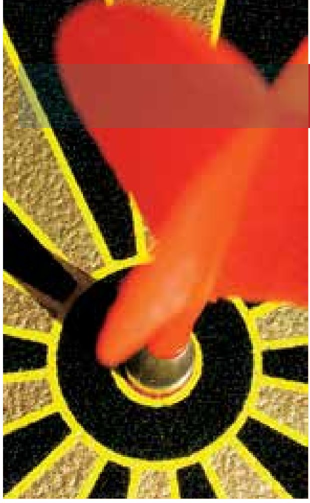

Editor: **Daniel Zeng**, University of Arizona and Chinese Academy of Sciences, zengdaniel@gmail.com

Fully Autonomous Driving: Where Technology and Ethics Meet

Dieter Birnbacher, *University of Duesseldorf*
Wolfgang Birnbacher, *IBEO Automotive Systems GmbH*

Machines tend to be superior to humans in terms not only of strength and precision but also of reliability in controlling complex processes and the capacity to learn from mistakes. In motorized traffic, the advantages of fully autonomous driving are evident. Trucks driving in a densely packed convoy would greatly reduce gas consumption by reducing air resistance; autonomous cars that allow passengers to make good use of their travel time would greatly reduce time wasted. Above all, autonomous car traffic holds the promise of massive increases in safety. Especially in situations where presence of mind and reaction time are crucial, machines can be expected to perform better than humans. Instead of emotional and reflex-like reactions, a machine can analyze a situation in a split second and make decisions based on an algorithm established long in advance.

The prospect of automatized car traffic, however, confronts ethics, law, and politics with novel and far-reaching questions. Even if autonomous cars are constructed in a way that makes traffic safety the top priority, critical situations in which loss of life and limb are inevitable could still arise, requiring them to negotiate between two or more evils. How, for example, should the autonomous car react if it must choose between risking serious damage to one or more passers-by or risking serious damage to one or more of its passengers?

At this point, it is important to clearly distinguish the tasks of technology from ethics and the haunting challenges confronting each. On the technology side, an autonomous car's system of sensors and control algorithms must be able to substitute for a human driver and even surpass a driver in relevant capacities. To achieve this, normal signal processing techniques (sensor fusion, object classification, and so on) are not enough. The algorithms must incorporate large parts of a human driver's accumulated experience. Current sensor data must be integrated and processed together with the acquired understanding of complex contextual relations. The result of this process will be a statistical situation assessment with probabilities (risk and gain) for different possible reactions. Given the complexity and diversity of possible scenarios, the "right" reaction cannot be programmed in advance for every concrete situation, but has to be calculated according to a defined algorithm.

Despite the enormous complexity of these tasks, the challenges are purely technical. As with a driver's test, requirements for the quality of this situation assessment can be defined and tested with real and virtual test drives.

The ethical tasks to be mastered are no less challenging. Several hard questions must be answered:

- Who should have the authority to decide whether the preference rules and learning skills programmed into the system are acceptable?
- How much differentiation in ethical programming should be allowed to different stakeholders (producers, users, societies)?
- Who is responsible in case of damage?

Errata

In the article, "Using Process Mining to Model Multi-UAV Missions through the Experience," by Juan Jesús Roldán, Jaime del Cerro, and Antonio Barrientos (http://ieeexplore.ieee.org/document/8012327/), there are two errors in Table 1, specifically in column 3, rows 7 and 8. The IM-Inf should be "Model," whereas the IM-Inc should be "No model." We regret any confusion this error has caused.

Although decisions in dilemmatic situations have become popular as exercises in ethical judgment in university courses and even in high schools, they concern rare cases and are marginal in comparison with the more central questions confronting society. After all, programming a certain risk behavior into a machine not only has consequences in critical situations but also defines the driving style generally. How safe is safe enough? How safe is too safe? Excessive safety would paralyze road traffic and seriously hamper acceptance of autonomous vehicles. Giving leeway to risky driving styles would jeopardize the safety objectives. How egalitarian does an automatized driving system have to be? Is a manufacturer allowed to advertise with fast cars at the price of lowered safety for other road users?

Empirical studies suggest that a great majority of people prefer a decision algorithm that minimizes overall damage. At the same time, they seem to be prepared to accept reductions of their own safety as long as everyone accepts these same risks. This seems to imply that the level of acceptable risk can be calculated as a utilitarian optimum over all users, and that the risks produced by individuals is equal for all autonomous vehicles. It goes without saying that an egalitarian decision algorithm along these lines would lead to a radical shift of responsibility from the individual to the public. Neither the owner nor the passengers could be held responsible for the behavior of the vehicle any longer since risk preferences and conflict solving are determined in advance by societal consensus, leaving no room for individual intervention. The same holds for producers. Since the vehicle's decisions and reactions follow socially established norms, producers can no longer be held responsible for damages that occur as a consequence of these norms.

This means that society, represented by its respective legislators, must establish the general principles of the decision algorithm and assign weights to the different kinds of evils in a socially acceptable way. More than any other area, technology, once put into use, is inseparable from ethics.

Even if a social consensus might be achieved on a broadly utilitarian ethical framework, there is plenty of room for controversy. One question is how to deal with questions of fairness—for example, between vehicles that are more and less vulnerable in crashes. A decision algorithm programmed to minimize harm would choose collision with a smaller vehicle in an unavoidable swerve maneuver over collision with an "armored" SUV because the latter would involve greater risk for its passengers. Owners of smaller cars who cannot afford bigger ones would likely therefore feel discriminated against. There might be controversy also about the weight assigned to different goods in potential conflict situations, such as the bodily integrity of humans and animals.

Acceptance of autonomous driving will depend on how far a consensus on these norms can be found, first among experts, then in society at large. One ethical condition, however, should be crucial: in no case should the ethical algorithms be put in practice as nontransparent black boxes. The built-in norms should, as far as possible, be understood and commonly shared. ■

Dieter Birnbacher is a professor of philosophy at the University of Duesseldorf. His research interests include ethics and applied ethics. Birnbacher has an honorary doctorship in philosophy from the University of Muenster, Germany. He is a member of Leopoldina, National Academy of Sciences. Contact him at dieter.birnbacher@uni-duesseldorf.de.

Wolfgang Birnbacher is an FPGA system designer at IBEO Automotive Systems GmbH. His research interests include sensor algorithms and signal processing. Birnbacher has an MASc in microelectronic systems from HAW Hamburg. He is a member of IEEE. Contact him at wolfgang.birnbacher@ibeo-as.com.

This article originally appeared in IEEE Intelligent Systems, *vol. 32, no. 5, 2017.*