Ethics and Information Technology (2020) 22:335–344
https://doi.org/10.1007/s10676-018-9445-9

ORIGINAL PAPER

Automated cars meet human drivers: responsible human-robot coordination and the ethics of mixed traffic

Sven Nyholm[1] · **Jilles Smids**[1]

Published online: 27 January 2018

Abstract
In this paper, we discuss the ethics of automated driving. More specifically, we discuss responsible human-robot coordination within mixed traffic: i.e. traffic involving both automated cars and conventional human-driven cars. We do three main things. First, we explain key differences in robotic and human agency and expectation-forming mechanisms that are likely to give rise to compatibility-problems in mixed traffic, which may lead to crashes and accidents. Second, we identify three possible solution-strategies for achieving better human-robot coordination within mixed traffic. Third, we identify important ethical challenges raised by each of these three possible strategies for achieving optimized human-robot cordination in this domain. Among other things, we argue that we should not just explore ways of making robotic driving more like human driving. Rather, we ought also to take seriously potential ways (e.g. technological means) of making human driving more like robotic driving. Nor should we assume that complete automation is always the ideal to aim for; in some traffic-situations, the best results may be achieved through human-robot collaboration. Ultimately, our main aim in this paper is to argue that the new field of the ethics of automated driving needs take seriously the ethics of mixed traffic and responsible human-robot coordination.

Keywords Human-robot coordination · Automated driving · Ethics · Responsible robotics · Agency

Introduction

Before 2015, discussions of crashes involving automated vehicles were largely hypothetical. However, with increased road-testing of automated vehicles, real world crashes soon started happening, with just under 20 cases in 2015. The initial crashes were primarily instances of conventional cars rear-ending slow-moving automated vehicles. And there was little damage done (Schoettle and Sivak 2015a). However, in 2016 there were some more dramatic developments. On Valentine's day (February 14), there was a not very romantic encounter between a "self-driving" Google-car and a bus. The former crashed into the latter. And on this occasion, Google had to assume responsibility for the collision, which was the first time that happened (Urmson 2016). More tragically, the first person was killed in a crash with a vehicle operating in automated mode in May. A Tesla Model S in "autopilot" mode collided with a truck that the car's sensors had not detected (Tesla 2016). What all these crashes so far—both those in 2015 and 2016—have in common is that they were collisions between automated cars and conventional cars. They were crashes in "mixed traffic."

This paper is a contribution to the new field of the ethics of automated driving (e.g. Goodall 2014a, b; Lin 2015; Hevelke and Nida-Rümelin 2014; Gurney 2016; Gogoll and Müller 2016; Nyholm and Smids 2016; Nyholm forthcoming). Its aim is to argue that this field should take mixed traffic very seriously. There are distinctive ethical issues related to how to achieve compatibility between automated vehicles and human-driven conventional vehicles that do not reduce to the main issues thus far mostly discussed in the ethics of automated driving. That is, there are ethical issues related to compatibility-challenges that do not reduce to how automated cars should be programmed to handle crash-scenarios or who should be held responsible when automated vehicles crash.[1] The ethics of automated driving also needs to deal with other key issues. And among those is the issue of responsible human-robot coordination: how to adjust robotic

✉ Sven Nyholm
s.r.nyholm@tue.nl

[1] Eindhoven University of Technology, Eindhoven, The Netherlands

[1] As that sentence implies, the ethics of automated driving has so far primarily focused on how to program automated vehicles to react to crash-scenarios (e.g. Goodall 2014a, b; Lin 2015; Gurney 2016;

driving and human driving to each other in a way that is sensitive to important ethical values and principles.[2]

It might be suggested that this is a minor issue. Eventually, we might only have automated vehicles on our roads. So this is just a transition-period worry. To this we respond as follows. Even if highly or even fully automated vehicles will at some later time come to dominate the roads, there will still be a long transition-period during which mixed traffic will be a problem that needs to be dealt with (van Loon and Maartens 2015). Nor should we assume that full automation in all vehicles is an end-point towards which we are moving with necessity (Mindell 2015); mixed traffic may come to mean a mix of vehicles with different levels and types of automation interacting with each other on the road (Wachenfeld et al. 2015; Yang et al. 2016).

In either kind of mixed traffic, there will be different types of vehicles on our roads with different levels and types of automation.[3] This will have two important consequences, similar to what we are already seeing today. Firstly, there will be incompatibilities in the ways these cars function and interact with each other, which will create new traffic-risks. Secondly, the vehicles on the road will have different crash-risk levels: certain kinds of cars will pose greater threats to others; and certain kinds of cars are going to be safer to be in when crashes occur than other cars are (Cf. Husak 2004). In light of these two observations, we do the following three things in this paper.

Firstly, we describe in general terms why there are incompatibilities between what we will call robotic driving, on the one hand, and human driving, on the other. That is, we describe why we think the functioning of automated cars and the driving-styles of human beings lead to compatibility-problems, meaning that there is a need to think about how greater compatibility might be achieved within mixed traffic.

Footnote 1 (continued)

Gogoll and Müller 2016; Nyholm and Smids 2016) and who should be held responsible for crashes (e.g. Hevelke and Nida-Rümelin 2014; Gurney 2016; Nyholm forthcoming).

[2] Ideally—as an anonymous reviewer suggested—human drivers would naturally adapt to automated cars, while purely technical solutions could be found to help automated cars adapt to human drivers. This, it might be thought, would help us to avoid turning human-robot coordination within mixed traffic into an ethical issue. We share the hope that drivers will ultimately turn out being able to adapt well to automated cars, but think that one cannot simply stand by and hope that this will happen. That would be irresponsible. Secondly, given the risks involved, and the further reasons we present below, the choice among technical solutions for adapting automated cars to human-driven cars is not an ethically neutral, purely technical matter.

[3] Additionally, automated cars also have to coordinate their driving with the behavior of pedestrians, animals, and people on bikes and motorcycles. Like human-driven conventional cars, pedestrians and bikers also don't behave like robots. So these are further human-robot coordination problems.

This takes us to the second thing we do, which is to present some of the main options there are for how to achieve better human-robot coordination in this domain. Thirdly, we consider what types of general ethical issues and challenges we need to deal with when we make these choices about how to achieve greater compatibility between automated cars and conventional cars within mixed traffic. For example, we will consider issues to do with respecting people's freedom and human dignity, on the one hand, but also positive duties to promote safety and to manage risks in responsible ways, on the other hand.

Human-robot coordination-problems in mixed traffic

The reasons why incompatibilities arise are fairly easy to explain and understand (van Loon and Maartens 2015; Cf. Yang et al. 2016). They have to do with the different ways in which automated cars and human drivers function as "agents" (i.e. as entities that act according to certain basic goals and principles). This includes the different ways in which automated cars and human drivers form expectations about other vehicles on the road. In explaining these incompatibilities, we will start with key differences in how goals are pursued and then continue with differences in how expectations are formed by automated cars and human drivers.

First of all, automated cars are a kind of artificial or robotic agents of at least a basic kind. They pursue goals, and do so in a way that is responsive to continually updated representations of the environment they operate in. This makes them into a kind of robotic agents, though of course ones designed by human agents (Nyholm forthcoming).[4] More specifically, automated cars are designed to reach their destinations in ways that are optimally safe, fuel-efficient, and travel time-efficient (e.g. by reducing congestion) (van Loon and Martens 2015).

This optimization-goal has a profound impact on the driving-styles of automated cars, making them markedly different from those of most human drivers. For example, in order to achieve fuel-efficiency and avoid congestion, automated cars will not accelerate vigorously, and brake very gently. Safety-enhancing aspects of their driving-styles include avoiding safety-critical situations, e.g. by staying longer behind a cyclist before overtaking (Goodall 2014b). More generally, at least at present, automated cars are programmed to follow the traffic rules very strictly in most situations.

[4] For more discussion of how to apply the concept of agency to entities that are not individual human beings, see also (Floridi and Sanders 2004; List and Pettit 2011).

One major function of these rules is precisely to enhance safety. Thus, under current engineering ideals, automated cars always give way when required, avoid speeding, always come to a stand-still at a stop-sign, and so on.[5]

Let us consider how this contrasts with human drivers. Human beings are, of course, also agents who pursue driving goals in traffic-situations they have to adequately perceive and represent. And humans also act on the basis of principles and rules (Schlosser 2015). Unlike robotic cars, however, humans exhibit satisficing rather than optimizing driving behavior (van Loon and Martens 2015). That is, they drive just well enough to achieve their driving-goals. This may include all kinds of driving-behavior that is not optimal in terms of safety, fuel-efficiency, and traffic flow: speeding, aggressive accelerating and decelerating, keeping too short following-distances, and so on. Moreover, this often involves bending or breaking traffic rules. Hence automated cars and human drivers have rather different driving-styles. The former are optimizers and strict rule-followers, the latter satisficers and unstrict rule-benders.

Consider next how self-driving cars and human beings perceive one another and form expectations about how other cars are likely to behave in different traffic-situations (van Loon and Martens 2015; Wolf 2016). Automated cars will become able to communicate with other automated cars using car-to-car information- and communication-technologies. But they will not be able directly communicate with human drivers in that way.

Instead, according to traffic-psychologists Roald van Loon and Marieke Martens, automated cars will typically form their expectations about the behavior of conventional cars on the basis of externally observable behavioral indicators, such as speed, acceleration, position on the road, direction, etc. The problem here is that, currently, "our understanding of these behavioural indicators lacks both quantification and qualification of what is safe behaviour and what is not" (van Loon and Martens 2015, p. 3282). We don't yet know how best to program automated cars to predict what is, and what is not, safe human behavior on the basis of the external indicators that automated cars can observe.

One potential way of making progress with respect to automated cars' ability to communicate with human drivers is indirect in nature. Human-driven cars could be made to closely monitor and to try to predict the behavior of their human drivers. The human-driven cars could then communicate these predictions to the automated cars. That way, the automated cars could make use both of their own observations and the predictions communicated to them by the human-driven cars, and then base their own predictions of the likely behaviors of the human drivers on this dual basis.[6] This could constitute an improvement. But it would still not be direct communication between automated cars and human drivers. Rather, it would be communication between the automated cars and the human-driven cars, where the latter would join the automated cars in trying to predict what the human drivers are likely to do, also based on externally observable behaviors.

For human drivers forming expectations about automated cars, the problem is slightly different.[7] In the process of becoming habitual drivers, humans acquire lots of expectations regarding driving-behaviors of other cars in various situations. These expectations often do not fit very well with automated cars. For example, an automated car might keep waiting where the human driver behind expects it to start rolling. So, in order to be able to fluently interact both with other human drivers and automated cars, humans need to simultaneously operate on the basis of two parallel expectation-forming habits. They would have to operate on the basis on expectation-forming dispositions applying to conventional cars, on one hand, as well as ones applying to automated cars, on the other hand. That is a heavy cognitive load for human drivers to deal with.

Of course, in the case of other conventional cars, human drivers can communicate with other human drivers using various different improvised signals, such as hand- and arm-gestures, eye contact, and the blinking of lights (Färber 2016; Schoettle and Sivak 2015b). This helps human drivers to form expectations about how other human drivers will behave. But as things stand at the moment, human drivers cannot communicate with robotic cars in these improvised ways.

Given these differences between robotic driving and human driving, mixed traffic is bound to involve a lot of compatibility- and coordination-problems. The equation here is simple: clashing driving-styles + mutual difficulties in forming reliable expectations = increased likelihood of crashing cars.[8] So the question arises of how we ought to make automated cars and human-driven conventional cars

[5] However, as we discuss below in "Options for better human-robot coordination in mixed traffic", various different stakeholders are already debating whether sel f-driving cars should be programmed to break the law in order to better coordinate interaction with conventional cars.

[6] Thanks to an anonymous reviewer for this suggestion.

[7] Moreover, we do not know how well human drivers in general are able to understand and predict the behavior of fellow drivers (van Loon and Martens 2015, p. 3283).

[8] Some authors stress our current lack of understanding of the compatibility-problem. However, we have already sufficient reasons to be worried about negative consequences for the safety in mixed traffic. In fact, there are some first indications that automated cars are more often involved in crashes and collisions than conventional cars are (Naughton 2015; Schoettle and Sivak 2015a). Note also that whereas in all cases reported in the Schoettle & Sivak-study, human drivers

maximally compatible with each other. We need to achieve good human-robot coordination, and avoid crashes and accidents caused by various different forms of incompatibilities. What types of options are there? And what ethical issues are raised by the different types of options we face?

Options for better human-robot coordination in mixed traffic

In 2015, after the first mixed traffic collisions started being reported and analyzed, a debate about how to achieve better compatibility arose in various different domains. Opinions were expressed and debated in the media, engineering and traffic psychology-labs, consulting firms, in policy-making teams, and elsewhere, though not yet in the context of philosophical ethics. Most of the smaller crashes in 2015 were generally judged to be due to human error (Schoettle and Sivak 2015a). However, automated driving as it is currently functioning was nevertheless criticized. And some of the more recent incidents—particularly the 2016-crashes we mentioned in our introduction—have also been blamed on perceived shortcomings in the automated vehicles.

The main type of solution to human-robot coordination problems within this domain that one most commonly sees being discussed is the following: to try to program automated cars to function more like human drivers or to have them conform their robotic driving-styles to human driving-styles.[9] For example, one influential media outlet reporting on technology developments ran an op-ed in which automated cars were said to have a "key flaw" in being programmed to follow rules rigidly and drive efficiently: this causes humans to drive into them. The suggested solution: make automated cars less strict in their rule-following and less efficient in their driving (Naughton 2015). Similarly, a consultant advising the Dutch Ministry of Infrastructure and the Environment's Automated Vehicle Initiative (DAVI) suggested, at an interdisciplinary event on the ethics of automated driving, that automated cars should be equipped with "naughty software": software that makes automated cars break rules in certain situations in which many humans do so (Wagter 2016). This solution is also advocated by engineering-researchers Christian Gerdes and Sarah Thornton. They argue that, because human drivers do not treat traffic rules as absolute, automated cars should be programmed to do the same. Otherwise, they cannot coexist in human traffic and will not be accepted by human drivers (Gerdes and Thornton 2016).

Others have also mainly focused on this general option, while adopting a more skeptical approach to whether it should be taken. In a media interview, Raj Rajkumar, the head of the Carnegie-Mellon laboratory on automated driving, was quoted as saying that his team had debated both the pros and the cons of programming automated cars to break some of the rules humans tend to break (e.g. speed-limits). But for now, the team had decided to program all their experimental cars to follow the traffic-rules (Naughton 2015). Google, in turn, at one point announced that although they would have all their test-vehicles follow all rules, they would nevertheless try to program them to drive more "aggressively" to better coordinate with human driving (Ibid.).[10]

As we see things, there are three important problems with this strong focus on whether to program automated cars to behave more like human drivers, and with treating this as the main option to consider for how to achieve better human-robot coordination. Firstly, this assumes that full automation is the optimal solution for all traffic-situations and that if cars are going to behave like humans, this necessarily has to happen by means of programming the cars to be more human-like in their functioning. As David Mindell argues in a recent book about the history of automation, this assumption overlooks the more obvious solution for how to handle at least some forms of situations (Mindell 2015; Cf.; Kuflik 1999). It overlooks the option of not aiming for complete automation in all sorts of traffic-situations, but instead trying to create a fruitful human–machine collaboration whereby both the driver's human intelligence and the car's technology are put to work.[11] (Cf. Bradshaw et al. 2013) The best way to make automated cars function more like humans—if this

Footnote 8 (continued)

were at fault, we cannot yet conclude that human drivers have more problems interacting with automated cars than vice-versa. Human overseers of self-driving cars may have prevented additional accidents. And automated cars have so far been test-driven in fairly "safe" testing environments, e.g. Mountain View, California. Human-robot coordination will be much more difficult in really busy big cities, and in harsher weather conditions.

[9] It is also possible to pursue technological solutions for adapting automated cars to human-driven cars that do not make robotic driving more like human driving. We focus here on the idea of making robotic driving more like human driving for two reasons: firstly, this idea is frequently suggested, and secondly, it raises ethical issues of the specific sorts we particularly wish to highlight in this paper. However, a fuller discussion than what we can fit into this paper would also explore possible ethical issues related to adapting robotic driving to human driving in ways that do not involve making the latter more like the former.

[10] At another point, however, Reuters reported that Google was then willing to program their self-driving cars to speed up to 16 kph if safety were served by doing so. (Ingrassia 2014).

[11] Suppose, for example, that an automated car carrying a perfectly normal human adult is facing the following situation: the road is otherwise empty, but there is a large branch on in the car's lane. There is a double-line, meaning that strictly speaking, it is against the traffic-rules to briefly cross into the oncoming lane as a way of avoiding hitting the branch. For the artificial intelligence in the car's technology,

is a good idea in certain situations—may often be to simply involve the human, rather than to try to create artificial human reasoning or reactions in the car. As Mindell argues, we shouldn't simply assume that for all types of driving- or traffic-problems, full automation is always the ultimate ideal.[12]

Secondly, some of the human traffic-behaviors that automated cars' envisioned "naughty software" is supposed to conform to may be morally problematic and therefore not very appropriate standards to conform robotic driving to. Speeding is a key example here. Because it greatly increases risks beyond democratically agreed upon levels, speeding is a morally problematic traffic-offence (Smids forthcoming). As such, it is not a good standard to conform the functioning of automated cars to.

In general, we want to suggest that when different aspects of human driving vs. robotic driving are compared, and ways of conforming these to each other are sought, we should avoid any solutions that conform one type of driving to immoral and/or illegal aspects of the other type of driving. We should instead use morally and legally favored aspects of robotic or human driving as the standards to conform to, if possible. In many cases, this will mean that conforming robotic driving to human driving will be a bad idea.[13]

Thirdly, in primarily—if not exclusively—considering whether or not to conform certain aspects of robotic driving to human driving, there is another important alternative is also overlooked (…that is, in addition to the option of not always aiming for complete automation.). And that other option that we think ought also to be taken seriously is: to seek means for potentially conforming certain aspects of human driving to robotic driving. This could be done with changes in traffic-laws and regulations. But it could also be done with the help of certain kinds of technologies.

To use the speeding-example again, one way of making people more likely to adhere to speed-limits, in the ways that more "well-behaved" automated cars do, is to mandate speed-regulating technologies in conventional cars (Smids forthcoming). New conventional cars can be equipped with speed-regulating technologies; most old cars can be retrofitted with such technologies at reasonable cost (Lai et al. 2012). This would help to make humans drive more like robots, and there are sound reasons to expect that this will help considerably to solve speed-induced compatibility problems.[14] Or, to use another example, alcohol-interlocks in cars could also make humans drive a little more like robots. If all human drivers use alcohol-interlocks, they would consistently be more fully alert and concentrated than if they sometimes also have the option of driving while under the influence of alcohol (Grill and Nihlén Fahlquist 2012). Still another option is equipping conventional cars with forward collision warning-technologies.[15] This may potentially enhance drivers' prospective awareness of the risks they are facing. A heightened risk-awareness could enable human

Footnote 11 (continued)

it is a tough challenge to figure out whether this is a situation where it is a safe and a good idea to break the rules, but for the human in the car it is a no-brainer. This is one kind of situation in which rather than to program a completely automated car to think and behave like a human, the human driver can work together with the car to deal with this situation (cf. Färber 2016, p. 143). This does not need to amount to a complete hand-over of all functions, but could potentially be solved in some other way. For example, in airplanes, when pilots switch off some of the autopilot-features, pilots do not typically start performing all functions manually, but rather simply take over certain aspects of the operation of the airplane (Mindell 2015).

[12] An anonymous reviewer challenged us to come up with a general principle for when it is a good idea to involve the human and for when it is not a good idea to involve the human. We think that beyond trivial answers such as "we should do this whenever this would bring about better outcomes and capabilities", it is unlikely to be possible to come up with general principles that apply to all cases alike. Given different types of challenges and situations, different claims will apply; there is not an informative "one size fits all" type of principle that we can apply across the board to determine when it will be good, and when it will not be good, to involve humans.

[13] Of course, the argument for programming automated cars to break rules (e.g. to speed) is typically that this might enhance safety. However, that argument is typically made on the assumption that the only options we have is to do nothing (which might be unsafe) or program automated cars to break rules (which may enhance safety). As we now go on to argue in the next paragraph, there is another option that should also be discussed: namely, trying to conform human driving to robotic driving. For more on the ethics of speeding in particular, and possible technical solutions for how to deal with it, see (Smids forthcoming).

[14] Firstly, if conventional cars slow down, the need to program automated cars to speed in situation like merging with speeding traffic disappears, while the safety of its occupants is not jeopardized. Of course, this is only one traffic situation. More generally, retrofitting conventional cars with speed limiters strongly eases interpretation and prediction of the behavior of conventional cars on the part of automated cars and vice versa. For, secondly, if conventional cars cannot speed, there will be a significant reduction of the range of actual and potential behavior of conventional cars that automated cars need to interpret and predict. In addition, in cases where they still misinterpret or make the wrong prediction, conventional cars sticking to the speed limit allow automated cars more time to adjust. Taking the perspective of the human drivers, thirdly, these will no longer face situations in which, due to a lack of time caused by speeding in particular, they fail to adequately interpret (unfamiliar) behavior of automated cars. Having more time to consider and interpret the situation is one of the benefits of speed-limiters reported by participants of intelligent speed adaptation (ISA)-trials (Oliver Carsten, personal communication). In addition, since no cars will speed any more, the driving-styles of automated and conventional cars become more alike, and one source of ill-applied driver's expectations is eliminated. We are indebted to ISA-expert Oliver Carsten for valuable discussion of these points.

[15] For an example of this technology, see: http://www.mobileye.com/technology/applications/vehicle-detection/forward-colision-warning/.

drivers to better coordinate with robotic cars, which also have enhanced risk-detection-systems as part of their overall makeup.[16]

Ethical concerns regarding attempts to create better human-robot collaboration within mixed traffic

In the foregoing section, we identified three general solution-strategies for promoting better human-robot coordination in mixed traffic:

1. Trying to make certain aspects of robotic driving more similar to human driving;
2. Not assuming that complete automation is the optimal state, but also exploring ways of involving the human driver so as to create better human-robot coordination;
3. Seeking means for making certain aspects of human driving more like robotic driving.[17]

All three of these ways of improving human-robot coordination in mixed traffic raise potential ethical concerns. The aim of this section is to draw attention to some of the main concerns that need to be confronted when human-robot coordination issues in mixed traffic are explored and investigated in more systematic ways. We will here keep the discussion on a fairly general level as our chief aim in this paper is not to advocate any particular solutions, but rather to motivate further discussion of the ethics of mixed traffic.

As we have already noted, conforming robotic driving to human driving can be ethically problematic if the particular aspects of human driving we would be trying to adapt to are morally or legally problematic. In other words, we would not want to create a robotic agent that replicates morally problematic or illegal human behaviors (Cf. Arkin 2010). The main way in which this option ought to be evaluated morally, then, is through investigation of whether the human traffic-behaviors we would seek to conform robotic behavior to are morally and legally problematic. If they are, then it may be better to seek alternative solutions to the given human-robot coordination problems.[18]

What about the second option considered above, viz. investigating whether some coordination-issues might be better handled via human-robot collaboration rather than through attempts to make robotic driving more human-like? What sorts of ethical issues might this way of promotion human-robot coordination give rise to? The most obvious ethical issue here is whether the responsibilities drivers would be given would be too much to handle, or whether the average driver could reasonably be expected to discharge these responsibilities, whatever they might be.

In other words, it may be that some ways of achieving greater compatibility between highly automated cars and conventional cars is by keeping the former from being completely automated, and requiring the human driver to "help" the automated cars with some of the tasks they need to perform within mixed traffic. But at the same time, perhaps some of the ways in which humans could help out would be too difficult for most drivers.[19] If so, it would be ethically problematic to place those responsibilities on their shoulders.

This same general type of worry has already been discussed in relation to how automated cars should respond to dramatic crash- and accident-scenarios. For example, Alexander Hevelke and Julian Nida-Rümelin argue that it would be unfair to require people to step in and take over in crash-scenarios, because people cannot be expected to be able to react quickly enough (Hevelke and Nida-Rümelin 2014). In order for it to be fair and reasonable to expect humans to "help" their automated cars in accident-scenarios, it needs to be likely that the average driver would be able to perform the given tasks ("Ought implies can").

We agree with the general thrust of Hevelke and Nida-Rümelin's worries about requiring people to take over in crash-scenarios. However, it is important not to draw too close of an analogy between handing over control to the

[16] Technologies like speed-limiters and alcohol-interlocks have been around for a long time, yet they have not been widely adopted. Why? We suspect that there is a "status quo bias" at work here, whereby people are intuitively biased towards the way things are, even if it is not an optimal state of affairs (Bostrom and Ord 2006). The widespread introduction of automated cars has a disruptive potential, however, whereby widely held attitude towards currently available used and un-used traffic technologies are likely to change. Hence the introduction of a supposedly safer alternative—viz. highly or fully automated driving—will give drivers reason to re-think their attitudes towards safety-technologies not currently used, but already available for, conventional cars. For more on this type of argument, see the last few paragraphs of "Ethical concerns regarding attempts to create better human-robot collaboration within mixed traffic" below.

[17] A fourth possible solution—suggested by an anonymous reviewer—would be to separate automated cars from conventional cars, having them drive in different lanes, or on different roads. This would certainly solve the problem of having to coordinate human and robotic driving, and it might be possible in certain places. However, given limitations in available space for roads and people's preferences for where they will want to be able to get to using their cars, this solution-strategy will be unrealistic in many places.

[18] Since strategy 1 (viz. to try to conform robotic driving to human driving) was already criticized and fairly extensively discussed in the foregoing section, we are here rather brief about the first solution-strategy, instead focusing more on the other two.

[19] Cf. the concept of "controllability", as defined by the International Organization for Standardization (ISO). See, for instance, paragraph 1:19 of their ISO 26262 report on functional safety, available here: https://www.iso.org/obp/ui/#iso:std:iso:26262:-1:ed-1:v1:en.

human driver in accident-scenarios and all forms of human involvement in attempts to create better human-robot coordination within mixed traffic. Some conceivable ways of promoting human-robot coordination by involving the human driver in the operation of highly automated cars may be too demanding to be reasonable. However, there can surely also be ways of involving the human driver that are not too demanding.[20] More specific ethical evaluation of different possible ways of involving the human driver would first need to look at what exactly the humans would be required and expected to do. The next step would then be to make an assessment of whether these are tasks most operators of automated cars would be able to perform.

Turn now to the third solution-strategy under discussion: seeking means for conforming certain aspects of human driving to robotic driving. As we noted above, this could be done, for example, by means of speed-controlling technologies. They could help to align the speeds at which people drive with the speeds at which robotic cars drive. Or it could be done—to use another example we also mentioned above—with the help of things such as alcohol-locks.[21] Whatever means might be suggested, what sorts of ethical issues might be brought to bear on the evaluation of this general strategy for achieving better human-robot coordination within mixed traffic?

This is perhaps the strategy most likely to generate heated debate if it is taken seriously and it receives the attention we think it deserves. On the critical side, obvious objections to be anticipated are likely to concern worries about potential infringements upon drivers' freedom and, at the extreme, perhaps even worries about infringements upon drivers' human dignity. On the other side, considerations such as the duty of care that we typically associate with traffic and related duties of responsible risk-management also need to be taken very seriously.

In other contexts, when discussions about mandating things such as speed-regulation technologies spring up—either for all drivers or some sub-class, such as truck-drivers—one of the issues that tends to be raised is the worry that this takes away the driver's freedom to choose how he or she wants to operate his or her vehicle. For example, one Canadian truck-driver who had been ordered to use a speed-limiter in his truck took the matter to court. There, he argued that his fundamental freedoms would be compromised if he couldn't himself be in charge of deciding how fast or slow he was going when driving his truck.[22] It is to be expected that similar objections will be raised if a serious discussion arises about the idea of trying to conform human driving to robotic driving by requiring human drivers to use technologies such as speed-limiters in their conventional cars.

The idea of trying to conform human traffic-behaviors to robotic traffic-behaviors might perhaps also, as we suggested above, strike some as an assault on human dignity. This would take the choice of whether or not to follow rules such as speed-limits (and thereby better coordinate one's driving with robotic driving) out of the hands of the human driver. The human driver could not self-apply the law. And being afforded the opportunity to self-apply laws—as opposed to being made to follow laws—has sometimes been said to be contrary to human dignity in general. For example, legal theorists Henry Hart and Albert Sachs see the self-application of law as a crucial part of human dignity (Hart and Sachs 1994). Legal philosopher Jeremy Waldron also joins them in associating this idea with human dignity in his recent book on dignity based on his Tanner Lectures on the subject (Waldron 2012, p. 55).

It is to be expected that these kinds of worries will be raised. But upon closer inspection, would it really offend against values such as freedom and human dignity to suggest that we try to achieve better human-robot coordination in mixed traffic by seeking technological means for conforming at least certain non-ideal aspects of human driving to robotic driving-styles?[23] Also, what sorts of countervailing arguments might be presented on the opposite side of the issue, that would qualify as positive arguments in favor of this general idea?

Here, we wish to make three main points. Firstly, from a legal and moral point of view, we do not currently enjoy neither a legal nor a moral freedom to speed or to otherwise drive in ways that expose people to greatly increased risks (Royakkers and Van Est 2016). We have a legal freedom to

[20] Recall, for instance, the above example of the human deciding whether or not to cross the double line that was mentioned in footnote 8 above. That was an example of the sort of situation that is not too demanding for the human driver, and where human input can improve the car's handling of the situation at hand.

[21] We are not here interested in investigating—or defending—any specific technological means for making human driving more like robotic driving; we're more interested in the general idea and the question of what sorts of ethical issues are relevant in relation to this sort of idea.

[22] At first, the court ruled in favor of the truck-driver. However, another Canadian court later overturned that decision, ruling that requiring the truck-driver to use a speed-limiter did not offend against his fundamental freedoms. See, e.g., http://www.todaystrucking.com/court-upholds-ontario-truck-speed-limiter-law.

[23] An anonymous reviewer suggested that the freedom-worry could be solved by offering human drivers a voluntary contract, whereby they would agree to using speed-limiters if they want to use manually driven cars. The problem here is that a substantial sub-set of drivers might reject this contract, just like the above-mentioned truck-driver who did not want to use a speed-limiter. This directly re-opens the question of whether such drivers have a justified claim to being afforded a freedom to speed.

do something if the law permits it, and a moral freedom to do something if morality permits it. Driving in ways that create great risks is neither permitted by law nor by good morals. So it could be argued that if we try to make people drive more like robots by putting speed-regulators in their cars, and thereby achieve better human-robot coordination within mixed traffic, then we do not take away any legal and moral freedom that people can currently lay claim to. What we would block would rather by a purely "physical" freedom to drive in certain dangerous ways that are neither legally nor morally sanctioned and that make it much harder to create good human-robot coordination within mixed traffic.[24] To clarify: the point is not that being free is the same as doing what is legally and morally permitted. The point is rather that there is a significant distinction between freedoms that people ought to be afforded and freedoms that they ought not to be afforded. And from a legal and moral point of view, people are not—and ought not to be—afforded freedoms to drive in ways that greatly increase the risks involved in traffic.[25]

Secondly, it may indeed be that in general, one important part of human dignity has to do with being afforded the freedom to self-apply laws. But it is not so clear that this ideal requires that people always be given a choice whether or not to self-apply all laws, across all different domains of human activity, whatever the costs (cf. Smids forthcoming; Yeung 2011). In some domains, other values may be more salient and more important for the purposes and goals specific to those domains. Traffic, for example, which is the domain we're currently discussing, is not obviously a domain where the most important value is to be afforded the opportunity to self-apply traffic-regulations.

Values much more salient in this domain include things such as safety, and mutual respect and concern, or more mundane things such as user-comfort and overall traffic-efficiency. It is not so clear that being afforded the choice of deciding whether or not to follow traffic-rules intended to save lives is a key value that stands out as being what we typically most value within this domain of human activity. Furthermore, there are still a lot of traffic rules to follow, giving ample room to self-apply the law. Moreover, being kept safe by laws and norms that seek to protect us and our life and limb can surely also be seen—and is surely often seen—as an important part of what it means to enjoy a dignified status in human society (Cf. Rosen 2012). So upon closer inspection, seeking means for making people drive more like robots may not be such a great offense to human dignity after all, even if the basic idea might sound a little strange at first.

Thirdly, there is another very important thing about the choices drivers face that should also be kept in mind, if it is indeed true that highly automated driving would be a very safe form of driving.[26] And that is that the introduction of this supposedly much safer alternative can plausibly be seen as changing the relative moral status of some of the choices drivers face.

If highly automated driving is indeed safer than non-automated conventional driving, the introduction of automated driving thereby constitutes the introduction of a safer alternative within the context of mixed traffic. So if a driver does not go for this safer option, this should create some moral pressure to take extra safety-precautions when using the older, less safe option even as a new, safer option is introduced.[27] As we see things, then, it can plausibly be claimed that with the introduction of the safer option (viz. switching to automated driving), a new moral imperative is created within this domain. Namely, to either switch to automated driving (the safer option) or to take or accept added safety-precautions when opting for conventional driving (the less safe option). If automated cars are established to be a significantly safer alternative, it would be irresponsible to simply carry on as if nothing had changed and there were no new options on the horizon.[28]

[24] Moreover, by making mixed traffic safer and thereby making the option of using a car available to, and more eligible for, a wider range of people (e.g. the elderly and the severely disabled), we could be seen as extending the freedom people enjoy in this domain (Cf. Bradshaw-Martin and Easton 2014). Let more people become able to exercise the option of using a car (either an automated car or a conventional car); and let this become a safe and reliable option for all. If these two conditions are fulfilled, then the option to use a car become more like a basic freedom. This requires that people who use conventional cars be willing to accept measures to create a more inclusive type of traffic, which can include accepting measures that help to create better human-robot coordination within mixed traffic. Cf. Pettit 2012 on "co-exercisability" and "co-satisfiability" as two of the requirements for counting something (e.g. an option all might be afforded within a society) as a basic liberty. See especially pp. 93–97.

[25] We thank an anonymous reviewer for prodding us to clarify this point.

[26] Legal theorists Marchant and Lindor (2012) argue that automated cars will not be legally viable unless they can be shown to be safer, if not much safer, than conventional cars. Hence, they argue, any discussion of traffic-scenarios involving automated cars likely to occur can treat automated cars as safer than conventional cars. Our argument in these last paragraphs of this section rests on the assumption that Marchant and Lindor are right about this. In other words, for the sake of this third argument, we here assume that automated cars will represent a safer alternative as compared to conventional cars.

[27] Cf. Scanlon (1998) and Lehnman (2008) on the general idea in risk ethics that one thing that can make imposing risks on others acceptable is that we take due precautions.

[28] Moreover, if the introduction of automated vehicles can extend the option of using a car independently to a greater number of people (e.g. the elderly and severely disabled people), then this seemingly also adds a further duty drivers for drivers of conventional cars: namely, a duty help to enable these new car-users to participate in mixed traffic in a safe way (Cf. Bradshaw-Martin and Easton 2014). That duty can be discharged by accepting certain means for making

Concluding summary

The widespread introduction of automated vehicles will create mixed traffic, involving both automated cars and conventional cars, and the automated cars are likely to feature different levels and types of automation. Automated cars are programmed to drive in optimizing ways, and are strict rule-followers; humans drive in a satisficing way, and are flexible rule-benders. Therefore, mixed traffic will create various human-robot coordination-issues, which can create dangerous situations and lead to crashes and accidents.

One suggestion about how to achieve greater human-robot coordination is to try to make robotic driving more like human driving. Another suggestion is to seek fruitful ways of involving the human in the operation of highly automated vehicles. A third is to seek means, which might be technological means, for making human driving more like robotic driving. All general solution-strategies, we have argued, deserve to be taken seriously and investigated further. We should not only focus on the first strategy.

Responsible human-robot coordination within mixed traffic needs to confront the various different ethical issues that these different solution-strategies give rise to. For example, if we want to conform robotic driving to human driving in a responsible way, we should try to avoid conforming robotic driving to morally problematic and illegal aspects of how many people drive. If and when we create new responsibilities for human drivers, we should not create responsibilities most humans are unlikely to be able to handle. And when it comes to conforming human driving to robotic driving, we need to be mindful of key ethical values such as freedom and human dignity. However, we must also be open to the positive ethical reasons there can be to try to conform human driving to robotic driving. If automated cars represent a much safer alternative, as it is widely hoped that they will do, then this seems to create a new duty for those who use conventional cars. And that duty is to either switch to automated cars (=the safer alternative) or to use extra precautions when using the otherwise less safe alternative.

The widespread introduction of automated vehicles—especially highly or fully automated vehicles—amounts to the introduction of a large number of robotic agents into a domain of human activity where the stakes are very high whenever there are accidents. This is an exciting development, but also one that creates new responsibilities and ethical challenges. In this paper, we have argued that one distinct and very important challenge is responsible human-robot coordination within this risky area of human life.

Footnote 28 (continued)

human driving more like robotic driving in certain ways, so as to achieve better human-robot coordination within mixed traffic.

Acknowledgements We thank the journal's anonymous reviewers and audiences in Eindhoven and Utrecht for their helpful feedback on this material.

References

Arkin, R. (2010). The case for ethical autonomy in unmanned systems. *Journal of Military Ethics, 9*(4), 332–341.

Bostrom, N., & Ord, T. (2006). The reversal test: Eliminating status quo bias in applied ethics. *Ethics 116*, 656–679.

Bradshaw, J., Hoffman, R., Johnson, M., & Woods, D. (2013). The seven deadly myths of "autonomous systems". *IEEE Intelligent Systems, 28*, 54–61.

Bradshaw-Martin, H., & Easton, C. (2014). Autonomous or "driverless" cars and disability: A legal and ethical analysis. *European Journal of Current Legal Issues, 20*(3). Retrieved from http://webjcli.org/article/view/344.

Färber, B. (2016). Communication and communication problems between autonomous vehicles and human drivers. In M. Maurer, J. C. Gerdes, B. Lenz, & H. Winner (Eds.), *Autonomous Driving* (pp. 125–144). Berlin: Springer.

Floridi, L., & Sanders, J. W. (2004). On the morality of artificial agents. *Minds and Machines, 14*(3), 349–379. https://doi.org/10.1023/B:MIND.0000035461.63578.9d.

Gerdes, J. C., & Thornton, S. M. (2016). Implementable ethics for autonomous vehicles. In M. Maurer, J. C. Gerdes, B. Lenz, & H. Winner (Eds.). *Autonomous driving* (pp. 87–102). Berlin: Springer

Gogoll, J., & Müller, J. (2016). Autonomous cars: In favor of a mandated ethics setting. *Science and Engineering Ethics*. https://doi.org/10.1007/s11948-016-9806-x.

Goodall, N. J. (2014a). Ethical decision making during automated vehicle crashes. *Transportation Research Record: Journal of the Transportation Research Board, 2424*, 58–65.

Goodall, N. J. (2014b). Machine ethics and automated vehicles. In G. Meyer & S. Beiker (Eds.) *Road vehicle automation* (pp. 93–102). Berlin: Springer.

Grill, K., & Nihlén Fahlquist, J. (2012). Responsibility, paternalism and alcohol interlocks. *Public Health Ethics, 5*(2), 116–127.

Gurney, J. K. (2016). Crashing into the unknown: An examination of crash-optimization algorithms through the two lanes of ethics and law. *Albany Law Review, 79*(1), 183–267.

Hart, H., & Sachs, A. (1994). *The legal process*. Eagan, MN: Foundation Press.

Hevelke, A., & Nida-Rümelin, J. (2014). Responsibility for crashes of autonomous vehicles: An ethical analysis. *Science and Engineering Ethics, 21*, 1–12.

Husak, D. (2004). Vehicles and crashes: Why is this issue overlooked? *Social Theory and Practice, 30*(3), 351–370.

Ingrassia, P. (2014). Look, no hands! Test driving a Google car. Reuters. Retrieved from http://www.reuters.com/article/us-google-driverless-idUSKBN0GH02P20140817.

Kuflik, A. (1999). Computers in control: Rational transfer of authority or irresponsible abdication of autonomy? *Ethics and Information Technology, 1*(3), 173–184. https://doi.org/10.1023/A:1010087500508.

Lai, F., Carsten, O., & Tate, F. (2012). How much benefit does intelligent speed adaptation deliver: An analysis of its potential contribution to safety and environment. *Accident Analysis & Prevention, 48*, 63–72. https://doi.org/10.1016/j.aap.2011.04.011.

Lenman, J. (2008). Contractualism and risk imposition. *Politics, Philosophy & Economics, 7*(1), 99–122. https://doi.org/10.1177/1470594X07085153.

Lin, P. (2015). Why Ethics Matters for Autonomous Cars. In M. Maurer, J. C. Gerdes, B. Lenz & H. Winner (Eds.), *Autonomes Fahren: Technische, rechtliche und gesellschaftliche Aspekte* (pp. 69–85). Berlin: Springer.

List, C., & Pettit, P. (2011). *Group agency: The possibility, design, and status of corporate agents*. Oxford: Oxford University Press.

Marchant, G., & Lindor, R. (2012). The coming collision between autonomous cars and the liability system. *Santa Clara Legal Review, 52*(4), 1321–1340.

Mindell, D. (2015). *Our robots, ourselves: Robotics and the myths of autonomy*, New York: Viking.

Naughton, K. (2015). Humans Are slamming into driverless cars and exposing a key flaw: Bloomberg business. Retrieved February 23, 2016, from http://www.bloomberg.com/news/articles/2015-12-18/humans-are-slamming-into-driverless-cars-and-exposing-a-key-flaw?utm_content=buffer16029&utm_medium=social&utm_source=facebook.com&utm_campaign=buffer.

Nyholm, S., & Smids, J. (2016). The ethics of accident-algorithms for self-driving cars: An applied trolley problem? *Ethical Theory and Moral Practice 19*(5), 1275–1289.

Nyholm, S. (forthcoming). Attributing agency to automated systems: Reflections on responsible human–robot collaborations and responsibility-loci, *Science and Engineering Ethics*. https://doi.org/10.1007/s11948-017-9943-x.

Pettit, P. (2012). *On the people's terms*. Cambridge, MA: Cambridge University Press.

Rosen, M. (2012). *Dignity*. Cambridge, MA: Harvard University Press.

Royakkers, L., & Van Est, R. (2016). *Just ordinary robots: Automation from love to war*. Boca Raton, FL: CRC Press.

Scanlon, T. (1998). *What we owe to each other*. Cambridge, MA: Belknap Press of Harvard University Press.

Schlosser, M. (2015). *Agency, Stanford Encyclopedia of philosophy (Fall 2015 Edition)*. https://plato.stanford.edu/archives/fall2015/entries/agency/.

Schoettle, B., & Sivak, M. (2015a). *A preliminary analysis of real-world crashes involving self-driving vehicles (No. UMTRI-2015-34)*. Ann Arbor, MI: The University of Michigan Transportation Research Institute.

Schoettle, B., & Sivak, M. (2015b). Road safety with self-driving vehicles: General limitations and road sharing with conventional vehicles. Retrieved from http://deepblue.lib.umich.edu/handle/2027.42/111735.

Smids, J. (forthcoming). The moral case for intelligent speed adaptation. *Journal of Applied Philosophy*. https://doi.org/10.1111/japp.12168.

Tesla (2016). A tragic loss. https://www.tesla.com/blog/tragic-loss.

Urmson, C. (2016). Report on traffic accident involving an autonomous vehicle, California, D. M. V. https://www.dmv.ca.gov/portal/wcm/connect/3946fbb8-e04e-4d52-8f80-b33948df34b2/Google+Auto+LLC+02.14.16.pdf?MOD=AJPERES.

van Loon, R. J., & Martens, M. H. (2015). Automated driving and its effect on the safety ecosystem: How do compatibility issues affect the transition period? *Procedia Manufacturing, 3*, 3280–3285. https://doi.org/10.1016/j.promfg.2015.07.401.

Wachenfeld, W., Winner, H., Gerdes, J. C., Lenz, B., Maurer, M., Beiker, S., Fraedrich, E., & Winkle, T. (2016). Use cases for autonomous driving. In M. Maurer, J. C. Gerdes, B. Lenz, & H. Winner (Eds.), *Autonomous driving* (pp. 9–37). Berlin: Springer.

Wagter, H. (2016). "Naughty Software", presentation at Ethics: Responsible driving automation, at Connekt, Delft.

Waldron, J. (2012). *Dignity, rank, and rights*. Oxford: Oxford University Press.

Wolf, I. (2016). The interaction between humans and autonomous agents. In M. Maurer, J. C. Gerdes, B. Lenz, & H. Winner (Eds.). *Autonomous driving* (pp. 103–124). Berlin: Springer.

Yang, Q., Gao, Y., & Li, Y. (2016). Suppose future traffic accidents based on development of self-driving vehicles, In S. Long & B. S. Dhillon (Eds.), *Man-machine-environment system engineering, lecture notes in electrical engineering*. New York: Springer

Yeung, K. (2011). Can we employ design-based regulation while avoiding brave new world? *Law Innovation and Technology, 3*(1), 1–29. https://doi.org/10.5235/175799611796399812.