CrossMark

On the problem of making autonomous vehicles conform to traffic law

Henry Prakken[1,2]

Abstract Autonomous vehicles are one of the most spectacular recent developments of Artificial Intelligence. Among the problems that still need to be solved before they can fully autonomously participate in traffic is the one of making their behaviour conform to the traffic laws. This paper discusses this problem by way of a case study of Dutch traffic law. First it is discussed to what extent Dutch traffic law exhibits features that are traditionally said to pose challenges for AI & Law models, such as exceptions, rule conflicts, open texture and vagueness, rule change, and the need for commonsense knowledge. Then three approaches to the design of law-conforming AV are evaluated in light of the challenges posed by Dutch traffic law, which includes an assessment of the usefulness of AI & Law models of non-monotonic reasoning, argumentation and case-based reasoning.

Keywords Autonomous vehicles · Traffic law · Legal reasoning

1 Introduction

Autonomous vehicles are one of the most spectacular recent developments of Artificial Intelligence. While currently allowed technology is limited to features such as adaptive cruise control, parking assistance with automated steering and lane keeping assistance, fully autonomous vehicles, which can drive to any location where it is legal to drive and make their own decisions without human intervention,

✉ Henry Prakken
  h.prakken@uu.nl

1   Department of Information and Computing Sciences, Utrecht University, Utrecht,
    The Netherlands

2   Faculty of Law, University of Groningen, Groningen, The Netherlands

or near-fully autonomous vehicles, which can control the vehicle in all but a few environments such as severe weather and which do not require human driver attention, may well be able to take part in ordinary traffic within the next decades (Anderson et al. 2016). Among the problems that need to be solved is the one of making the behaviour of autonomous vehicles (AV) conform to the traffic laws. Solutions to this problem may well profit from computational models of legal reasoning but so far the field of AI & Law has hardly addressed this issue. The present paper aims to put this topic on the AI & Law research agenda by way of a case study of Dutch traffic law and its implications for the design of fully autonomous self-driving cars. In particular, the challenges will be discussed that Dutch traffic law poses for AV and how existing AI & Law techniques are relevant for dealing with these challenges. While several features of Dutch traffic law are specific to the Dutch jurisdiction, its main features are similar to traffic laws in other western jurisdictions, especially European ones, so the present case study is more generally relevant than just for Dutch law. One of the few earlier studies of the problem of making AV conform to traffic law is Leenes and Lucivero (2014), who primarily take a legal point of view but hint at the possibility of programming AV according to AI & Law reasoning models. One purpose of the present paper is to further investigate this possibility. In the literature on AV design there have to the best of my knowledge so far not been any systematic studies of the problem of making AV conform to traffic law through its design (as also observed by Leenes and Lucivero 2014). For example, Thrun (2010) does not even mention the problem while yet he discusses the DARPA Urban challenge, in which the AV had to obey California traffic rules. Instead, most research concerns issues like liability for accidents, regulation of AV design and production, regulation of (experimental or regular) use of AV, data protection and privacy issues, and whether the law should be changed to accommodate AVs. See e.g. Anderson et al. (2016), Leenes and Lucivero (2014), Palmerini et al. (2014) and Vellinga (2017) for discussions of these issues and for further references. Therefore, the present study, while still a conceptual one, fills an important gap in the literature.

The problem of making AV conform to traffic law is a special case of the more general problem of making intelligent autonomous systems conform to the relevant laws. The present study of the problem for autonomous vehicles may therefore also have relevance for the study of the more general problem. Computer systems are increasingly being employed in practice with some degree of autonomy. Their behaviour is not fully specified by the programmer but is the result of the implementation of more general cognitive or physical abilities. Such artificially intelligent systems can do things which, when done by humans, are regulated by law. Apart from self-driving cars, some examples are care robots that help sick or elderly people and whose actions can damage property or the health of the person (spilling coffee over an iPad, failing to administer medication on time), intelligent fridges that can order food or drinks when the supplies run out and thus have to conform to contract law, and financial trading programs that have to conform to financial regulations.

When such autonomous systems are being used, legal rules cannot any more be regarded as regulating human behaviour, since it is not the humans but the machines

who act. This raises the problem of how the autonomous systems can be designed in such a way that their behaviour complies with the law. Note that this question needs to be asked irrespectively of the legal question whether machines can be assigned responsibility in a legal sense. Even if a human remains legally responsible or liable for the actions of the machine, the human faces the problem of ensuring that the machine behaves in such a way that the responsible human complies with the law. Currently, this kind of problem is mainly studied under the heading of 'machine ethics' (Anderson and Anderson 2011). While this may be the appropriate field for studying the related problem of making intelligent autonomous systems behave ethically responsibly, the problem of making them conform to the law arguably belongs to AI & Law.

Accordingly, I will first in general terms compare the new task of designing law-conforming intelligent autonomous systems to the legal tasks so far modelled in AI & Law research and discuss the implications of any differences for the design of such systems. Then I will by way of a case study zoom in on the problem how autonomous vehicles can be made to conform to Dutch traffic law. I will first give an overview of Dutch traffic law and discuss to what extent it exhibits features that are traditionally said to pose challenges for computational models of legal reasoning, such as exceptions, rule conflicts, open texture and vagueness, rule change, and the need for commonsense knowledge. Then I will evaluate three different approaches to the design of law-conforming AV in light of the challenges posed by Dutch traffic law. Among other things, this involves an assessment of the usefulness of AI & Law models of nonmonotonic reasoning, argumentation and case-based reasoning.

As for the nature of this paper it should be noted that the aim is not to present concrete technical advances but rather to clarify and organise a number of issues concerning the problem of making AV conform to traffic law. Several parts of the paper, such as the legal knowledge-representation issues and methods, may be obvious to AI & Law specialists. However, I believe that it is important that these issues and methods also become known to specialists in AV design, who generally have little or no knowledge of the field of AI & Law. Moreover, the general clarification and organisation of the relevant issues may be of help to AI & Law researchers who want to start research on making AV in particular, and autonomous systems in general, conform to the relevant regulations.

2 The classic AI & Law problems versus the new challenge

The task of making intelligent autonomous systems conform to the law has some similarities and differences with more traditional tasks modelled in AI & Law. Just as in, for instance, legal decision or argumentation support, the task is to apply norms to facts in order to legally classify behaviour. Therefore, the task of making autonomous systems conform to the law faces some of the same challenges as any task in which behaviour has to be legally classified, in particular the possibility of rule conflicts, ambiguities in the formulation of legal rules, the open-textured and

vague nature of many legal concepts and the possibility of unforeseen exceptions on the basis of purpose or principle.

However, there are also differences. First, while AI & Law research has traditionally focused on support tools for humans carrying out legal tasks, with autonomous systems this is different: they do not support humans in their legal tasks (although they may support humans in other tasks) but they have to decide about the legal status of their own actions. In many cases it will be impossible for humans to check or override the system's decision.

Moreover, while tasks supported by traditional AI & Law tools often concern the application of the law to past cases, to determine the legal status of some past behaviour or some existing state of affairs, autonomous systems have to determine the legal status of their future actions. Among other things, this means that in contrast to traditional legal settings, autonomous systems do not face evidential problems in the legal sense. Even when traditional AI & Law supports legal tasks with an eye to the future, such as deciding on social-benefits or on permit applications, drafting regulations or contracts or designing tax schemes, there are differences with autonomous systems. While traditionally supported future-oriented tasks concern behaviour in the non-immediate future and often contain classes of actions (as with contract drafting or regulation design), autonomous systems have to 'run-time' consider individual actions in the immediate future.

Another difference is that while the legal tasks traditionally modelled in AI & Law require explanation and justification of decisions, with autonomous systems there is less need for this, since the primary problem is to generate legally acceptable behaviour. Therefore, the black-box nature of data-mining or machine-learning techniques may, unlike with traditional legal tasks, be less of a problem for the task of making autonomous systems conform to the law. Of course, when an autonomous system does something legally wrong, its behaviour might have to be explained in a court case. However, this does not require that the system itself can do that; it may suffice to have a log file recording the system's internal actions.

Next, while much AI & Law research studies legal tasks in an adversarial setting (primarily a legal proceeding), with the task of letting autonomous systems conform to the law there are no adversaries: all that counts is to let the system do what it has to do within the bounds of the law. Thus there will be less need for argumentation than usual in AI & Law applications.

Yet another difference is that one may expect that the bulk of the cases encountered by an autonomous system will from a legal point of view be standard, mundane cases. For example, autonomous cars will not have to determine the legal responsibility for car accidents but will have to decide about driving from A to B in a way that conforms to the traffic regulations. While processing legislation in public administration also usually concerns standard cases, in the court room this is different.

Finally, many of the tasks traditionally modelled in AI & Law are strictly legal while autonomous systems have to balance legal considerations against other considerations. Autonomous systems are not designed to obey the law but for other purposes, such as driving from A to B. Sometimes behaviour fulfilling this purpose is from a legal point of view illegal but still socially acceptable; for example,

slightly speeding in a queue of cars that all drive a few miles above the maximum speed, or passing a vehicle that is standing still where changing lanes is forbidden but there is no approaching traffic. This means that the behaviour of autonomous systems should not be seen as rule-governed but as rule-guided. Legal rules are just one factor influencing socially optimal or permissible behaviour. Other factors are e.g. social conventions, individual or social goals or simply common sense. And sometimes these other factors override the legal factors. Having said so, even rule-guided models of autonomous systems will have to specify what the law requires, and this is the problem to be discussed in the rest of this paper.

3 Dutch traffic law

In this section Dutch traffic law will be summarised as far as it results in cues for traffic behaviour of autonomous vehicles. First the Dutch traffic regulations will be reviewed, after which civil liability rules will be discussed as far as they give rise to cues for traffic behaviour.

3.1 Traffic regulations

A general feature that the Dutch traffic regulations share with any set of traffic regulations is that they apply to a relatively simple, closed and predictable world, at least compared to may other legal domains. This is one reason why the Dutch traffic regulations are rather precise and concrete (but with some exceptions to be discussed below). Another reason is the need for humans to safely and efficiently coordinate their actions in traffic, which requires clear and precise rules. For this reason "reasoning from first principles" is in traffic undesirable (MacCormick 1998). Moreover, in contrast to domains in which AI & Law has been applied successfully, like social security and tax law, the amount of Dutch traffic legislation is rather modest and is not subject to frequent change.

The main Dutch traffic regulations are the Road Traffic Act 1994 (Wegenverkeerswet 1994, WVW) and the Traffic Rules and Signs Regulations 1990 (Reglement Verkeersregels en Verkeerstekens 1990, RVV). The WVW hierarchically precedes the RVV. The Dutch traffic regulations are (like presumably in all jurisdictions) designed to promote two purposes: safe and efficient traffic. Needless to say that these purposes can in concrete situations conflict. These purposes are codified in Article 5 WVV, which states that 'It is forbidden to behave in such a way that danger on the road is or may be caused or that the road traffic is or may be impeded'. Only three other articles from the WVW arguably yield cues for traffic behaviour.

Article 12 para 1 WVW states that road users are required to follow instructions given by authorised officials like police officers. Article 6 WVW is the only rule in the WVW or RVV that refers to a driver's mental state. It says that it is forbidden in traffic to behave in such a way that a traffic accident happens due to one's fault by which serious bodily harm or death is caused to someone else. The main purpose of this article is not to define a separate cue for traffic behaviour but to allow for more

severe sanctions when violation of other traffic rules causes serious bodily harm. Finally, article 185 WVW is a special rule concerning civil liability for damages caused by traffic accidents, broadening the conditions for liability according to general Dutch tort law. It will be discussed further in Sect. 3.2.

The final relevant part of the WVW is a list of definitions. For present purposes the definitions of various kinds of vehicles and traffic agents and of 'roads' are the most relevant. They are stated in concrete terms, so they are quite precise.

An obvious potential obstacle for attempts to make AV conform to traffic law (also mentioned by Leenes and Lucivero 2014) is the occurrence of vague and open-textured terms in traffic regulations. Articles 5, 6 and 185 WVW are the only genuinely open norms of Dutch traffic law, with clearly vague and open-textured concepts like causing danger on the road, impeding traffic, fault, serious bodily harm, and causation of harm or damage. As just noted, Article 6 WVW is for present purposes irrelevant. Moreover, the vague notion of causing danger in Article 5 WVW is to a large extent made precise in the RVV, since its rules can be regarded as an attempt to enforce safe traffic behaviour. Article 5 WVW thus has two roles: as a fall-back option in case dangerous behaviour is not forbidden by specific RVV rules, and as a general implicit exception to specific RVV rules in case otherwise permitted or obliged behaviour would cause danger to or impediment of traffic.

An interesting case in this respect is that of a self-driving Google car which was stopped by the California police for driving too slowly. Google had for safety reasons set the car's maximum speed for roads in a 35mph zone at 25mph and one of its cars was causing a big queue of traffic while driving 24mph.[1] This behaviour was held by the police to violate the following rule from the Californian traffic regulations:

> No person shall drive upon a highway at such a slow speed as to impede or block the normal and reasonable movement of traffic, unless the reduced speed is necessary for safe operation, because of a grade, or in compliance with law.

Under Dutch law this behaviour would be a violation of Article 5 WVW. Incidentally, this would arguably not be a case of Article 5 WVW making an exception to the speed limit rules, since the speed limit rules are arguably not permissions to drive any with speed up to the speed limit but prohibitions to drive faster than the speed limit.

Because of the purposes of promoting safe and efficient traffic, the rules in the RVV are generally quite concrete, so the problem of open texture and vagueness is not a serious as in other legal domains. Moreover, the special values of promoting safe and efficient traffic also underlie an important difference between criminal traffic law and general criminal law. Unlike in general criminal law, which consists of prohibitions, traffic law mostly specifies obligations, that is, it specifies how traffic participants should behave in particular situations. (Exceptions to obligations are specified as explicit permissions. For example, Article 3 para 1 WVW says 'Drivers shall keep to the right as much as possible' while Article 3 para 2 WVW says 'Cyclists may ride in pairs'.)

---

[1] http://www.bbc.com/news/technology-34808105, accessed 21 December 2016.

Nevertheless, the RVV still contains several vague or open-textured terms. Moreover, the term 'driver' raises a legal problem if used in norms, such as in Article 3(1) RVV. With AV the legal problem is who counts as the driver (Vellinga 2017): the human sitting behind the steering wheel, the AV, or perhaps the manufacturer of the AV? However, this is only a problem if observed traffic behaviour has to be legally classified. For present purposes, where the focus is on cues for traffic behaviour, all that counts is how to program the AV such that it will keep right as much as possible.

I now give an overview of the open-textured or vague terms in the RVV.

First of all, the term keeping right *as much as possible* in Article 3 RVV is open-textured. It should arguably not be read in a physical sense but in the sense of what is reasonably right as much as possible given the traffic context. This reading makes the term open-textured.

Article 17(1a) RVV: drivers who want to make a turn are permitted to get in lane by keeping left or right *as much as possible*.

Article 28 RVV: sound or light signals are only allowed to avert a *threat of danger*.

Article 43(3) WVW: on motorways, use of the shoulder is only allowed *in case of emergency*.

Article 57 RVV is a prohibition for drivers to make *unnecessary noise*.

Finally, various rules on carrying particular types of light contain the term *serious inhibition of visibility*.

Other concepts are clear for humans but require some recognition or judgement by AV, on the recognition of objects (such as tunnels, bicycle lanes, crossings and roundabouts) or of spatial relations (on or just before, at a short distance, blocking, making free), or of behaviours of humans (indicating to turn, moving with difficulty) or on observability (that he can clearly see, serious inhibition of visibility). Here are some examples.

Article 11(2) RVV: overtaking drivers *who have indicated that they want to turn left*.

Article 13 RVV is an exception to Article 3(1) RVV for *traffic jams* on motor ways.

Article 14 RVV: drivers are forbidden to *block* crossings.

Article 15a (1): road users are permitted to enter a railway crossing if *they can directly continue and make the crossing fully free*.

Article 19 RVV states that 'A driver must at all times be able to bring his vehicle to a standstill within the distance *that he can see to be clear*'.

Article 23 RVV: drivers are not allowed to stand still at a crossing, railway crossing, bicycle lane, *tunnel*, ...

Article 24: a prohibition to park at a place *meant for immediate loading or unloading of goods*.

Article 32(2c): carrying long distance light is forbidden when following another vehicle *at a short distance*.

Article 47 RVV: a permission not to keep right on *or just before* roundabouts.

Article 49: RVV: an obligation for drivers to give way to persons *who move with difficulty*.

Article 68(1b) RVV: a yellow traffic light means stop unless the distance to the traffic light is so short that stopping is *not reasonably* possible.

Article 75: A yellow flashing light means, dangerous point: *carefulness* required.

Article 83 requires road users to stop when a *stop sign (red lamp or red letters 'STOP')* is shown from a police vehicle.

A general structural feature of the WVV and RVV is that they allow for rule conflicts and for their resolution with general conflict resolution principles (some of them are also noted by Leenes and Lucivero (2014)). Article 63 RVV says that traffic signs have priority over traffic rules, Article 63a says that temporary traffic signs have priority over permanent traffic signs, Article 63b says that if various rules concerning the maximum speed conflict, the lowest maximum speed is the valid one, and Article 64 says that as regards giving way, traffic lights have priority over traffic signs. Article 84 says that instructions of authorised officials have priority over traffic signs and rules. As noted above, Article 5 WVW yields an implicit exception to all obligations and permissions in the RVV in case the obliged or permitted behaviour creates danger for or impediment of traffic. This conflict is resolved by *Lex Superior Derogat Legi Inferiori*, since the WVW is a statute and the RVV a lower regulation. Moreover, the RVV was drafted according to an explicitly adopted convention to separate general rules from exceptions. So the principle of *Lex Specialis Derogat Legi Generali* applies straightforwardly. An example discussed by Leenes and Lucivero (2014) is Article 15 para 1 RVV, which says that 'At road junctions, drivers must give way to traffic approaching from the right', while Article 15 para 2 RVV says that 'a. drivers on unpaved roads must give way to drivers on paved roads; b. all drivers must give way to tram drivers'. An example of conflicting rules with a not so clear rule-exception structure is Article 14 RVV: 'drivers must not block road junctions' versus Article 82 para 1 RVV: 'Road users are required to follow instructions given verbally or by means of gestures by: authorised officials who are identified as such ...' (which is a refinement of Article 12 WVW). These rules conflict when a police officer orders a car to stop at a road junction. This conflict is resolved by Article 84 RVV that signals of authorised officials have priority over traffic signals and rules. Finally, an example of a rule conflict not resolved by statutory principles is the conflict between the general priority rules and Article 54 WVW, which says that motor vehicle drivers carrying out special maneuverings, such as starting to drive, driving backwards, driving onto the road from an entrance way, turning, changing lanes, changing between the main lane and a merging lane, should give way to all other traffic. Here it was the Dutch Supreme Court that decided that Article 54 WVW sets aside all other traffic priority rules (HR 17 September 2002, NJ 2002, 549).

Finally, the RVV contains various applicability rules concerning road signs.

3.2 Civil liability rules

The main function of civil liability rules is to determine liability once something has happened. Currently, there is much discussion among legal scholars on liability issues concerning AV when an AV causes an accident (Anderson et al. 2016; Vellinga 2017). However, for present purposes traffic liability rules are relevant in a

different way: they can pragmatically entail additional cues for driving behaviour in cases where no specific rule from the RVV or WVW is violated, since rational drivers will want to avoid damages. Below, Dutch traffic liability law will be discussed from the latter perspective.

Article 185 WVW is a special rule concerning civil liability of owners of motor vehicles for damages caused to drivers of non-motorised vehicles or pedestrians. Its underlying principle is that motor vehicles are inherently dangerous and motor vehicle drivers deliberately accept the resulting risks, so that more vulnerable road users should be protected. Accordingly, 185 WVW says that the owner of a motor vehicle involved in an accident is civilly liable for damages to persons or goods not transported by a motor vehicle, unless he can prove that the accident was due to force majeure. Note that liability according to this article does not require negligence or violation of a traffic rule.

Existence of force majeure is seldomly accepted by courts (Rutten and Oskam 2016). For example, according to the Dutch Supreme Court, a technical defect of the vehicle does not count as force majeure. The same holds in general for unusual weather. Finally, mistakes of other road users count only in some very special cases as force majeure (although they can be a reason to distribute the amount of damages over the plaintiff and defendant). To start with, the driver's behaviour should not be blamable in any way. Next, mistakes of road users over 14 years old only result in force majeure for the motor vehicle driver if these mistakes were so improbable that it would be unreasonable to require of the motor vehicle driver to take this possibility into account. According to Dutch case law, not improbable in this sense are, for example, cases of failure to give way by the non-driver, cycling without light and suddenly crossing a road. For mistakes made by road users less than 14 years old the criterion is even more strict; they only count as force majeure if the child acted with recklessness bordering to intent. Dutch commentators give as a possible example of force majeure that older children deliberately play the dangerous game of crossing the street just before approaching traffic (Rutten and Oskam 2016, p. 323). On the other hand, if a younger child suddenly crosses the street because of carelessness, this does not classify as force majeure. In consequence, force majeure in case of collisions with road users less than 14 years old is by Dutch courts almost never accepted.

Apart from this, the Dutch case law on article 185 WVW is quite diverse and does not yield clear rules on what counts as force majeure (Rutten and Oskam 2016). Nevertheless, a relevant cue for motor vehicle drivers entailed by Article 185 is that collisions should be avoided at all reasonable costs, where a cost is only in very exceptional cases regarded as unreasonable, and where the motor vehicle driver is expected to anticipate possible mistakes by pedestrians or cyclists. According to Rutten and Oskam (2016), for the extent to which a motor vehicle driver should anticipate mistakes, at least the following two factors are relevant. First, if there are special traffic characteristics, such as the presence of obstacles, pedestrian crossings, bicycle path or bus stops, then more care than usual needs to be taken. Second, if the motor vehicle driver saw or should have seen the other, then force majeure is in practice virtually never assumed. On the other hand, case law also teaches that civil liability on the basis of article 185 WVW does in general not

require conclusions about the mental state of the offender (negligence, intent). Generally, all that counts is whether observed behaviour violated a specific traffic rule or was otherwise blamable.

If the accident is between two motor vehicles, then there are no special liability rules and the general rules for tort liability apply. A full analysis of the relevant Dutch case law is beyond the scope of this paper (see e.g. van Wijk 2014) but the general standard emerging from it is that of the perfect driver. Even if two motor vehicles collide, the norms are stringent in that only perfect drivers can fully escape liability. The underlying principle is that cars are inherently dangerous and the consequences of even minor traffic mistakes can be very serious, so people who decide to use cars should be encouraged to drive according to the highest possible standards.

3.3 Summary on Dutch traffic law

Summarising, the main way in which Dutch traffic law promotes the purposes of safe and efficient traffic is with concrete traffic rules in the RVV, generally specifying which behaviour is obligatory, with exceptions in the form of explicit permissions. This is why it is so important that AV are designed in such a way that they conform to the traffic rules. In addition there is the general prohibition clause of Article 5 WVW which has two functions: it prohibits dangerous or impeding behaviour that is not (clearly) prohibited by specific traffic rules, and it provides a general implicit exception to the RVV rules in case otherwise obliged or permitted behaviour is dangerous or impeding. Moreover, civil liability rules, although primarily meant to allocate damages after something has gone wrong, yield additional cues for behaviour, which can be regarded to have the same functions as Article 5 WVW. These liability rules are meant to protect the more vulnerable road user against the stronger. Accordingly, motor vehicle drivers are almost always held liable towards non-motorised traffic participants. Even if two motor vehicles collide, the norms for liability are stringent in that only perfect drivers can fully escape liability. The underlying principle here is that cars are inherently dangerous objects and the consequences of even minor traffic mistakes can be very serious, so people who decide to use cars should be held to the highest possible driving standards.

More generally, looking at the Dutch law on liability for traffic offenses, the striking thing is that liability (whether criminal or civil) to a very large extent depends on just two factors: whether a traffic law rule was violated and (in some cases) whether some bad consequence of the violating act happened. If these two conditions are satisfied, chances are very high that the actor is held legally liable. Claims as to exceptional circumstances, like force majeure, are seldomly accepted. The general assumption underlying this is that dangerous behaviour is forbidden and the traffic rules define what is dangerous behaviour.

4 Autonomous vehicles and Dutch traffic law

In this section I discuss some challenges that Dutch traffic law presents for autonomous vehicles. I will first consider three different approaches to the design of law-conforming AV, namely, the *regimentation* approach (designing the system in a way that guarantees that the system will not exhibit unwanted behaviour), the *reasoning* approach (giving the system the ability to reason about the lawfulness of its own behaviour) and the *training* approach (letting the system acquire the ability to behave legally correctly implicitly by training). I will then evaluate each of these approaches in light of the challenges posed by Dutch traffic law.

4.1 Three approaches to achieving norm compliance

One solution to the problem of making AV conform to traffic law is to design the system in a way that guarantees that the system will not exhibit unwanted behaviour. This is the conventional solution when non-autonomous machines, tools or systems are used, sometimes called *regimentation*. A similar approach has been proposed for autonomous systems by Broersen (2014), who proposes to verify the behaviour of systems off-line with so-called model-checking techniques. Here the verification of the system's behaviour takes place off-line by comparing the system's actual behaviour with an external model of normatively correct behaviour, while the system itself does not reason with that model. Note that good practice in AI & Law (Bench-Capon and Coenen 1992; van Engers et al. 2001) requires that the regimented AV design must be linked to the relevant law for purposes of validation and maintenance. In this respect, Leenes and Lucivero (2014) observe that current designs of AV do not have an explicit traffic law model but that a usual approach to test AV is with 'play books', which contain scenarios the AV must be able to handle. They warn that the traffic law model implicit in such scenarios may be incomplete or oversimplified, while testing whether this is the case is difficult since the traffic model remains implicit.

When the relevant traffic model is made explicit to overcome these problems, then one issue is whether the model should be machine-processable in the same way as when designing explicit legal reasoners or whether it can simply consist in the selection of the relevant natural-language sources. The structure of Dutch traffic law gives a compelling reason why the specification has to be formal. Given the structure of the RVV and its hierarchical relation to the WVW, it is in general impossible to link specific behaviours or designs to specific rules, since the correct conclusion on the deontic status of a behaviour depends on the priority relations between the various rules. So even in the design phase the correct AV behaviour has to be found by reasoning with the relevant traffic laws and doing this in a reliable way requires formal specification of these laws. A second reason to have such a formal specification is that it could also be used to test the traffic regulations, for example, whether indeed all rule conflicts can be resolved with the relevant conflict resolution principles. The need for such legislative drafting support is arguably more

pressing when AV enter the roads, since AV arguably need more precise cues for behaviour than human traffic participants.

A limitation of the regimentation approach is that when systems are increasingly autonomous and have to operate in increasingly complex environments, their input and behaviour cannot be fully predicted, so that regimentation or advance off-line model checking are impossible or of limited value. How can norm compliance then be ensured? The question then arises whether an autonomous system should be designed to *reason* about how to behave lawfully by letting it reason about what is normatively correct behaviour, or whether it can be *trained* to do so with machine-learning techniques applied to a large number of training cases. In the reasoning approach there is the obvious need for explicit representation of legal information in the system and for giving the system explicit reasoning and decision making capabilities. This is still somewhat similar to the traditional AI & Law systems for supporting human decision making, except that the human is taken out of the loop. An important issue then is whether the mundane nature of cases faced by the autonomous system can reduce the complexity of the classification and interpretation problems to such an extent that the machine can fully take over. On the other hand, the reasoning can, unlike in the traditional settings, be opaque in that there is less need for explaining or justifying why the behaviour is legally correct.

The third approach is that the ability to behave legally correctly is acquired implicitly by training. It seems that the currently usual approaches to designing AV to a large extent rely on machine learning approaches. For example, the March 2016 edition of the Google Self-Driving Car Project Monthly Report[2] says

> ...rather than teaching the car to handle very specific things, we give the car fundamental capabilities for detecting other road users or unfamiliar objects, and then we give it lots of practice in a wide range of situations.

The training approach is similar to regimentation in that it aims to equip the AV with law-conforming behaviours without giving it explicit normative reasoning capabilities. However, this approach differs from regimentation in that it does not aim to fully *guarantee* correct behaviour. Accordingly, validation of correct behaviour is not done by formal means but by empirical testing. For very advanced autonomous systems, like robots operating in daily life, this approach might be equivalent to solving the notorious AI common-sense problem, but for more modest systems this approach might be more realistic. One interesting question is how autonomous vehicles classify on this scale. Below we will discuss some interpretation and classification problems in Dutch traffic law that are relatively easy for humans but seem very hard for the current generation of autonomous vehicles. The training approach does not necessarily avoid the need for explicit representation of legal rules and regulations. They must now be represented as part of the design specification. Like with the regimentation approach, the AV design should in agreement with good practice in AI & Law be linked to the relevant law

---

[2] https://static.googleusercontent.com/media/www.google.com/lt//selfdrivingcar/files/reports/report-0316.pdf (accessed January 5, 2017).

and for the reasons discussed above the specification of the relevant law should arguably be formal.

4.2 Required abilities of AV

I now discuss which abilities an AV that can fully autonomously drive in all Dutch traffic situations should have. A survey of the extent to which current AV technology provides these abilities is beyond the scope of this paper, but a few observations on this can still be made.

As regards interpreting sensor data, current AV design mainly seems to be concerned with cognitively speaking relatively 'low level' capabilities like determining the own and other road users' location, speed and direction, distinguishing drivable from non-drivable areas and recognising obstacles. According to Leenes and Lucivero (2014) current AV technology does not yet allow for advanced object recognition, while yet this is required by Dutch traffic law. Leenes and Lucivero (2014) discuss the example of Article 16 RVV, which states that 'Road users must not cut across military columns and motorised funeral processions'. This requires the ability to classify vehicles. Similarly, as we saw in Sect. 3, the AV needs to recognise other types of objects, such as tunnels, or bus stops. For humans such classification tasks are straightforward; the (legal or commonsense) definitions of the various types of vehicles and other objects are generally precise. The added complexity compared to traditional AI & Law is that the AV needs to perform such classification from its sensor data. This is a major problem, since according to Anderson et al. (2016) making sense of sensor data is probably the hardest part of designing fully autonomous AV.

Important kinds of objects to be recognised are traffic signs, traffic lights and road lining. A usual approach here is letting the AV use maps on which these things are indicated. This has two obvious limitations: permanently changed traffic situations not yet incorporated in the maps, and temporary changes, such as the directions given by authorised officials on the basis of Article 12 WVW or Article 82 WVV. According to Gomes (2014):

> Google's cars can detect and respond to stop signs that aren't on its map, a feature that was introduced to deal with temporary signs used at construction sites.

According to Gomes (2014), Google's reply to this is as follows:

> Google says that its cars can identify almost all unmapped stop signs, and would remain safe if they miss a sign because the vehicles are always looking out for traffic, pedestrians and other obstacles.

But this has the potential problem that the resulting behaviour does not comply with the second purpose of Dutch traffic law, namely, to promote efficient traffic. Overly cautious behaviour might result in a violation of Article 5 WVW's prohibition to impede other traffic. See also the case of the Google car driving too slowly discussed above in Sect. 3.1, and see Gomes (2014), who remarks that

But in a complex situation like at an unmapped four-way stop the car might fall back to slow, extra cautious driving to avoid making a mistake.

A step further than object classification and recognition of traffic signs and lights and road lining is situational awareness and interpretation. For example, an AV should be able to distinguish between ordinary pedestrians (merely to be avoided) and police officers giving directions. Gomes (2014) discusses how the Google car currently deals with this:

Pedestrians are detected simply as moving, column-shaped blurs of pixels – meaning, (...), that the car wouldn't be able to spot a police officer at the side of the road frantically waving for traffic to stop.

While the Google car may thus avoid colliding with a police officer, it may fail to obey the officer's directions.

Article 19 RVV, which states that 'A driver must at all times be able to bring his vehicle to a standstill within the distance that he can see to be clear', seems within the capabilities of current AV technology. However, the behavioural cues that can be derived from the liability rules of Article 185 WVW and the general tort liability rules require that an AV has sophisticated means to interpret a situation and its context from its sensor data. For example, an AV must be able to adjust its speed to the state of the road, the type of environment (country side, busy shopping street, and so on), to unusual weather conditions, to the presence of special features like bus stops, pedestrian crossings, objects besides the road that block or impair the car's view, and to the presence or absence of foot paths or bicycle paths. Moreover, knowledge of the local context may also be relevant. For example, in one Dutch case, the court remarked that in a large city like Amsterdam it is not unusual for cyclists to ride their bicycle in the dark without light (ECLI:NL:GHS-GR:2001:AK4463). If capabilities like object classification or recognising types of persons is still to a large extent beyond current AV technology, then the same can be inferred from this for these more advanced capabilities to interpret a situation from the AV's sensor data.

Finally, as remarked by Leenes and Lucivero (2014), the purpose to promote both safe and efficient traffic requires that an AV has the ability to recognise and conform to social cues. For instance, it should be able to interpret gestures by or eye contact with human traffic participants. While for humans this is generally easy, for AV this, too, seems a hard problem. More generally, AVs should have the capability to anticipate the behaviour of human traffic participants, since this capability is essential in avoiding accidents (Surden and Williams 2016). Perhaps the AV could employ a general expectation that traffic participants will more or less behave as they should. This implies that the AV should be able to recognise the type of participant to make the correct predictions on their behaviour, since different categories of road users are governed by different rules. However, this expectation comes with exceptions. For instance, as we saw in Sect. 3.2, Dutch case law on Article 185 WVW motor vehicle drivers should anticipate possible mistakes by non-motorised traffic participants.

4.3 Knowledge representation issues

Above we saw that in all three design approaches there is the need for formal specification of traffic law. In the reasoning approach this is obvious, while in the regimentation and training approach this need stems from the need to maintain links between the AV design and the law for purposes of validation and maintenance. Therefore it becomes relevant to discuss the formal specification of traffic law from the point of view of AI & Law research on knowledge representation.

4.3.1 Representing the regulations

Computational representation of the relevant regulations does from an AI & Law point of view seem straightforward.

First, unlike in other domains, the problem of knowledge acquisition or maintenance is not very complex. Compared to, for instance, social security law, tax law or immigration law (three areas where legal knowledge-based systems have been applied in The Netherlands), the body of relevant traffic law is manageable in size and relatively stable over time.

Next, since the safety and efficiency purposes of traffic law require it to be precise, the rules have a clear conditional and deontic structure, with arguably no syntactic ambiguities. So any language with the means to express if-then rules with boolean conditions and a way to express deontic qualifications and their relations of conflict (either through modalities or through first-order predicates) will do. Definitional and structural knowledge could be represented in a first-order language or encoded in logic-programming-like rules. There seems no need to have systematic accounts of time but some form of spatial traffic and commonsense knowledge should be represented. Perhaps some current research in the AV design world on ontologies can be reused here, such as of Morignot and Nashashibi (2013) and Zhao et al. (2015). We also saw that representing and reasoning about mental states is in general not required by the relevant laws, since the traffic rules hardly ever refer to a driver's mental state.

Third, as we saw above, the rules from Dutch traffic laws may conflict and such conflicts can often be resolved with explicit priority rules. Although Leenes and Lucivero (2014) note that subtle conflicts may arise that cannot be resolved with established priority principles, looking at the types of conflicts that can arise, it seems that they can generally be handled well by the specific conflict resolution rules in the RVV and Dutch case law and by the general *Lex Specialis* and *Lex Superior* principles. Now the problem of representing and reasoning with conflicting rules and their priority relations has essentially been solved by AI & Law. Two approaches are available. An approach that is the most faithful to the actual structure of the Dutch traffic regulations is to adopt a system that allows for rule conflicts and priorities, such as the one of Prakken and Sartor (1997) or the more recent $ASPIC^+$ framework (Modgil and Prakken 2013), or a suitable variant of defeasible logic (Antoniou et al. 2000; Governatori et al. 2010). Alternatively, rule priorities can be compiled away in the use of explicit exception clauses, as in e.g. the approaches of

McCarty and Cohen ([1994](#)) and Kowalski and Toni ([1996](#)). Both methods are well understood in AI & Law and in nonmonotonic logic in general.

As for the choice between the two methods, there is a subtlety. Above, I argued that given the hierarchical structure of and the relation between the WVW and the RVV, it is often impossible to link specific behaviours or designs to specific rules, since the correct conclusion depends on the priority relations between the various rules. Above I concluded that even in the design phase the correct AV behaviour has to be found by reasoning with the relevant traffic laws and that doing this in a reliable way requires formal specification of these laws. However, if the formal specification uses the method of implicit exceptions and priority rules, the problem remains. This speaks for a compilation approach, at least in the regimentation and learning approaches. In such a compilation approach, the initial formal representation is made as faithful to the structure of the WVW and RVV as much as possible, so with implicit exceptions and rule hierarchies. Then this initial formal specification is automatically compiled into another formal specification with explicit exception clauses and the AV design is linked to that second design. For evidence that the compilation approach is feasible see Delgrande and Schaub ([2000](#)), who systematically translate prioritised default logic with and without defeasible priorities (Brewka [1994](#)) into Reiter ([1980](#))'s original version of default logic.

To give an idea of how this could work, consider Article 3 RVV para 1 that drivers shall keep right as much as possible, Article 13 RVV that in case of traffic jams on motor ways keeping the rightmost lane is not obliged, and Article 82 para 1 RVV that drivers are obliged to obey the instructions of authorised officials. Schematically, a direct formalisation with implicit exceptions and priorities is:

3(1) RVV: IF Driving on public road THEN Obliged to keep rightmost lane
13 RVV: IF Driving on motor way in a traffic jam THEN NOT obliged to keep rightmost lane
82(1) RVV: IF Authorised official indicates to X THEN Obliged to X

Article 82(1) RVV has priority over both 3(1) RVV and 13 RVV on the basis of the special priority rule Article 84 RVV. Moreover, 13 RVV has priority over 3(1) RVV by the *Lex Specialis* principle.

This can be compiled into the following representation with explicit exception clauses:

3(1) RVV: IF Driving on public road AND NO Exception to 3(1) RVV THEN Obliged to keep rightmost lane
E1: IF Driving on motor way in a traffic jam THEN Exception to 3(1) RVV
E2: IF Authorised official indicates not to keep rightmost lane THEN Exception to 3(1) RVV
13 RVV: IF Driving on motor way in a traffic jam AND NO Exception to 13 RVV THEN NOT obliged to keep rightmost lane
E3: IF Authorised official indicates not to keep rightmost lane THEN Exception to 13 RVV

Moreover, for all three rules a technique like predicate completion can be used to list all exceptions to the rule. For example, for Article 3(1) this would yield:

> Exception to 3(1) RVV IFF Driving on motor way in a traffic jam OR Authorised official indicates to not to keep rightmost lane OR ...(other exceptions to 13(1) RVV).

And for 13 RVV it would yield:

> Exception to 13 RVV IFF Authorised official indicates to not to keep rightmost lane OR ...(other exceptions to 13 RVV).

Assume now that an AV designer wants to implement a policy to choose the fastest moving lane on a motorway in a traffic jam. This policy can be linked to the combination of the version of 13 RVV with an explicit exception clause and the predicate completion of this exception clause.

4.3.2 Modelling interpretation

While nonmonotonic reasoning techniques as previously developed or applied in AI & Law are useful, they also have limitations. Such techniques do not offer ways to recognise the need for an exception to a legal rule or to recognize the best way to resolve a conflict between regulations, unless this has been programmed into the system in specific terms. Moreover, if the rules contain general exception clauses like causing danger or impediment, the classification and interpretation problem will be too big. More generally, rule-based nonmonotonic techniques do not provide means to deal with open-textured and vague concepts and with unforeseen exceptions. In Sect. 3.2 we saw that the Dutch case law on liability for traffic accidents is quite diverse, with few general rules and much attention for the specifics of a case. At best, a number of relevant factors can be recognised. In AI & Law terms this body of case law is thus factor-based.

AI & Law provides sophisticated models of arguing with cases in factor-based domains (Ashley 1990; Aleven and Ashley 1995; Bench-Capon and Sartor 2003) but it is doubtful whether these models are useful for present purposes. To start with, some of the most influential AI & Law models of case-based reasoning were explicitly meant to be used in adversarial contexts. For example, Ashley (1990), in motivating his HYPO system, sketched a vision of a system which could support an advocate charged with preparing a case at short notice. The system would be able to accept the facts of the case and then generate arguments for the two sides to the case and counterarguments to them, together with the precedents on which they are based. HYPO's successor CATO (Aleven 2003) has the same adversarial focus, now in teaching contexts. As noted above in Sect. 2, with the task of letting autonomous systems conform to the law there are no adversaries: all that counts is to let the system do what it has to do within the bounds of the law. Thus there will be less need for argumentation than in as usual in AI & Law. Moreover, the task of generating safe and efficient traffic behaviour means that the case-based reasoning should result in reliable and useful conclusions on how to drive. Here a problem is that the amount and nature of relevant Dutch case law is such that it does not

support useful conclusions for a run-time reasoning AV. In fact, for this reason, machine learning techniques for statistically analysing a body of case law, as applied by e.g. van Opdorp et al. (1991) and Bench-Capon (1993), will also not be useful, since their models will not be reliable enough. Although the drawback that they have for more traditional AI & Law applications (that they are notoriously bad in explaining their output) does not apply to the present application, the relevant case law does not meet the three requirements for making them work: a known and stable set of relevant factors, many decided cases, and little noise among or inconsistency between these cases.

Having said so, these negative conclusions on the usefulness of AI & Law models of case-based reasoning only apply if the AV is designed to reason with the traffic law. In the regimentation and learning approaches it may still be useful to represent case law in the traffic model and link it to components of the AV design. For example, cases may be used to generate test situations for the play books.

Finally, as regards the open-textured terms in the RVV (which are not so vague as in Articles 5 and 185 WVW and in general Dutch tort law) a different approach may be possible. It may be possible to specify a behaviour or test case constraint that operationalises the concept. A similar approach may be possible (but with greater difficulty) for the general fall-back clauses and liability rules. In case of vague or open-textured terms the full legal meaning does not need to be specified; all that is needed is a policy for compliance which operationalises the concept. Here the industry and government could collaborate in developing standards for AV behaviour (a possibility discussed by Anderson et al. (2016) for the USA). And perhaps internal guidelines of insurance companies on how they deal between each other with liability issues can be used (another possibility discussed by Anderson et al. (2016) for the USA). Nevertheless, there is still the specification issue, since such operationalisations and standards should still be linked to the relevant regulations.

5 Related research

Arguably the main publication relevant to the present study is Leenes and Lucivero (2014). It was a source of inspiration for the current paper and it was discussed in several places above. As for other relevant work, in the early days of AI & Law research, Den Haan (1996) and Den Haan and Breuker (1991) applied knowledge representation and reasoning techniques to a part of Dutch traffic law. Den Haans research was partly a theoretical exercise and partly intended to develop support tools for legislative drafting. Although her specific model of reasoning with rules hierarchies was, in light of later developments, somewhat ad hoc and outdated, this is understandable given the state of the art at that time. In her modelling of the Dutch traffic law domain, Den Haan separated world knowledge (descriptive, that is, terminological and structural) and causal/intensional, that is, actions plus processes) from normative knowledge. She regarded statutory definitions as world knowledge. In her actual world knowledge model, she only modelled descriptive knowledge and no causal or commonsense (for instance, spatial) knowledge. Rule

conflicts were in her model resolved with the Lex Specialis principle, which as noted above is feasible since the Dutch RVV was drafted with this principle in mind. Den Haan's model supported reasoning with alternative viewpoints on the basis of alternative semantic interpretations of open-textured terms, but she did not model argumentation about choosing specific interpretations. Moreover, she did not model the step from fact to legal concept, so she also left the problem of generating alternative interpretations to future research. Finally, she modelled only a small part of the RVV, dealing with the place on the road (Articles 3–10 RVV) and giving way (Article 15 RVV).

As remarked in the introduction, there have to the best of my knowledge so far not been any systematic studies of relating traffic regulations to AV designs. There is some work on combining ontologies with rule languages, but the example rules are still ad hoc and no general modelling method is proposed. Morignot and Nashashibi (2013) propose a symbolic approach to modelling AV behaviour. They combine a traffic ontology represented in Protégé[3] with traffic behaviour rules modelled in SWRL (Horrocks et al. 2004). Morignot and Nashashibi want to allow for behaviour that violates the traffic rules. Accordingly, the behavioural rules can have as a condition that a considered behaviour is illegal but there are no rules for deriving conclusions about this. Zhao et al. (2015) have two simple examples of what they call traffic rules in SWRL. These rules seem to be behavioural rules that are only implicitly based on traffic regulations.

There is currently much attention for so-called 'moral algorithms' for letting AV deal with situations akin to moral dilemmas. For example, Bonnefon et al. (2016) study moral decision problems that AV could face by presenting people with several simple situations and asking them what the AV should do. In one such situation the AV has to choose between killing a pedestrian or killing the AV's passenger by driving into a wall. Bonnefon et al. conclude that it is important to study which moral algorithms should be programmed into AV. From our legal point of view, this importance can be challenged. As we saw above, the cue for driving behaviour implied by article 185 WVW in this example is to avoid collisions with pedestrians at all reasonable cost, where a cost is not easily determined to be unreasonable. The main emphasis in Dutch case law is not on what to choose in dilemmas such as the ones discussed by Bonnefon et al. but on anticipation, that is, on how much care has to be taken to avoid such situations. Therefore, instead of focusing on moral algorithms for situations of the kind studied by Bonnefon et al. the first priority should be to study 'legal algorithms' for avoiding such situations, that is, for careful and prudent driving behaviour that yet does not make the traffic too slow to be efficient. In Dutch case law, situations of the kind discussed by Bonnefon et al. hardly ever occur. The closest is a case in which a driver at night suddenly swerved to the left lane because a roe suddenly crossed the road from the forest, after which the car crashed into another car. According to van Wijk (2014, pp. 86-87), the Dutch Supreme Court held in this case that the driver could not be expected to choose for the real risk to seriously harm himself.

---

[3] http://protege.stanford.edu.

It might be asked in a more general sense to what extent the problems of making autonomous systems conform to the law and making them conform to ethical rules and principles differ. For many applications the differences may be considerable but for autonomous vehicles the legal and ethical requirements arguably almost completely coincide, since traffic ethics has to be derived from the same values that traffic law is meant to promote, namely safe and efficient traffic.

One issue here concerns the relation between legal and ethical rules. At first sight, it would seem that the aim to make an AV (or another kind of autonomous system) conform to the law runs the risk of ignoring ethical reasons for overriding the law. However, this risk is not very real since law application also involves considering the social context and issues of fairness, justice, common sense and the like. One way the law accounts for this is by way of general exception categories and in many cases ethical reasons for overriding a legal rule will classify under a general legal exception category. For example, if slightly speeding or safely driving through a red light is necessary to transport a wounded person to hospital as quickly as necessary, then in Dutch law the traffic rule to stop in front of a red light is set aside by a general principle that violating a criminal provision is allowed if this is necessary for protecting a legitimate interest. Even if no statutory exception category is available, then legal systems allow rules being set aside by unwritten principles, as in the well-known case discussed by Dworkin (1977) of a grandson who had killed his grandfather and then claimed his share of the inheritance under US inheritance law. The court rejected his claim on the ground that nobody should profit from their own wrongdoing.

A final piece of related research relates to my suggestion in Sect. 4.2 with reference to Surden and Williams (2016) that AV should have the capability to anticipate the behaviour of human traffic participants, since this capability is essential in avoiding accidents. In fact, Surden and Williams discuss the reverse problem of humans traffic participants having to interpret the behaviour of AV, with the same underlying motivation. Surden and Williams observe that current AVs are unpredicable to human traffic participants since AVs are both self-controlled and have a large degree of freedom of choice, while currently their design is 'technically opaque' to ordinary humans. They then suggest technical design solutions to make AV more predictable to ordinary humans in traffic, such as the ability of the AV to communicate to people that they have been detected and to communicate the AV's intentions to these people. They also discuss how government policy and the law can encourage such technical solutions.

6 Conclusion

This paper has studied the problem of making autonomous vehicles conform to traffic law by way of a case study of Dutch traffic law. While several features of Dutch traffic law are specific to the Dutch jurisdiction, its main features are similar to traffic laws in other western jurisdictions, especially European ones, so the main conclusions are arguably generalisable beyond Dutch law. The paper's main contributions are as follows. For AI & Law it has put a new topic on its research

agenda. For AV designers it has hopefully created a deeper awareness of the challenges that traffic law poses for AV design. To start with, an account was given of the similarities and differences between the tasks traditionally modelled in AI & Law and the new task of autonomous driving on public roads. Next, the need was identified for formal specification of traffic law as a component of AV design even if the AV does not itself reason with such specifications. As regards the applicability of knowledge representation and reasoning techniques previously developed in AI & Law, the conclusion was that logic-based techniques for rule-based (monotonic or nonmonotonic) reasoning are largely suitable for representing the logical and hierarchical structure of the regulations but cannot deal with the interpretation problems arising from open texture, vagueness and the need for new exceptions. On the other hand, existing argumentation-based techniques for dealing with open texture, vagueness and the need for new exceptions turned out to be probably largely inapplicable, partly since the adversarial setting presumed by these techniques is lacking in the AV problem and partly because the existing case law is too sparse and inconclusive. A more promising approach seems to aim at developing standards and guidelines for implementing law-conforming behaviours in a collaborative effort between the government, industry and possibly insurance companies. However, this does not avoid the need for formal representation of the relevant traffic regulations. Finally, the present paper has provided reasons for the relevant research communities to focus less on moral algorithms and more on careful and anticipatory driving behaviour (while not sacrificing traffic efficiency).

The present study has focused on fully autonomous self-driving cars. For less advanced stages of automation-assisted driving, some of the negative conclusions may not hold or may hold to a lesser extent while other conclusions may still hold with the same force, depending on the degree and kind of autonomy of the AV. Moreover, in the future traffic laws may be adapted to the presence of AV on the road. Nevertheless, the present paper has in any case laid the foundations for a systematic study of the problem of making autonomous vehicles conform to the law.

Finally, the question arises to what extent this paper's conclusions apply to the more general problem of making autonomous systems conform to the law. Here it is worth noting that the idea of following rules and how best to do so is really a rather deep and complicated problem more generally, not just in the domain of AI (Brozek 2102). So a general study of making autonomous systems conform to the law could arguably benefit from a general theory of what rule following amounts to. However, for present purposes such a general theory was not required, since, as our case study has indicated, traffic law is rather precise and uncontroversial, while it hardly refers to the traffic participants' mental state: given the underlying purposes of promoting safe and efficient transportation, all that counts in traffic law is that the right behaviour is generated. It remains to be seen whether similar special-purpose approaches are suitable other kinds of autonomous systems or whether a more general theoretical account of rule following is needed.

References

Aleven V (2003) Using background knowledge in case-based legal reasoning: a computational model and an intelligent learning environment. Artif Intell 150:183–237

Aleven V, Ashley K (1995) Doing things with factors. In: Proceedings of the fifth international conference on artificial intelligence and law, ACM Press, New York, pp 31–41

Anderson M, Anderson S (eds) (2011) Machine ethics. Cambridge University Press, Cambridge

Anderson JM, Kalra N, Stanley K, Sorensen O, Samaras C, Oluwatola O (2016) Autonomous vehicle technology. A guide for policy makers. RAND Corporation, Santa Monica

Antoniou G, Billington D, Governatori G, Maher M (2000) A flexible framework for defeasible logics. In: Proceedings of the 17th national conference on artificial intelligence (AAAI-00), pp 405–410

Ashley K (1990) Modeling legal argument: reasoning with cases and hypotheticals. MIT Press, Cambridge

Bench-Capon T (1993) Neural networks and open texture. In: Proceedings of the fourth international conference on artificial intelligence and law, ACM Press, New York, pp 292–297

Bench-Capon T, Coenen F (1992) Isomorphism and legal knowledge based systems. Artif Intell Law 1:65–86

Bench-Capon T, Sartor G (2003) A model of legal reasoning with cases incorporating theories and values. Artif Intell 150:97–143

Bonnefon JF, Shariff A, Rahwan I (2016) The social dilemma of autonomous vehicles. Science 352:1573–1576

Brewka G (1994) Reasoning about priorities in default logic. In: Proceedings of the 12th national conference on artificial intelligence (AAAI-94), pp 247–260

Broersen J (2014) Responsible intelligent systems. The REINS project. Kuenstliche Intell. 28:209–214

Brozek B (2102) Rule-following. From imitation to the normative mind. Copernicus Center Press, Krakow

Delgrande J, Schaub T (2000) Expressing preferences in default logic. Artif Intell 123:41–87

Den Haan N (1996) Automated legal reasoning. Doctoral dissertation. Faculty of Law, University of Amsterdam

Den Haan N, Breuker J (1991) A tractable juridical KBS for applying and teaching traffic regulations. In: Breuker J, Mulder RD, Hage J (eds) Legal knowledge-based systems. JURIX 91, Model-based reasoning, Koninklijke Vermande, Lelystad, pp 5–16

Dworkin R (1977) Is law a system of rules? In: Dworkin R (ed) The philosophy of law. Oxford University Press, Oxford, pp 38–65

Gomes L (2014) Hidden obstacles for Google's self-driving cars. MIT Technology Review. https://www.technologyreview.com/s/530276/hidden-obstacles-for-googles-self-driving-cars/

Governatori G, Rotolo A, Rubino R (2010) Implementing temporal defeasible logic for modeling legal reasoning. In: JSAI-isAI 2009 workshops, LENLS, JURISIN, KCSD, LLLL, Tokyo, Japan, November 19–20, 2009, revised selected papers, Springer, Berlin, no. 6284 in Springer lecture notes in AI, pp 45–58

Horrocks I, Patel-Schneider PF, Boley H, Tabet S, Grosof B, Dean M (2004) SWRL: a semantic web rule language combining OWL and RuleML. Submission to W3C

Kowalski R, Toni F (1996) Abstract argumentation. Artif Intell Law 4:275–296

Leenes R, Lucivero F (2014) Laws on robots, laws by robots, laws in robots: regulating robot behaviour by design. Law Innov Technol 6:193–220

MacCormick N (1998) Legal reasoning and interpretation. In: Routledge encyclopedia of philosophy. Taylor and Francis, pp 525–530

McCarty L, Cohen W (1994) The case for explicit exceptions. Methods Log Comput Sci 1:19–50

Modgil S, Prakken H (2013) A general account of argumentation with preferences. Artif Intell 195:361–397

Morignot P, Nashashibi F (2013) An ontology-based approach to relax traffic regulation for autonomous vehicle assistance. In: Proceedings of the 12th IASTED conference on artificial intelligence and applications, Insbrück, Austria

Palmerini E, Azzarri F, Battaglia F, Bertolini A, Carnevale A, Carpaneto J, Cavallo F, Carlo AD, Cempini M, Controzzi M, Koops BJ, Lucivero F, Mukerji N, Nocco L, Pirni A, Shah H, Salvini P, Schellekens M, Warwick K (2014) Guidelines on regulating robotics. RoboLaw deliverable D2.6, FP7-SiS project on regulating emerging robotic technologies in Europe: robotics facing law and ethics

Prakken H, Sartor G (1997) Argument-based extended logic programming with defeasible priorities. J Appl Non Class Log 7:25–75

Reiter R (1980) A logic for default reasoning. Artif Intell 13:81–132

Rutten R, Oskam V (2016) Overmacht in het verkeer; een zeldzaamheid of toch niet? Verkeersrecht pp 322–329 **(in Dutch)**

Surden H, Williams MA (2016) Technological opacity, predictability, and self-driving cars. Cardozo Law Rev 38:121–181

Thrun S (2010) Toward robotic cars. Commun ACM 53:99–106

Vellinga N (2017) From the testing to the deployment of self-driving cars: legal challenges to policymakers on the road ahead. Comput Law Secur Rev. doi:10.1016/j.clsr.2017.05.006

van Engers T, Gerrits R, Boekenoogen M, Glassée E, Kordelaar P (2001) POWER: using UML/OCL for modeling legislation—an application report. In: Proceedings of the eighth international conference on artificial intelligence and law, ACM Press, New York, pp 157–167

van Opdorp G, Walker R, Schrickx J, Groendijk C, van den Berg P (1991) Networks at work. A connectionist approach to non-deductive legal reasoning. In: Proceedings of the third international conference on artificial intelligence and law, ACM Press, New York, pp 278–287

van Wijk M (2014) Verkeersaansprakelijkheid. Het Civiele en Strafrechtelijke Schuldbegrip bij Verkeersongevallen. Celsus juridische uitgeverij, Tilburg **(in Dutch)**

Zhao L, Ichise R, Mita S, Sasaki Y (2015) Ontologies for advanced driver assistance systems. In: Proceedings of the 35th semantic web & ontology workshop, pp 1–6