Meaningful Human Control over Automated Driving Systems

THIS ISSUE'S TOPICS

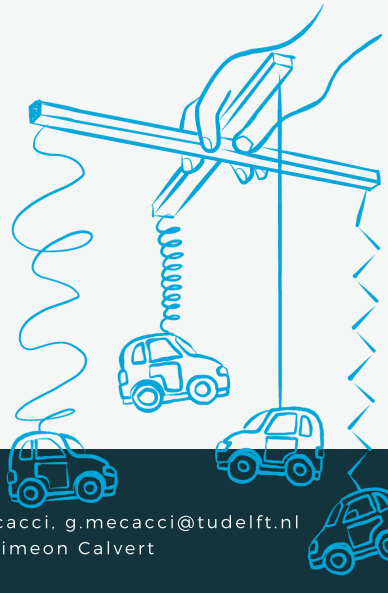
THE OPINION
Time to get real: automated vehicles are a game changer, but are we all playing the same game?
by *Simeon Calvert*

THE EDITORIAL
Is meaningful human control even possible?
by *Giulio Mecacci*

THE REPORT
Part of our team visited San Francisco's Automated Vehicle Symposium to discuss meaningful human control.

NEWS FROM OUR PROJECT
Our updates on research and things we do in our daily life of hard working researchers.

PUBLICATIONS AND DISSEMINATION
A short list of our most recent publications, the events we attended, and those we are planning to attend.



COLOPHON

Meaningful Human Control newsletter
TU Delft

Editor-in-chief: Giulio Mecacci, g.mecacci@tudelft.nl
Editors: Daniël Heikoop, Simeon Calvert

Interior view of General Motor level 4 vehicle (source: GM)

THE OPINION

TIME TO GET REAL: AUTOMATED VEHICLES ARE A GAME CHANGER, BUT ARE WE ALL PLAYING THE SAME GAME?

by *Simeon Calvert*

Earlier this year when reading about yet another Tesla autopilot crash, this time colliding with a parked police vehicle in California, the despair of so many warnings came over me again. With increasing regularity, we can read about similar accidents occurring (and not just with Tesla's). And it has long been common knowledge in (traffic) psychology that human drivers do not possess sufficient ability to monitor automated systems, while being left with few other tasks to perform. Time and time again, this has been demonstrated to lead to inattention, distraction and drowsiness. Furthermore, giving a driver a piece of technology that has abilities beyond what is permitted is like giving a toddler a box of chocolates and only allowing them to eat one. The abilities and benefits of Autopilot and other similar driving technologies are well publicised, so how can we expect a driver not to want to use them to their advantage!? It would seem to me that we are not being realistic in regard to what we are asking drivers.

When such accidents occur, both manufacturer and even police investigators will often refer to the operational design domain (ODD) and stated restrictions of the automated driving technology that is, for example, laid out in the user's manual. By doing this, they place responsibility of control firmly at the feet of the driver. "Oh, the driver was meant to monitor, but did not and is to blame!", "Oh oh, the driver accepted responsibility and should have been aware that the system is not designed to be used in that way". Is this fair? I would say that proper responsibility is not something that should merely be attributed, but also has to depend on reasonable ability. In a paper we recently submitted, we argued that responsibility should follow reasonable and meaningful human control. Such meaningful human control would ensure that drivers are assigned responsibilities that reasonable and would address the design of automated vehicles to ensure proper control and responsibility can be attributed. And not just any control, but control that also leads back to ethically acceptable human norms and values.

Enough on control and responsibility and back to the accidents, and again from a viewpoint of realism: accidents with automated vehicles seem to be big news. Maybe that's fair, but much of the surprise and sensation that goes with it is maybe not as fair. Automated vehicles are not and will never be infallible, therefore accidents will happen. And to the dismay of many, new types of accidents that wouldn't occur with manual vehicle will also occur. Of course this throws up ethical and legal questions, which I will not address here. But the bigger picture needs to be more focussed on the overall effect that automated vehicles will have. If our focus is on safety, then that should entail the totality of accidents, injuries and fatalities. If vehicle automation leads to a total reduction, then surely that is a good thing! But accidents will still occur and that is something that needs to be accepted (note: this does not mean though that we should not aim for increased safety!).

So to conclude, I probe society and industry alike to retain a broad perspective on what vehicle automation will bring and what the consequences may be. In general, automation will be good for mobility, but we need to lose much of the exaggerated utopian perspectives. Automated vehicles should be implemented to their strengths and not where they are not (yet) ready or suitable. Likewise humans should be incorporated in the system to make use of their strengths. Merely asking them to monitor does not do this. Meaningful human control can allow us to understand this to a better extent. It is certainly not all doom and gloom. My conclusion is that automation in road traffic has much potential, therefore: Let us all be realistic and proceed with optimistic realism, but let us also be mindful of avoiding blind utopianism.

Disclaimer: this article expresses the author's personal opinion, and does not necessarily represent the position of the research team.

Copyright Tim Hayes, Huffington Post

THE EDITORIAL

A POSSIBLE MEANINGFUL HUMAN CONTROL

by *Giulio Mecacci*

A few days ago, Humanitarian Law & Policy blog published a piece by Elke Schwarz titled "The (im)possibility of meaningful human control for lethal autonomous weapon systems".War drones are not our core business, but self-driving cars share in many senses identical issues. To some extent, those too can be dangerous if out of control. In this editorial, I would like to comment on that piece, and explain why meaningful human control might not be as impossible as the author seems to suggest.

The author outlined three very good reasons why a meaningful form of human control might be hardly achievable over all kinds of systems that involve artificial intelligence for their functioning. The first reason concerns cognitive limitations produced in human-machine interactions. This has to do with the fact that us humans tend to deploy superficial and automatic reasoning when interacting with highly autonomous systems. This would drastically limit the quality of our contribution to the overall decision-making process. The second problem the author highlights is that humans cannot access and effectively make use of the amount of information that an artificial intelligence can process. That is why we use computers and assistive technologies, after all. Making meaningful decisions is hard if we cannot see the grander picture, so to speak. A third issue would regard temporal limitations. Humans are slow and clunky when it comes to timely and resolutely intervene to steer a decision that a machine has just suggested (or plainly made).

If we are unable to timely understand, question and act upon artificially intelligence mediated (moral) decisions, how will we ever be able to reasonably keep those machines under control? And how are we, as humans, going to bear the responsibility for those decisions, if they are not really ours? I have been toying with artificial intelligence and human cognition for long enough to confidently say that I fully agree with Schwarz's points. Yet, I do not entirely agree with her conclusion.

The three arguments the author makes are only good for as long as one endorses a certain notion of meaningful human control. "Meaningful" does not equal to "more". Surely not more of the same kind. If we stick with a classic notion of control, where a human is in control for as long she is directly, operationally so, then Schwarz is right. Artificial intelligence will make that kind of control pretty hard to achieve and maintain. Almost by definition. More AI means more hidden variables between us and the system's behavior, and thus less predictability. After all, the reason why we use artificial intelligence is that we want it to think and act in our stead, and we want it to do that better than we do, too.

But meaningful human control, we believe, is not a quantitatively superior form of control. Rather, it's a qualitatively different one that carves the nature of control at different joints. It is not about how well a system responds to our actions, which is a good way of measuring control for "dumb" devices, but focuses on a different kind of responsiveness. Namely, responsiveness to human (moral) reasons. It is not the kind of control that you have on a coffee machine, that waits for you to press the switch. It is rather the kind of control we can have on a well trained horse, where you can loose the reins a bit while being reasonably confident that she will always do what you would do. This is what meaningful human control over intelligent systems will look like. A kind of control that ought to be obtained not only through careful design and cautious experimentation, but also by profoundly revisiting our entire idea on what controlling really amounts to. And yes, we are quite confident that it is not that impossible.

Disclaimer: this article expresses the author's personal opinion, and does not necessarily represent the position of the research team.

THE REPORT

A VERY CAUTIOUS AVS SYMPOSIUM

by *Giulio Mecacci*

We have been dreaming of California for a while, and we finally got to visit San Francisco. The City (as those from the Bay call it) gifted us with one of its brightest weeks. We didn't get to see the infamous fog, only the beautiful sun of northern California.

The Automated Vehicle Symposium (AVS) took place in a luxurious venue, the Union Square's Hilton Hotel. There, we were co-organizing a 'breakout session', a small discussion group on a particular topic. In this case, it was indeed meaningful human control. We held this meeting under the Chatham House Rule, therefore I won't be allowed to make names or specific references.

Different topics were treated, most of them regarding the impact of autonomous driving systems on human values. Not only safety, but also accountability and the fundamental civil rights that autonomous driving systems can threaten or promote. Mobility in general was discussed in a broad fashion, as an important component in what grants us freedom and individual autonomy. A number of societal values were recognized to be at stake if not accounted at the earliest possible stages of the development of this technology. That's what we were there for.

This general attitude of caution was in a way the gist of the symposium. Public acceptance of the technology is an important step to achieve its full potential, and that can only be built through transparency, education, and the involvement of the greater possible number of stakeholders. As reminded by the US Department of Transportation Secretary Elaine Chao in her keynote, the limits of current technology should be made clear to the public, and expectations should be kept in line with reality.

In the days after the symposium, several outlets picked up this caution message quite well, as showed for instance by Aarian Marshall's good piece on Wired*, Home from the honeymoon, the self-driving car industry faces reality*. We will keep doing our share to ensure a steady, yet value-centered growth of self-driving technology. We are glad to see that the community is growingly recognizing our approach, and that we will actually be able to make a difference in the years to come.

Hilton hotel, Union Square

NEWS FROM OUR TEAM

Dr. Simeon Calvert, traffic engineer

Much of the recent work on the MHC project from the engineering track has focussed on two things: development of the modelling framework, and on gaining greater empirical evidence of driver vehicle dynamics, which will allow us to model with greater accuracy.
The modelling framework is being designed from the perspective of including driver cognitive behaviour as many traditional traffic simulation models omit many behaviour dynamics and therefore simulate very robotic like traffic. If we are modelling automated vehicles, which are inherently more 'robotic', then a new type of simulation model for regular vehicles is also required (This was proposed earlier in the year by myself and Hans van Lint). A main focus in regard to automated driving has been on the transition of control from automated to manual driving and on different vehicle dynamics. Progress has been steady and is beginning to produce promising results. We aim to produce a paper with the conceptual description to be submitted by the end of 2018 and to be able to use the model to analyse cases during 2019.
Obviously, such a model requires insights into what actual behaviours and dynamics of drivers and vehicles are. In practice, these insights based on empirical evidence, are scarce. Therefore, effort is being made to collaborate with various FOT's and also to set-up our own experiments to feed the model calibration and design. A planned experiment for later this year will analyse the transition of control from a task demand perspective.
A final note may be made that the project is gaining visibility within the scientific world, and I have accepted an invitation to give an additional presentation at a workshop during the IEEE ITSC conference later in this year, besides other invites that other project members have had for various gatherings.

Dr. Daniël Heikoop, behavioral scientist

The investigation and quantification of human behaviour in relation to driving with automated driving systems has now been completed and written down, ready to be subjected to scientific opinion. The results from this research will be presented at the Driver Distraction and Inattention conference in Goteborg, Sweden, coming October, and preliminary results have already been presented at the HUMANIST conference in The Hague, The Netherlands.
The next step, acquiring expert opinion regarding a novel, human-oriented approach of a framework for automated driving systems, has been set in collaboration with CBR, where a focus group meeting has been held with a group of highly enthusiastic driving examiners. Their input from that evening will now be used to develop a recommendation towards a framework of levels of automation that defines how automated a driving system is from a human perspective, rather than from a technical perspective, and empirical assessment of that framework is consequentially envisioned in the near future as well.
Relatedly, active collaboration for research proposals that build upon and extend the findings from this project has so far been productive, and this endeavour will be continued to safeguard this project's legacy.
Furthermore, internal collaboration within our project has now produced several submitted articles and empirical research plans of which its execution will commence in the near future. We have only just begun.

Dr. Giulio Mecacci, philosopher of technology

Philosophia ancilla scientiarum. One of the main aims of the philosophy of technology section is to help out in framing scientific research within a bigger picture, interpreting its societal relevance and its ethical implications. Philosophy also deals with conceptual clarification, trying to make hard things a bit easier and more accessible. With that in mind, we have offered substantial contributions to a number of different investigations within our project. In the meantime, the foundational philosophical paper has finally come to an end, awaiting final revisions before submission. There, the notion of meaningful human control is further investigated. The paper provides tools for the assessment of control in (semi) autonomous system and indications on possible ways to design for meaningful human control. If on the one hand it is important to have clear ways to understand control in autonomous systems, on the other hand this is only relevant for the benefits it brings about. For instance, by determining the degree of control, one can determine the degree of responsibility. Our next step is to investigate how control and responsibility could, and should, be connected when autonomous driving systems come into the equation.

PUBLICATIONS AND DISSEMINATION


Maui

- Attendance at the 6th International Conference on Driver Distraction and Inattention, October 15-17, 2018, Gothenburg, Sweden.
- Presentation at Radboud University. September 26, 2018, Nijmegen, The Netherlands.
- Attendance at the Kennisagenda @ Connekt. September 11, 2018, Delft, The Netherlands.
- Attendance at the Automated Vehicles Symposium. July 9-12, 2018, San Francisco, USA.
- Presentation at meeting of the Management Board of Italian State Railways. June 25, 2018, Rome, Italy.
- Presentation at the Universidad Nova Lisbon. May 4, 2018, Lisbon, Portugal.
- Presentation at the University of Turin. April 5, 2018, Turin, Italy.
- Attendance at the NWO-MVI conference, January 19, 2018, Muntgebouw Utrecht, The Netherlands.
- Attendance at the Kennisagenda @ Connekt, September 5, 2017, Delft, The Netherlands.
- Invited lecture at Politecnico Milano, Italy. "Ethics in the Age of Autonomous Transportation".
- Co-organized workshop on "Social and Ethical Implications of Autonomous Vehicles" at University of Vienna.

- D. D. Heikoop, M. Hagenzieker, G. Mecacci, S. Calvert, F. Santoni de Sio, & B. van Arem (submitted). "Human behaviour with automated driving systems: A quantitative framework for meaningful human control".
- S. C. Calvert, G. Mecacci, B. van Arem, F. Santoni de Sio, D. D. Heikoop & M. Hagenzieker (submitted). "Gaps in the control of automated vehicles on roads".
- S. C. Calvert, D. D. Heikoop, and B. v. Arem. (submitted) "Core components of automated driving for meaningful traffic flow and safety".
- G. Mecacci & F. Santoni de Sio (in progress). "Meaningful Human Control, practical reasoning, and dual-mode vehicles".
- S. C. Calvert, G. Mecacci, D. D. Heikoop & F. Santoni de Sio (accepted). "Full platoon control in Truck Platooning: a Meaningful Human Control perspective". IEEE ITSC conference, Nov 2018, Maui, USA.
- S. C. Calvert, and D. D. Heikoop, (accepted). "Vehicle Automation Design to Maintain Meaningful Human Control for Driver Inattention", *6th International Conference on Driver distraction and Inattention*, October 15-17, 2018, Gothenburg, Sweden.
- D. D. Heikoop., & M. Hagenzieker, (accepted). "Working Towards a Meaningful Transition of Human Control over Automated Driving Systems", *6th International Conference on Driver Distraction and Inattention*, October 15-17, 2018, Gothenburg, Sweden.
- F. Santoni de Sio & J. v. d. Hoven. (2018) "Meaningful Human Control Over Autonomous Systems: A Philosophical Account", *Frontiers in Robotics and AI*.
- D. D. Heikoop, S. C. Calvert, G. Mecacci, M. Hagenzieker, F. Santoni de Sio, & B. van Arem. "Meaningful Human Control over Automated Driving Systems", 6th HUMANIST Conference, June 13-14, The Hague, The Netherlands.


San Francisco Bay (panorama)