanie Swan (DIYgenomics, USA)
Katarzyna Wac (Stanford University, USA and University of Geneva, Switzerland)
Ikuko Eguchi Yairi (Sophia University, Japan)

Discovery Informatics and Health/Cognitive Modeling Committee

Chirag Patel (Stanford University, USA)
Rui Chen (Stanford University, USA)
Ryota Kanai (University of Sussex, UK)
Yoni Donner (Stanford, USA)
Yutaka Matsuo (University of Tokyo, Japan)

Designing Health and Cognitive Enhancement Committee

Eiji Aramaki (University of Tokyo, Japan)
Pamela Day (Stanford, USA)
Tomohiro Hoshi (Stanford, USA)

Application, Platform, Field Study Committee

Miho Otake (Chiba University, Japan)
Yotam Hineberg (Stanford, USA)
Yukiko Shiki (Kansai University, Japan)

Advisory Committee

Atul J. Butte (Stanford University, USA)
Seiji Nishino (Stanford University, USA)
Katsunori Shimohara (Doshisha University, Japan)

Contents

Annotated Decision Trees for Simple Moral Machines

Oliver Bendel

School of Business FHNW, Bahnhofstrasse 6, CH-5210 Windisch
oliver.bendel@fhnw.ch

Abstract

Autonomization often follows after the automization on which it is based. More and more machines have to make decisions with moral implications. Machine ethics, which can be seen as an equivalent of human ethics, analyses the chances and limits of moral machines. So far, decision trees have not been commonly used for modelling moral machines. This article proposes an approach for creating annotated decision trees, and specifies their central components. The focus is on simple moral machines. The chances of such models are illustrated with the example of a self-driving car that is friendly to humans and animals. Finally the advantages and disadvantages are discussed and conclusions are drawn.

Introduction

More and more semi-autonomous and autonomous machines have to make decisions with moral implications. Machine ethics analyses the chances and limits of moral machines (Anderson and Anderson 2011; Wallach and Allen 2009; Bendel 2014e; Bendel 2012). It is a design discipline located between artificial intelligence (AI), robotics, computer science and philosophy. The phase of brainstorming for ideas is long since over. Today this discipline works on the concept of moral machines. Prototypes have already been presented (Aegerter 2014). Slowly but steadily the design discipline is living up to its classification and its own claim.

Decision trees represent rules of decision making. They are widely used in economics, computer science, and artificial intelligence. They are made up with root nodes and internal nodes linked to one another and to decision-making options. The forms of representation are plenty. They often begin from a defined starting point and then a question is raised with "Yes" or "No" as possible answers. These answers lead to new questions until several options are reached at the end. Branch structures with additional information deriving and reasoning the questions can be considered annotated decision trees (Bendel 2015a). The nodes, or the links between the nodes, are described below in more detail.

So far, decision trees have been used rarely only for modelling moral machines (Bendel 2015a; Bendel 2015b). Modelling efforts on a meta level have been documented in (Anderson and Anderson 2011), for instance the "MoralDM Architecture" by (Deghani et al. 2011). (Azad-Manjiri 2014) drafts an "architecture of moral agents", including a "decision tree algorithm to abstract relationships between ethical principles and morality of actions" (Azad-Manjiri 2014, 52). (Bostrom and Yudkowsky 2014) reason that "a machine learner based on decision trees or Bayesian networks is much more transparent to programmer inspection".

In this article, following the explanation of the term of the simple moral machine, a concept for creating annotated decision trees in the context of machine morality is proposed, and their central components are specified. A concrete modelling is presented and illustrated with the example of a self-driving car that is friendly to humans and animals. In this set-up it can be considered a simple moral machine. The modelling is explained in detail. Finally the advantages and disadvantages of such decision trees are discussed.

Simple Moral Machines

Simple moral machines mean (semi-)autonomous systems that follow a few simple rules in the standard situations they have been developed for, or make correct decisions by means of observations and analysis of memorized cases, and in consequence act morally (well) (Bendel 2013a). Complex moral machines on the other hand have to master a large number of morally charged situations. Examples are self-driving cars involved in accidents with humans in conventional road traffic or martial drones programmed to eliminate target persons. In order to show the problematic of moral machines, frequent reference is made to exactly such complex machines and the associated conflicts. The fact that humans have a tendency of failing in this kind of situation, which on principle can hardly be mastered at all,

not even with high moral competency, high rationality and high empathy, tends to be forgotten.

The proposed terms are not for classification but for orientation. Therefore it is not necessary to draw a clear demarcation line. A few examples should suffice to clarify the idea of simple moral machines (Bendel 2013a; Bendel 2015a):

- Chatbots or chatterbots on websites inform people about products and services, they provide entertainment and customer allegiance. Most of them would respond inadequately to an announcement of suicide plans. SGT STAR of the U.S. Army is a "good" bot in this respect as it mentions the phone number of the National Suicide Prevention Lifeline. A "better" bot would give out the hotline for the country of the user, or would connect the user to a contact person. This kind of behavior can be realized by extending the knowledge base and by evaluating the IP address. The GOODBOT of School of Business FHNW meets these requirements (Aegerter 2014). Before handing over to a human being it would escalate on several levels.
- Servicebots such as carebots, therapybots, household- and gardenbots are available in many different designs. Robot mowers and robot vacuum cleaners are widely in use. A standard robot vacuum cleaner ingests all that is in front of or under it, not only things and dirt but small and smallest beings as well. Many people believe animals should not be hurt or eliminated. Robots could be furnished with image recognition and motion sensors, and could be taught to spare the lives of beings (Bendel 2014a). Robot mowers too could be improved in this manner however lawns are relatively complex environments and meadows even more so.
- Private drones such as unmanned aerial vehicles (UAV) are used by companies, media, scientists and the police forces and are growing more and more popular. They can transport goods, and when furnished and upgraded adequately, they can photograph or film objects and people. Most people object to being photographed secretly and having their privacy invaded. Cities like Zurich have already responded with rigorous restrictions (Häuptli 2014). Drones can be supplemented with image and pattern recognition to make them refrain from taking photos (Bendel 2015c). This leaves them fully functional but limited in the best interest of persons who are affected.
- Self-driving cars, also known as robotic or robot cars, are already underway in several US-American and European urban and rural regions. They unburden or replace the driver, they can prevent accidents and save the lives of passengers and other road users. Night vision devices and image recognition systems can distinguish between humans and animals and set priorities if this function is enabled. This kind of advanced driver assistance systems (ADAS) today already allows for moral machines in the widest meaning of the term (Bendel 2014b).
- Wind power stations are often built on hills or mountains or in the open sea, with giant rotors on high pylons. Collisions with birds and bats occur frequently. When combined with ultrasonic systems and image and pattern recognition, the turbines would be able to shut down when necessary. They could alert each other of the movements of individuals or swarms and make the relevant decisions. Sensors would make it possible to set-up an early warning system with sounds and light stimuli in the wider environment. A few animal-friendly prototypes are already in operation (Federle 2014).
- 3D printers have been launched on mass markets some time ago. They are capable of "printing out" all kinds of objects. Typical materials used for 3D printing are plastics, metals and plaster in the form of powder, granules, solid pieces or liquids. The materials are glued or melted and hardened or dried. Small and large forms can be created, even entire rooms or houses (Emmerth 2013). Several firearms were modelled successfully out of plastics and withheld several shots. Even functional metal firearms have been printed in the meantime. 3D printers that analyze files and find out information on form and function of the targeted object could prevent the production of pistols or bomb components (Bendel 2013a).

This list could be continued at random. It has been shown that very different types of machines are on issue, softwarebots and hardwarebots, large processor-controlled systems with movable parts, or small electronic devises. Their decisions, good or bad, as well as their non-decisions affect humans and animals.

Annotated Decision Trees for Moral Machines

In the following a process for developing annotated decision trees is proposed. This process has been proven to result in consistent and feasible models. A simple version of an annotated decision tree was created in (Bendel 2015b). The present article describes the process of more complex versions as found in (Bendel 2015a). Surely refined versions can be a goal for the future.

The first step in the development of decision trees for simple moral machines is to define the system. It shall be reviewed whether the system can be considered simple wholly or partly, and whether it can be turned into a moral machine, capable of making intentional, purposive decisions of moral implications. Sometimes moral machines will have to be designed from scratch. Moral questions frequently turn up in vehicles in urban traffic or in infor-

mation systems as sociotechnical systems. More or less self-sufficient systems as used for instance for scientific missions in volcano craters or on Mars rarely face conflicts. Challenges are bound to occur when systems diffuse in our society.

The next step will be to define the relevant function or (sub-)task of the machine (Bendel 2015a). With a strong focus, even a complex machine can become a simple moral machine. For instance one can build a car with advanced driver assistance systems or an autonomous car that will generally emergency brake for people, but will consider the type, age and health of animals and align the braking accordingly (Bendel 2015a). This car will not have to weigh thousands of possible alternatives or determine the value of human individuals or groups. Its function will be well-defined and the complexity of its reality reduced. This is exactly what the following chapter deals with.

In another step the targets of this activity will be defined in more detail. Is saving lives the goal, or avoiding injuries? Are humans or animals in the focus of attention? The targets have moral connotations. Structures and annotations will be derived from them later. Before that, the endpoints should be noted. They are different alternatives for decision making, they correspond to the options of the machine, and they depend on the activity in the focus. They also help achieve the targets. For a photo drone programmed to take pictures of flora and fauna but not of humans, the final decision could be: "don't photograph", "photograph from great heights" or "photograph from different heights". Another consideration is that photographing might disturb and stress animals, therefore the moral drone would not only refrain from photographing humans, but would adjust to the situation of the animals as well (Bendel 2015a).

Then the root node is determined with the related questions. For an animal-friendly robot vacuum cleaner, the question might be: "Is there anything on my track?" (Bendel 2015a) In this example the starting point for the first branch is an exceptional situation (if there is no exception, the work will be completed as per routine). Binary decision trees, which have preference in this article, always have one branch titled "Yes" and one titled "No" that leads to another (internal) node or directly to an endpoint and hence to the decision. If, in the case of the robot vacuum cleaner, the answer is "Yes", then the next question could be "Is it an animal?". If the answer again is "Yes", a distinction could be made between size and type. The cleaner would not have to shut down for large animals (or for humans), because they could not be sucked in, but it might have to shut down for smaller animals. Moral questions (and answers) seem to impose in such matters. This leads to the next step.

The individual nodes can be annotated in order to analyze and reason the questions with the help of comprehensive and discussable assumptions to match and link them to the further proceedings as best possible. One can also start from the assumptions, and place the nodes later on. It is a dialectical process, the assumption creates the node and the node is annotated with the assumption until a satisfactory overall image is achieved. It is proposed to segregate the different kinds of assumptions from each other, and to make different assumptions, for instance from the perspective of morality, economic efficiency, and safety of operation (Bendel 2015a). One node can have several assumptions – from one perspective or from different points of view. The annotations could be numbered consecutively, best within a category (so in the end there would be for example moral assumptions 1 to 5, distributed over several nodes).

In the example of the drone, the question whether it is a human could be negated. The next questions could be: is it an animal? (node 2), a plant? (node 3) or a thing? (node 4). At node 3 the assumption could be that plants are hardly or not at all influenced by drones, so drones could approach them freely. More recent research however has found evidence of communication and reception abilities in the flora. At node 2 after a confirmation the next question could be: is it a bird? Birds in flight, one could assume, should not be injured by drones. These would have to avoid individual birds as well as flocks of birds, and refrain from photographing close up. Obviously, such annotations are useful in order to find the right (or at least reasoned) rules for decision making.

In the last step, more nodes are found and branches created. 10 to 20 nodes can be coped with in modelling, and can be represented fully on screen. The number should be limited also for reasons of technical reality. In our example, the photo drone has to recognize whether there is a human, an animal, a plant, or a thing, and it has to determine the animal species etc. Another consideration is that the observations and decisions have to be followed by technically feasible actions. The total structure is tested for consistency, reviewed for loops, dead ends or errors, and the final plausibility is verified in the end. In general every node should include a question which is unambiguous and verifiable with the available technical or other resources. This issue is not the object of this contribution.

Components of Annotated Decision Trees

The central components of annotated decision trees can be derived from the proposed process. They also depend on the applied modelling tool and language. An attempt is being made to keep the components as general as possible while integrating commonly used forms. The application to the moral machine and the extension of the decision tree into morality are of particular importance.

- At the edge of the decision tree, the class or type of machine is specified as a kind of header, and the target of the moral machine is described in a few words.
- The starter symbol – a rounded or standard rectangle – briefly and precisely denominates the task. Multiple tasks would require multiple decision trees.
- Behind the task and linked with an arrow element follows the root node with the first question. A rhombus is proposed as the symbol for the root node.
- Two branches (also in the form of arrows) branch off the root node towards two internal nodes or to one (but no more than one) decision. The branches are marked with "Yes" or "No" accordingly.
- The internal nodes are furnished with more questions for review. They go out to more branches, to more nodes, or to the final decisions at the end.
- Root nodes and internal nodes are annotated as far as possible and necessary. This should be done in the comment mode to make sure the annotation is unambiguously linked to the question(s).
- Annotations are made from the perspective of morality (obligatory), economic efficiency, operational safety etc. They can be numbered and ranked by priority.
- The endpoints – symbolized by rectangles – give decisions that can be implemented by the machine. They can be demarcated clearly and be described unambiguously, and they can lead to a continuum of alternatives.
- A caption mentions the abbreviations of the perspectives and explains them in more detail. It can also explain special characters such as the negation symbol – if used – and priorities.

The assumptions can be reasoned in more detail in an additional document that can also refer to the models of normative ethics, and classify the annotations in the cultural and social context.

Annotated Decision Trees for Robot Cars

The following focuses on a robot car or a car with advanced driver assistance systems, which under certain circumstances can be considered a simple moral machine. The general underlying assumption is that decisions towards human beings, especially if concerning their health and lives, are highly complex. In particular the choice between the well-being of different people will almost always present a moral dilemma. The issue could be whether the car, when the brakes fail, shall kill the man, the woman, the old person or the young person, a single person, a group of people and so on (Holzer 2015). When concentrating on animals the situations seem to be easier to oversee and the decisions simpler. Giving priority to certain species will hardly rebuke people, if lives can be saved or a species can be protected, although – and this will be discussed in more detail further below – animals are perceived and valued very differently. This could be a general stimulation for moral machines (Bendel 2015a).

Cars, busses and other vehicles use more and more advanced driver assistance systems. Some of them assist the driver, inform and support him, others convert the depending machine to a semi-autonomous one which temporarily and partly functions independently of the driver (Bendel 2014d). Traffic sign recognition, braking assistants, emergency braking assistance, lane changing assistance or construction zone assistance, autonomous cruise control systems and parking assistance are examples. ADAS are usually permanently installed in the car. Even fully autonomous systems, such as self-driving cars or trucks, no longer are science-fiction (Bendel 2015d). Prototypes are known, with the Google car as one example, as well as scientific or commercial projects. They can be seen in European cities (Kolhagen 2013; Stoller 2013). The automobile manufacturer Daimler urges its autonomous trucks, which have been cruising on US roads for some time, onto German roads with high speed (Bradl 2015). Autonomous systems are independent of humans for longer periods of time, in their decisions as well as in their motions and activities. Of course the rules are predefined for them to begin with. However such systems are capable of learning, also through their observations, prioritize and adjust rules accordingly.

It is possible to develop ADAS capable of making decisions relating to animals (Bendel 2014d). Animal-related actions are absolutely relevant, this is indicated by many pertinent road signs in many countries, where they warn of toad migration, hedgehog populations or deer crossing. Emergency braking systems should be able to respond appropriately and without human assistance to imminent dangers, always under consideration of tailgating cars and other factors. Modern image recognition and night vision systems can differentiate between animals and humans even in the dark. In interaction with emergency braking systems they are capable of making good and right judgements. In general, autonomous cars either have to respond adequately to avoid accidents or escalate to humans (Goodall 2014).

Decision trees for autonomous cars and advanced driver assistance systems can look back unto certain traditions (Bendel 2015a). (Kopf 1994) for instance presents a situative analysis with decision trees for assisting drivers on highways. (Lorenz 2014) also addresses this instrument in the context of concepts for ADAS and specifies precisely: "Entscheidungsbäume veranschaulichen hierarchisch aufeinanderfolgende Entscheidungen zur Klassifikation bestimmter Objekte oder Zustände (Decision trees visualize decisions in consecutive hierarchies for classification of certain objects or conditions)" (Lorenz 2014, 59) A deci-

sion tree aligned to morality for practical implementation was roughly sketched in (Bendel 2015b) and was much extended in (Bendel 2015a). Cars with advanced driver assistance systems or autonomous cars are versatile systems just like drones. One of their main tasks is driving. Many sub-tasks have to be mastered for this purpose (Pellkofer 2003), such as speed regulation and keeping or changing lanes as required. The following concentrates on braking under special consideration of animals. These limitations and settings of priorities allow considering the moral machine as a simple one.

The modelling (see Fig. 1) assumes the activity is driving (Bendel 2015a). The lane is checked for objects less than 40 meters away from the car (this value should be replaced by a formula as this distance might be too short for high speeds and to long for low speeds). If an object is detected on the road, and this is a human being, the system initiates emergency or danger braking.

If an animal is in danger, the system will proceed depending on the species as in the example of the drone. Collisions with bigger animals shall be avoided, and rare species shall have special consideration. Insects and mollusks are exempt. Braking for them would be uneconomical and mobility, the purpose of driving, would be very limited. If the detected object is not a being, then other factors will have to be considered. Bigger objects would require braking in order to avoid damage to the vehicle and risks for the lives of the passengers. Of course, reality can come in many other different forms: a tiny object like a nail could cause considerable damage, avoiding it might be sensible. This issue could be modelled. Leaving certain impacts unconsidered might be reasonable to keep the complexity manageable.

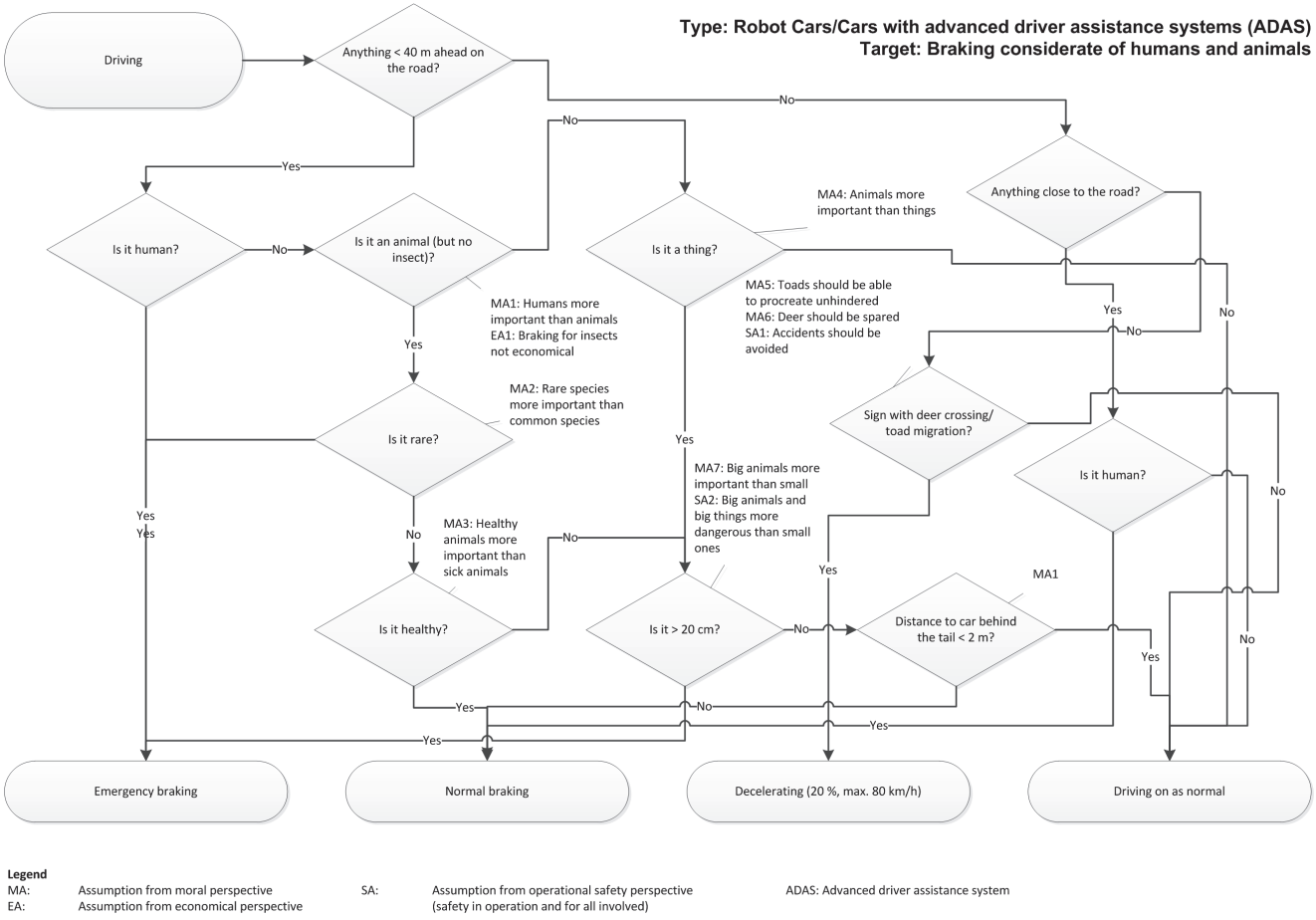

Fig. 1: Decision tree for robot cars (cf. Bendel 2015a)

Chances and Risks of Decision Trees

Chances and risks of annotated decision trees for moral machines are discussed in the following. One important benefit of decision trees is that the developer can design the system using transparent dialectic methods. Modelling is similar to loud brainstorming, weighing up, adding and excluding. It can be reviewed and edited. It visualizes different considerations with their backgrounds and in their context and allows defining the alternatives of decision making. After all, it's a very simple method.

Decision trees can be converted and executed by computer programs. They also contain hints to the required sensors and actors, they refer not only to the software of moral machines but also to the hardware. This is essential considering that moral machines usually act and move in open worlds (which of course have to be closed partly). In the mentioned examples optical analysis (image and pattern recognition) and measuring methods using ultrasound can be applied. Technological standards are not defined and this is another benefit.

In this context it is very important that the annotated decision trees provide an instrument for the development of moral machines. Not only through the annotations, helpful as they are, but also through selection of suitable machines as well as definition and selection of activities. The modelling illustrates the goal and the underlying assumptions while offering the highest possible freedom, not only in technological aspects. It does not define mandatory standards on the moral level. By requiring mandatory annotations it only requests that something be explained, and if possible also reasoned, in writing.

There are no obstacles for further extension of the system. The assumptions could be linked to detailed reasoning filed in an additional document. They can be sequenced by priority. Moral assumptions might be considered more important than economical assumptions (or vice versa). This issue was not intensified in this article, but it was made clear that the admission of certain questions (for instance for insects) would lead to highly unsatisfactory options (permanent braking activities, or even standstill in the warmer seasons). In this aspect priorities were set ex ante.

Drawbacks have also been found. The dialectic method cannot exclude creation of redundant nodes and branches or over-modelling. This is a problem not only in terms of programming elegance, but also for technological realization. It is not granted that the decision options in the end are the right ones, just because they are possible and because a continuum (e.g., of the types of braking) can be seen. The modelling itself offers not enough room for more detailed reasoning, at least in the standard formats and on standard screens. Having to refer to an additional document makes the model more difficult to survey. Zoom-in and zoom-out functions are available but add little clarity.

One could also complain that philosophy doesn't come to its rights here, and that modelling fails to consider models of normative ethics as well as duty ethics or consequence ethics or other options of funding and setting. There is the risk of applying too much of a hands-on mentality, with a layman understanding of ethics and morality. However the feasibility is indeed a benefit of annotated decision trees. They remove fears with respect to moral machines, and draw them closer to the range of what is possible and practicable. Information and classification can be presented in an additional document.

A further criticism is the purely rational reasoning of the decisions related to measurable dimensions and observable facts. The machine has no problem with determining the size of animals, their species or if they are rare. It can analyze the object according to the latest state of the art. Morality however is more than just measuring and valuing, and working off a list of rules. In the best case it includes instincts and empathy. The machine might have to be taught a different perspective towards animals. Many people love their pets, they would rather accept driving over a giant tarantula than over a newborn kitten. Surely the development of machines should not ignore the feelings of humans or their sets of values. At the same time there is a chance to raise the discourse on a more rational level. Not only pretty and trustful animals are worthy of protection, but also animals that are ugly, rare, or necessary for the ecobalance.

The final criticism is the technological basis and the overall architecture. What if some measurement results are uncertain in practice? What if environmental conditions are difficult or if people mislead the machines by using optical methods? What if decisions of the module responsible for the robot operation are in conflict with the moral module?

Summary and Outlook

Decision trees are suitable for the representation of decision making rules with moral implications. In this article they were applied to a simple moral machine. Completeness was not claimed. The intention was to illustrate and clarify the principle. The moral assumptions were visualized in annotations. Their being cogent or shared by a wide majority was not required. Again, emphasis was on understanding the principle. It was shown that further to moral reasoning, other reasons related to profitability and operations are possible and sensible.

Future research can tackle the further development of decision trees. These must be, for example, integrated in an overall architecture to ensure the optimal functioning and the conflict-free processing. The form of the annotations

can be standardized in meta documents. It must be underlined that the sensor systems as a basis for decisions have to be improved. Furthermore, one can try to use alternatives like the mentioned Bayesian networks, therefore probabilistic graphical models.

Generally, different routes can lead to the goal. Maybe only the finished machines and their behaviors will qualify the best methods. It is essential not to lose time in machine ethics as happened in AI. The discussion of moral machines is going on full speed, and it would not benefit the discipline to keep talking for 50 years instead of presenting results. If it wants to be acknowledged permanently as a design discipline it has to show successful outcomes that are backed up by philosophy and technology as well as compatible to society.

References

Aegerter, J. 2014. FHNW forscht an "moralisch gutem" Chatbot. *Netzwoche*, 4/2014, 18.

Anderson, M., and Anderson, S. L. eds. 2011. *Machine Ethics*. Cambridge: Cambridge University Press.

Azad-Manjiri, M. 2014. A New Architecture for Making Moral Agents Based on C4.5 Decision Tree Algorithm. *International Journal of Information Technology and Computer Science* (IJITCS), Vol. 6, No. 5, April 2014, 50–57.

Bendel, O. 2015a. Einfache moralische Maschinen: Vom Design zur Konzeption. *Proceedings der AKWI 2015.* Luzern.

Bendel, O. 2015b. Die Maschinenstürmer des Informationszeitalters. *ICTkommunikation*, March 5, 2015. http://ictk.ch/content/die-maschinenst%C3%BCrmer-des-informationszeitalters.

Bendel, O. 2015c. Private Drohnen aus ethischer Sicht: Chancen und Risiken für Benutzer und Betroffene. *Informatik-Spektrum*, February 14, 2015 ("Online-First" Article on SpringerLink).

Bendel, O. 2015d. Selbstständig fahrende Autos. *Gabler Wirtschaftslexikon*. Wiesbaden: Springer Gabler. http://wirtschaftslexikon.gabler.de/Definition/selbststaendig-fahrende-autos.html.

Bendel, O. 2014a. Die Roboter sind unter uns. *Netzwoche*, 22/2014, 28.

Bendel, O. 2014b. Advanced Driver Assistance Systems and Animals. *Künstliche Intelligenz*, Volume 28, Issue 4 (2014), 263–269.

Bendel, O. 2014c. Für wen bremst das Roboterauto? *Computerworld.ch*, May 16, 2014. http://www.computerworld.ch/marktanalysen/studien-analysen/artikel/.

Bendel, O. 2014d. Fahrerassistenzsysteme aus ethischer Sicht. *Zeitschrift für Verkehrssicherheit*, 2/2014, 108–110.

Bendel, O. 2014e. Wirtschaftliche und technische Implikationen der Maschinenethik. *Die Betriebswirtschaft*, 4/2014, 237–248.

Bendel, O. 2013a. Ich bremse auch für Tiere: Überlegungen zu einfachen moralischen Maschinen. *inside-it.ch*, December 4, 2013. http://www.inside-it.ch/articles /34646.

Bendel, O. 2013b. Buridans Robot: Überlegungen zu maschinellen Dilemmata. *Telepolis*, November 20, 2013. http://www.heise.de/tp/artikel/40/40328/1.html.

Bendel, O. 2012. Maschinenethik. *Gabler Wirtschaftslexikon.* Wiesbaden: Springer Gabler. http://wirtschaftslexikon.gabler.de/Definition/maschinenethik.html.

Bostrom, N., and Yudkowsky, E. 2014. The Ethics of Artificial Intelligence. In Frankish, K., and Ramsey, W. M. eds. *The Cambridge Handbook of Artificial Intelligence*. Cambridge: Cambridge University Press, 316–334.

Bradl, N. 2015. Autonome Lkw: Future Truck soll schnell auf die Straße. *LOGISTIK HEUTE*, July 28, 2015. http://www.logistik-heute.de/Logistik-News-Logistik-Nachrichten/Markt-News/13347/Daimler-Vorstandsmitglied-rechnet-mit-Genehmigung-in-den-naechsten-Wochen-Au.

Deghani, M.; Forbus, K.; Tomai, E.; and Klenk, M. 2011. An Integrated Reasoning Approach to Moral Decision Making. In Anderson, M., and Anderson, S. L. eds. 2011. *Machine Ethics*. Cambridge: Cambridge University Press, 237–248.

Deng, B. 2015. Machine ethics: The robot's dilemma. *nature*, July 1, 2015. http://www.nature.com/news/machine-ethics-the-robot-s-dilemma-1.17881.

Emmerth, D. 2013. Ein Zimmer aus dem 3-D-Drucker. *Tages-Anzeiger*, October 22, 2013. http://www.tagesanzeiger.ch/wissen/technik/Ein-Zimmer-aus-dem-3DDrucker/story/11214726.

Federle, S. 2014. Radar soll Zugvögel schützen. *Tierwelt*, Nr. 10, March 5, 2014, 22–23.

Goodall, N. J. 2014. Ethical Decision Making During Automated Vehicle Crashes. *Journal Transportation Research*, September 29, 2014, 58–65.

Häuptli, L. 2014. Kampf den Drohnen. *NZZ am Sonntag*, December 7, 2014. http://www.nzz.ch/nzzas/nzz-am-sonntag/kampf-den-drohnen-1.18439765.

Holzer, H. 2015. Wer programmiert die Moral für die Maschine? *Handelsblatt*, January 28, 2015. http://www.handelsblatt.com/auto/nachrichten/autonome-fahrzeuge-wer-programmiert-die-moral-fuer-die-maschine/11295740.html.

Kolhagen, J. 2013. Autopiloten auf Rädern. *Versicherungswirtschaft*, June 1, 2013, Nr. 11, 70.

Kopf, M. 1994. *Ein Beitrag zur modellbasierten, adaptiven Fahrerunterstützung für das Fahren auf deutschen Autobahnen.* Dissertation. Düsseldorf: VDI-Verlag.

Lorenz, L. M. 2014. *Entwicklung und Bewertung aufmerksamkeitslenkender Warn- und Informationskonzepte für Fahrerassistenzsysteme: Aufmerksamkeitssteuerung in der frühen Phase kritischer Verkehrssituationen.* Dissertation. München.

Pellkofer, M. 2003. *Verhaltensentscheidung für autonome Fahrzeuge mit Blickrichtungssteuerung.* Dissertation. München. http://athene-forschung.unibw.de/doc /85319/85319.pdf.

Stoller, D. 2013. Vollautomatisch und ohne Fahrer in der Stadt unterwegs. *Ingenieur.de*, July 15, 2013. http://www.ingenieur.de/Themen/Automobil/Vollautomatisch-Fahrer-in-Stadt-unterwegs.

Wallach, W., and Allen, C. 2009. Moral Machines: Teaching Robots Right from Wrong. Oxford: Oxford University Press.