Law, Innovation and Technology

ISSN: 1757-9961 (Print) 1757-997X (Online) Journal homepage: https://www.tandfonline.com/loi/rlit20

Laws on Robots, Laws by Robots, Laws in Robots: Regulating Robot Behaviour by Design

Ronald Leenes & Federica Lucivero

To cite this article: Ronald Leenes & Federica Lucivero (2014) Laws on Robots, Laws by Robots, Laws in Robots: Regulating Robot Behaviour by Design, Law, Innovation and Technology, 6:2, 193-220, DOI: 10.5235/17579961.6.2.193

To link to this article: https://doi.org/10.5235/17579961.6.2.193

Published online: 07 May 2015.

DOI: http://dx.doi.org/10.5235/17579961.6.2.193 (2014) 6(2) LIT 193–220

Laws on Robots, Laws by Robots, Laws in Robots: Regulating Robot Behaviour by Design

Ronald Leenes and Federica Lucivero*

I. INTRODUCTION

The idea of robots taking over and replacing humanity continues to spark people's imagination. This may partially be attributable on the fact that, since they were introduced in Karel Čapek's *RUR* (*Rossum's Universal Robots*) in 1920,[1] robots have often been presented as humanoids. This humanoid representation has raised questions concerning robots' abilities to act morally and respectfully towards human beings. Isaac Asimov's three laws of robotics, which first appeared in a story called 'Runaround',[2] are probably the best-known answer to these interrogations. According to these laws:

1. A robot may not harm a human being, or, through inaction, allow a human being to come to harm.
2. A robot must obey the orders given to it by human beings, except where such orders would conflict with the First Law.
3. A robot must protect its own existence, as long as such protection does not conflict with the First or Second Law.

Asimov's rules are compelling because they clearly place robots as subordinate to human beings. Robots are not to harm human beings, they must follow humans' orders, and

* Ronald Leenes is Professor in Regulation by Technology at the Tilburg Institute for Law, Technology and Society, Tilburg University, The Netherlands. Federica Lucivero is currently a Marie Curie Fellow in the Department of Social Science, Health & Medicine, King's College London, UK. Federica was a Post-Doc fellow at TILT at the time of writing this paper. The research leading to his paper received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no 289092 (RoboLaw). The authors wish to thank their research assistant Marlou Brokx. All websites accessed November 2014.

1 Karel Čapek, *RUR* (Pegasus, 2010 [1920]).

2 Isaac Asimov, 'Runaround' (first published 1942), reprinted in Isaac Asimov, *The Complete Robot* (Voyager, 1995).

only as a third priority may they look after their own safety. From a Science Fiction (Sci-Fi) novelist's perspective, these laws are ideal. They clearly define the relation between humans and robots. They are also easily broken (especially rule number one), creating situations that trigger rule conflicts—and these provide rich territory for Sci-Fi stories.

Within the ethical and philosophical debate, the question of robot morality and the meaning of moral laws regulating robot behaviours have been valid topics in reflecting on the meaning of humanity in light of emerging technology and vice versa.[3] Despite the widespread influence of the three laws of robotics and their role in shaping visions of future robo-dense worlds in Sci-Fi and philosophy, hands-on roboticists have largely neglected them. Indeed, the robots presented in literature and movies bear little resemblance to the state of the art in robotics. Even the most advanced current humanoid robots pale in comparison to the C-3POs (and even R2D2), Wall-es and Datas that we have come to love from the big screen. Contemporary roboticists face much more mundane questions than how to prevent robots from running haywire and attacking humans. Instead they have a hard time figuring out how to solve the problems of space navigation, locomotion, obstacle avoidance,[4] learning by example and grasping objects like eggs without pulverising them.[5]

This gap between Asimov's laws and mundane issues of functional robot design and task definition is, however, much narrower than it seems at first sight. Indeed, several scholars in philosophy, ethics and sociology of technology have highlighted the morally relevant nature of design choices.[6] For example, a decision not to provide users with access to a machine's control panel contributes to distributing roles and responsibilities among social actors concerning who is authorised to fix a problem.[7] Technology is never neutral and it promotes some values and demotes others in a process of mutual shaping

3 Mark Coeckelbergh, 'Moral Appearances: Emotions, Robots, and Human Morality' (2010) 12 *Ethics and Information Technology* 235; Patrick Lin, Keith Abney and George A Bekey, *Robot Ethics: The Ethical and Social Implications of Robotics* (MIT Press, 2012); Amanda Sharkey, 'Robots and Human Dignity: A Consideration of the Effects of Robot Care on the Dignity of Older People' (2014) 16 *Ethics and Information Technology* 63; Wendell Wallach and Colin Allen, *Moral Machines: Teaching Robots Right from Wrong* (Oxford University Press, 2009).

4 Chris Urmson *et al*, 'Autonomous Driving in Urban Environments: Boss and the Urban Challenge' (2008) 25 *Journal of Field Robotics* 425; Sebastian Thrun, 'Toward Robotic Cars' (2010) 53(4) *Communications of the ACM* 99.

5 See eg George A Bekey, Robert Ambrose, Vijay Kumar, David Lavery, Arthur Sanderson, Brian Wilcox, Junku Yuh and Yuan Zheng, *Robotics: State of the Art and Future Challenges* (Imperial College Press, 2008); Giuseppe Carbone (ed), *Grasping in Robotics* (Springer, 2012).

6 See, *inter alia*, Peter-Paul Verbeek, 'Design Ethics and the Morality of Technological Artefacts' in Pieter E Vermaas, Peter Kroes, Andrew Light and Steven A Moore (eds), *Philosophy and Design: From Engineering to Architecture* (Springer, 2008) 91–103; Bruno Latour and Couze Venn, 'Morality and Technology: The End of the Means' (2002) 19 *Theory, Culture & Society* 247; Ibo Van de Poel, 'Values in Engineering Design' in Anthonie WM Meijers (ed), *Philosophy of Technology and Engineering Sciences* (Elsevier, 2009) 973.

7 Madeline Akrich, 'The Description of Technological Objects' in Wiebe E Bijker and John Law (eds), *Shaping Technology Building Society: Studies in Sociotechnical Change* (MIT Press, 1992).

with the societal 'moral landscape'.[8] If design can be 'value-sensitive',[9] engineers' choices in designing robots require a normative and moral reflection on the types of values that will be embedded in them.[10]

The gap between norms and robots is even narrower if we consider the fact that they function in highly regulated normative contexts. Hence, not only can robots be expected to create a moral landscape, by displaying, enacting and promoting some values while demoting others, they will also have to adhere to and respect the existing normative landscape. Robots operating in human environments have to observe social and legal norms. For example, robots operating in public spaces, such as roads and pavements, should not cross the street whenever it suits them, but will have to comply with traffic regulation. At a higher level of complexity, companion robots should observe the socially accepted norm that it is inappropriate to simply intervene and hijack an ongoing conversation. Life for humans is full of social and legal norms and it might be reasonable to require robots to comply with at least a minimal set of these as well. These rules may be much more mundane and concrete than Asimov's moral imperative and yet they present a normative structure whose relevance in a specific situation has to be recognised by the robot, in order to adopt a compliant behaviour.

Given the limited capabilities of current robots to learn, these norms will have to be designed into the system. Norm compliance usually requires a mapping of concrete (intended) behaviour to relatively abstract norms. For instance, criminal provisions penalising theft usually do not mention concrete objects that can be stolen, but rather contain abstract notions, such as 'any good'. Concrete actions, such as placing a bag of sweets in one's pocket and leaving the candy shop without paying, will have to be characterised as an instance of appropriating a good without the owner's consent and hence falling within the scope of the theft provision, making the act 'theft'. Humans tend to cope reasonably well with these relatively abstract legal and social norms, with disputes regarding the meaning of norms and legal qualification of facts being the exceptions rather than the rule. Computers, on the other hand, despite nowadays spending most of their time processing data and information rather than crunching numbers, have difficulty coping with legal norms. Progress in the roughly 30-year-old field of Artificial Intelligence and Law compared to progress in domains such as image processing and

8 See eg Wiebe E Bijker, Thomas P Hughes and Trevor Pinch, *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology* (MIT Press, 1987); Tsjalling Swierstra, Dirk Stemerding and Marianne Boenink, 'Exploring Techno-Moral Change: The Case of the Obesity Pill' in Paul Sollie and Marcus Düwell (eds), *Evaluating New Technologies: Methodological Problems for the Ethical Assessment of Technology Developments* (Springer, 2009) ch 9, 119, http://link.springer.com/chapter/10.1007%2F978-90-481-2229-5_9.

9 Batya Friedman, Peter H Kahn and Alan Borning, *Value Sensitive Design: Theory and Methods*, University of Washington CSE Technical Report, 2 December 2001, www.urbansim.org/pub/Research/ResearchPapers/vsd-theory-methods-tr.pdf.

10 Aimee van Wynsberghe, *Designing Robots with Care: Creating an Ethical Framework for the Future Design and Implementation of Care Robots* (University of Twente, 2012).

search is rather limited.[11] Legal expert systems that apply codified legal knowledge to cases do exist,[12] but these are generally limited to small, well-defined domains and aim at solving specific tasks within these domains. In the case of robots, adding to the existing complexity of coping with symbolic information in the guise of rules and 'facts', come sensor input and actuator output that have to be mapped onto legal and social knowledge in order for them to generate norm-compliant behaviour.

If robots are to behave in a way that is compliant with certain legal and social norms, the question arises as to how this can be achieved. The present paper explores this question. We will first elaborate on the different ways in which technology can be regulated (section II), introducing four types of regulation. Next we will introduce the example of self-driving cars (section III), which will be used to describe the four types of robot regulation (section IV). In so doing we will point out the importance of moving from a focus on the regulation of human beings to a focus on the regulation of robot behaviour through design (techno-regulation and value sensitive design) (section V). Section VI discusses various conceptual, technical and normative issues related to the attempt to regulate robot behaviour through design.

II. REGULATING ROBOTS

In his recent work on the 'laws of robots', Ugo Pagallo[13] points out that in addressing the question of the laws of robotics scholars tend to either (1) look at how robots affect legal concepts and principles, as for example in the case of the concept of 'legal personhood'; or (2) analyse the way they create new principles and concepts, as in the case of liability for transactions conducted by autonomous agents; or (3) deny their specificity and turn the challenges presented by robots into questions addressable within the boundaries of the current system.[14]

In a similar vein, Bibi Van den Berg,[15] building on Roger Brownsword and Karen

11 For an overview of the various research strands in AI and law, see eg Andreas Margelisch, 'A State of the Art Report on Legal Knowledge-Based Systems', Swiss Life Information Systems Research Group, 2003, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.49.6315&rep=rep1&type=pdf. See also Trevor Bench-Capon *et al*, 'A History of AI and Law in 50 Papers: 25 Years of the International Conference on AI and Law' (2012) 20 *Artificial Intelligence and Law* 215; Radboud Winkels' ICAIL 2013 address, www.iaail.org/docs/PresidentialAddressICAIL2013.pdf, which shows the rise and decline of AI and Law research.

12 See Margelisch (n 11) 29.

13 Ugo Pagallo, *The Laws of Robots: Crimes, Contracts, and Torts* (Springer, 2013).

14 Pagallo's approach consists of describing how legal systems deal with robotic innovation 'through such a complex network of concepts as agency, accountability, liability, burden of proofs, clauses of immunity, or unjust damages'. In doing this, he analyses the conditions of legitimacy that are set for the design, construction and use of robots. The focus of this analysis is on the 'principle of responsibility before the law': the author asks how this principle is challenged in cases of criminal, civil (contract) and tort law concerning robots and reflects on existing sources of agreement. See Pagallo, *ibid*, 11.

15 Bibi van den Berg, 'Techno-Elicitation: Regulating Behaviour through the Design of Robots' in Bibi van den Berg and Laura Klaming (eds), *Technologies on the Stand: Legal and Ethical Questions in Neuroscience and Robotics* (Wolf Legal Publishers, 2011) ch 19, 403–22.

Yeung's distinction between regulation *of* and regulation *through* technology,[16] points out that most scholars in the field of law and robotics address questions of regulation *of* robotics. They focus primarily on whether robots raise legal problems and whether these can be addressed within the existing legal framework. Such legal analyses usually address questions of liability,[17] the legal status of robots[18] and rights for robots.[19] What according to Van den Berg is largely neglected is how human behaviour is being regulated *through* robotics. She claims that human behaviour will increasingly be regulated through technology and that the consequences of this indirect form of regulation should be studied more explicitly.

The idea that technology can regulate human behaviour is not new.[20] Langdon Winner pointed this out in a groundbreaking paper more than 30 years ago.[21] Famously, Winner argued that the low height design of the fly-overs on Long Island was intentional, serving to prevent public buses from passing under them—which, in turn, had the strong normative implication of discouraging or even forbidding low-income people from reaching the Long Island beaches. But the idea of regulating through technology was popularised among legal scholars through Lawrence Lessig's idea of 'code as law';[22] that is to say, the idea that code (architecture or design) of technological artefacts can be intentionally used to enforce legal norms. This concept of *regulation by technology* is also referred to as 'techno-regulation'.[23] Examples of such regulation include speed ramps that enforce drivers to comply with speed limits by imposing an immediate sanction (damage to the driver's car) in the event of attempted evasion of the norm, and technical measures such as 'region codes' implemented in DVD players in order to segment markets and thus regulate what people can watch on DVD. Van den Berg argues for a widening of studies of law and robotics by including a reflection on how human behaviour can be regulated through the design and functionalities of robots.

If the distinction between regulation *of* robots and regulation *through* robots has the merit of pointing out how robots can be both the regulated entity from the regulating means, there is another aspect of robot regulation that could be pointed out. The tradi-

16 Roger Brownsword and Karen Yeung, *Regulating Technologies: Legal Futures, Regulatory Frames and Technological Fixes* (Hart Publishing, 2008) 3.

17 Samir Chopra and Laurence F White, *A Legal Theory for Autonomous Artificial Agents* (University of Michigan Press, 2011) 119.

18 Lawrence B Solum, 'Legal Personhood for Artificial Intelligences' (1992) 70 *North Carolina Law Review* 1231.

19 eg Gunther Teubner, 'Rights of Non-Humans? Electronic Agents and Animals as New Actors in Politics and Law' (2007) 33 *Journal of Law and Society* 497; K Darling, 'Extending Legal Rights to Social Robots', 'We Robot' Conference, University of Miami, April 2012, http://ssrn.com/abstract=2044797.

20 It can in fact be traced back at least to Bentham's Panopticon (1787).

21 Langdon Winner, 'Do Artifacts Have Politics?' (1980) 109 *Daedalus* 121.

22 Lawrence Lessig, *Code and Other Laws of Cyberspace* (Basic Books, 1999) 6.

23 Roger Brownsword, 'What the World Needs Now: Techno-Regulation, Human Rights and Human Dignity' in Roger Brownsword (ed), *Global Governance and the Quest for Justice* (Hart Publishing, 2004) ch 13, 203; Ronald Leenes, 'Framing Techno-Regulation: An Exploration of State and Non-State Regulation by Technology' (2011) 5 *Legisprudence* 143.

tional focus of regulation scholars has been the regulation of human behaviour through law. Although Lawrence Lessig broadens the focus of regulation to include architecture (code)[24] as a means for regulation, those who are subject to regulation in his approach are still humans. However, when we bring robots into the equation, their behaviour, as well as human behaviour, is (or can be) the target of regulation. The means that not only human behaviour but also robot behaviour can be regulated by law as well as code. In view of having robots and humans occupy the same space in an orderly manner, we can distinguish the following categories:

1. Regulating *robot design, production* through *law*.
2. Regulating *user behaviour* through the *robot's design*.
3. Regulating the effects of *robot behaviour* through *law*.
4. Regulating *robot behaviour* through *code.*

While in 1 the regulatees are the human producers and designers (for example through ISO standards) and users (liability norms), in 3 we refer to (the legal effects of) robot behaviour regulated or rather constituted by law. For instance, whether robots can enter into contracts has to be defined by law.[25] If 2 is concerned with the regulation of human behaviour by designing robots that do not allow unlawful actions (of humans), 4 concerns the regulation of robot behaviour by embedding code that limits the robot to engaging only in certain (lawful) action. Before elaborating further on these four types of robo-regulation we will introduce an example of a type of robot currently being developed that touches upon all four categories: automated vehicles.

III. REGULATING ROBOTS: THE EXAMPLE OF SELF-DRIVING CARS

Car manufacturers and research institutes around the globe are experimenting with vehicles with different degrees of autonomy: partial, high and full automation.[26] Google has already covered several thousand kilometres on public roads with fully automated vehicles.[27] Public investment in automated cars is conspicuous. The EU has invested

24 And market, and social norms, but these are less relevant here.

25 See Pagallo (n 13) 101; Chopra and White (n 17) 29.

26 This categorisation has been developed by the German Federal Highway Research Institute (BASt): www.bast.de/EN/Home/home_node.html.

27 On 7 August 2012, Google reported that their self-driving cars ('about a dozen' of them) had completed more than 300,000 miles of testing: http://googleblog.blogspot.nl/2012/08/the-self-driving-car-logs-more-miles-on.html. On 28 April 2014, Google reported their cars to have covered 'thousands' of miles on the streets in Mountain View, California: http://googleblog.blogspot.nl/2014/04/the-latest-chapter-for-self-driving-car.html.

heavily in automated systems on the assumption that they address several challenges that have been set as strategic priorities by the Directorate General of Transport and Mobility.[28] Automated vehicles hold the promise of increasing traffic safety by reducing accidents due to human error, such as those resulting from driver distraction or reduced vigilance. They are also expected to reduce fuel consumption and reduce traffic congestion by optimising driving styles. Sustainability of energy consumption, safety on the road and accessibility to transport are indeed high on the European Commission's list of priorities.[29]

The US Defense Advanced Research Projects Agency (DARPA) also expects positive outcomes from scientific research on automated cars. In order to incentivise it, DARPA organised two 'Grand Challenge' events respectively in 2004 and 2005, wherein research groups from academia were encouraged to develop automated vehicles able to race on a desert trail. In 2007 a new competition was launched: the Urban Challenge. In this event autonomous vehicles had to be 'capable of driving in traffic, performing complex manoeuvres such as merging, passing, parking and negotiating intersections'.[30] The 11 vehicles that passed the National Qualification Event drove for 97 km through an urban environment interacting with other unmanned as well as manned vehicles (a total of 50) that were introduced into the circuit in order to increase traffic density.

DARPA challenges and other trials in more or less complex experimental sites have shown that automated cars are able to perform the three main functions they are designed for: perception, planning and control. As explained by Sebastian Thrun, former researcher in Artificial Intelligence at Stanford and head of the Stanford team that developed Stanley—the winning robot in the 2005 DARPA Grand Challenge—and currently employed at Google:

> Perception addresses the problem of mapping sensor data into internal beliefs and predictions about the environment. Planning addresses the problem of making driving decisions. Control then actuates the steering wheel, throttle, brake, and other vehicle controls.[31]

Despite the fast development of these functions in software architecture, the way to the public roads is still long, although governments are taking action to speed up develop-

28 The Commission communication of 15 February 2006 on the Intelligent Car Initiative—'Raising Awareness of ICT for Smarter, Safer and Cleaner Vehicles' COM(2006) 59 final (not published in the Official Journal)—highlights the potential of intelligent cars for the European Union and calls for the establishment of an Intelligent Car Initiative to support policies and innovation in this area. In May 2010 the Commission launched the Digital Agenda for Europe, tackling the issue of how digital technologies can help societies and policymakers address various challenges. One of the areas of application of the digital agenda concerns mobility, and intelligent cars have a prominent role in it. See http://ec.europa.eu/digital-agenda/en/node/76926 and http://ec.europa.eu/digital-agenda/en/node/76926 for a database featuring a variety of EC-funded projects developing ICT-based systems and services for the road.

29 See http://ec.europa.eu/transport/index_en.htm.

30 See http://archive.darpa.mil/grandchallenge.

31 Thrun (n 4) 100.

ments by allowing testing on public roads and looking into road law changes.[32] The automated car industry is currently facing several challenges. The Dutch Automated Vehicle Initiative (DAVI), which is developing tools to study automated driving on public roads (in real environments), has pointed out the need to address not only technical challenges (especially connected to sensing and reliable control) and traffic management challenges (related to vehicle-to-vehicle and vehicle-to-infrastructure communication, platooning and lane specific controls), but also legal challenges and challenges presented by human factors.[33] According to the DAVI team, 'human factors' concern human users' reactions to automated driving and the expected driving behaviour. The authors highlight the need to gather empirical evidence on how the automation of driving changes the role of the driver, who will no longer be a manual controller of the vehicle. Three types of legal challenges are identified. The first concerns the type of approval that automated cars require in order to be allowed to drive on European roads (and be sold in European markets); the second challenge is related to the need to legalise the shared driving responsibility between the driver and the vehicle; and the third challenge involves the need to reconsider current legislation concerning liability for material and immaterial damage in the case of partial or high automation.

Similarly, in a report on the social, ethical and legal issues involved in autonomous systems, the UK's Royal Academy of Engineering concludes:

> In order to avoid stalling a technology that could be of significant benefit, Government should engage in early consideration of regulatory policy so that such systems can be introduced. There is also need to consider changes needed to legal frameworks that establish culpability for accidents on the road. When a driver cedes control of their car to its own navigating or driving systems, or to a centralised system controlling speed and distance between cars, they also give away some responsibility. So where would that responsibility rest in the case of an accident? Who will regulate the technology employed and how will it be certified? Will insurance companies accept the risk?[34]

The different challenges pointed out in these reports fall into different categories introduced in the previous section and will be further explored below. In fact, as we will see, these regulatory challenges require responses such as regulation of the production of self-driving cars through law (approval types), the regulation of users' behaviour through code (shared responsibility), and regulation of the effects of robot behaviour through law (liability in case of damage). What seems to be missing here—and this refers to our fourth category—is an acknowledgement that robots should comply with existing legal codes, such as traffic law. This is a critical point that emerged in the 2007 DARPA

32 For instance, on 30 July 2014, the UK Business Secretary Vince Cable revealed that public road testing in the UK will be allowed from January 2015. See eg http://bbc.com/news/technology-28551069.

33 The goals of the initiative are published in a white paper available at http://davi.connekt.nl/pdf/white-paper-davi.pdf.

34 Royal Academy of Engineering, 'Autonomous Systems: Social, Legal and Ethical Issues' (2009) 19, www.raeng.org.uk/societygov/engineeringethics/pdf/Autonomous_Systems_Report_09.pdf.

Urban Challenge. On that occasion, the Stanford team's vehicle 'Junior' crossed the finish line first, followed one minute later by 'Boss', Carnegie Mellon's entry. However, DARPA highlights that the goal of the competition was not just to award the prize for speed, but also to evaluate whether the cars were able to follow Californian driving rules. After DARPA officials analysed each team's infractions of the California Driver Handbook, 'Boss' was awarded the first prize of $2 million.[35] This fourth type of regulation will be further discussed in section IV.4.

IV. TYPES OF ROBO-REGULATION

1. Regulating the Design and Production of Robots through Law

The design of robots is regulated by imposing norms and safety standards that have to be respected in order to guarantee that robots are non-harmful for users. In the context of manufacturing, there are standards and regulations pertaining to the construction of robots. Industrial robots are classified as 'machinery' and hence they fall within the scope of machinery regulation, such as the Machinery Directive 2006/42/EC in Europe. The aim of this directive is to provide for a healthy and safe environment for employees in industrial environments. These environments are relatively well structured. Machines, for instance, can be bolted to the floor and be surrounded by cages, thus limiting their reach. Risks introduced by machines are relatively easily managed in such controlled environments. The behaviour of industrial robots with respect to their human co-workers and others has to be, and can be, controlled. This may be achieved by passive measures, such as fences and emergency stop buttons, and by active means, such as sensors that signal to the robot that individuals are within reach of moving parts, forcing it to stop or at least not make dangerous movements in the direction of the human. The Machinery Directive and ISO standards (in particular ISO 10218 on industrial robots) provide guidance for the design and implementation of robots to meet the health and safety requirements.

The scene of robotics, however, is changing rapidly. The existing army of industrial robots is being supplemented by a variety of (personal) service robots (non-industrial robots). These range from care robots and companion robots to assist and keep the elderly company, all the way down to simple cleaning robots such as the Roomba. Due to this extension of functions and the operating environments of service robots, safety needs have changed and have become more complicated. The environments in which service robots operate are far less structured than those of industrial robots, and this will have to be reflected in health and safety standards. Instead of staying away from humans, service robots are meant to be near and even touch humans. They also need the capability to handle unexpected events and human movement rather than not hav-

35 See n 30.

ing to 'worry' about humans because they will (be forced to) stay at a safe distance.[36] Regulation relating to non-industrial robots lags behind regulation for industrial robots, although various specific areas and types of robots are covered by regulation at the EU level, such as the General Product Safety Directive 2001/95/EC and the Consumer Protection Directive 1999/44/EC, as well as at Member State level. An ISO standard for non-medical care robots is under development (ISO 13482).

What complicates matters further is the fact that safety in the consumer space depends not only on the behaviour of the machine, but also on that of the user, the consumer. Both robot and human user play a role in limiting risks, and given the flexibility of humans to adapt to unexpected events and changes in the environment compared to robots, their role is likely to remain important in the foreseeable future. To play their role properly, the user will have to be informed of the risks involved in using the robot and be technically and legally protected against any potentially harmful effects of the robot. These technical and legal protection measures will have to be regulated/implemented through other types of robot regulation, in particular through code embedded in robots (section IV.4) and regulating the effects of robot behaviour (section IV.3).

In the case of automated cars, the issue of regulating their design and production through law is clearly on the agenda as well. Even for road tests, but certainly for production vehicles, cars have to undergo testing by the road authorities. Two kinds of issues emerge here. First of all, a statistically significant number of on-road automated vehicle tests have to be done before the safety of these systems can be measured. This is not the case at the moment. Second, it is not clear how to test the safety of automated cars within existing regulation. For example, how should we test the safety of automated cars in the framework of Article 8 of the Vienna Convention, which states that a vehicle should always have a driver who is in control? Directive 2007/46/EC establishes a framework for the approval of motor vehicles, identifying 47 types of testing of passive safety (eg airbags) and active safety (for the prevention of accidents). Article 20 specifically concerns new technologies which are incompatible with one or more regulatory acts and enables Member States to grant provisional approval, valid only in their territory. In the case of the DAVI initiative, which saw the first automated vehicle driving on the Dutch highway on 12 November 2013, carrying the Dutch Minister of Transport as the 'driver', the RDW (the Dutch type-approval authority) had to find some ad hoc solution for testing the car. In this particular case, a derogation from the vehicle safety and traffic regulation was granted.[37]

36 We are anthropomorphising a little here. Industrial robots of course present no worry at all. Certain actions and events are simply not anticipated to be part of their normal behavioural repertoire and hence they have no behavioural responses for those (non-anticipated) events. On the other hand, should such a non-scripted event occur, the effects may be disastrous for the humans present, because the robot most likely will not respond in an appropriate way.

37 The derogation is very specific. It specifies which vehicle requirements are relaxed, but also on which roads (or sections of road) the vehicle is permitted, which manoeuvres it is permitted to make, and which 'drivers' are permitted to drive the vehicle.

2. Regulating User Behaviour through the Design of Robots

The behaviour of individuals (as the users of robots) is regulated not only by means of legal and social norms, but also by what technology permits and inhibits. The particular design of a robot can both limit what the user can do with it and limit potential harms caused by it. A very simple example of the latter could be a domestic (cleaning) robot shutting off all motors when lifted, thus limiting the risk of the user's fingers getting caught in places they should not go. The elicitation of behaviour through design includes a wide range of phenomena that for instance vary as to the *amount of choice* an individual has in ignoring the behaviour prescribed by the technology.[38] Going from almost total freedom we move from persuasive technologies,[39] through 'nudging'[40] and affordances,[41] to techno-regulation.[42]

Persuasion is prominent in what BJ Fogg has termed 'captology'.[43] Persuasive techniques are often used in ICTs to change people's attitudes and behaviour. An example is the use of suggestive technology, 'an interactive computing product that suggests a behaviour at the most appropriate time',[44] such as a Speed Monitoring Awareness Radar Trailer (SMART). SMART is a mobile speed detector that can be placed outside venues such as playgrounds to force motorists to reflect on their current speed and its appropriateness in the vicinity of playing children. The user is free to ignore the speed limit, but can hardly escape reconsidering their course of action. An example of persuasion in automated vehicles is the lane assist/lane departure feature as available in some cars, such as the Volvo V40. The car is capable of staying in its lane in certain conditions, and can also signal to the driver that the car is about to leave its lane through steering wheel vibration.[45]

Nudging, introduced by Richard Thaler and Cass Sunstein,[46] is related to captology, but has a slightly stronger normative connotation. All technology has behavioural effects and any design requires choices—hence designs are 'choice architectures'. Thaler and Sunstein claim that designers have a 'responsibility for organizing the context in which people make decisions';[47] they maintain that some choices are better than others; and

38 Bibi Van den Berg and Ronald Leenes, 'Abort, Retry, Fail: Scoping Techno-Regulation and Other Techno-Effects' in Mireille Hildebrandt and Jeanne Gaakeer (eds), *Human Law and Computer Law: Comparative Perspectives* (Springer, 2013) 74.

39 BJ Fogg, *Persuasive Technology: Using Computers to Change What We Think and Do* (Morgan Kaufmann, 2003).

40 Richard H Thaler and Cass R Sunstein, *Nudge: Improving Decisions about Health, Wealth, and Happiness* (Yale University Press, 2008).

41 Donald A Norman, *The Psychology of Everyday Things* (Basic Books, 1988) 9.

42 Brownsword (n 23) 203; Roger Brownsword, 'Code, Control, and Choice: Why East is East and West is West' (2005) 25 *Legal Studies* 1; Leenes (n 23).

43 See Fogg (n 39) 5.

44 *Ibid*, 41.

45 See eg www.euroncap.com/rewards/technologies/lane.aspx for an overview of status and limitations.

46 See Thaler and Sunstein (n 40) 3.

47 *Ibid*.

they believe that, by nudging (gently pushing) people in the (morally, environmentally, healthy, etc) right direction, better choices will be made. Their approach is one of 'libertarian paternalism'. Nudging may involve placing the desirable option at the top of a list, or preselecting the 'right' choice. People are free to behave differently, but cognitive psychology[48] and behavioural economics teach us that humans are not the rational actors economic theory assumes them to be; rather they are guided by heuristics and cognitive biases. Nudging can exploit these biases in behaviour through choice architectures.[49] In the context of autonomous vehicles one could think of turning on the 'lane departure warning' setting by default as an example of nudging.[50] If people have to take action to switch it off, many will likely refrain from doing so.

An even stronger form of hardcoding intended behaviour is captured in the notion of 'affordance'. This concept, which originated in animal psychology,[51] was popularised by Donald Norman[52] to describe how artefacts influence behaviour. By mounting a plate on a one-way opening glass door and a handle on the other side, designers can make immediately clear how people are to open the glass door.[53] The technical design affords certain actions and makes other actions more difficult, or even inhibits them. Not providing any controls on a robot would provide a clear signal that the user should not interfere with the robot.

Finally, there is techno-regulation. Techno-regulation stands out because it concerns norms that are intentionally incorporated into the technology.[54] Affordance may be an unintentional effect; the designer of the glass door may simply not have thought about the push plate, but just have picked it because it looked better to have it on the outside, which might just happen to be the side that can be pushed. Speed limiters in trucks are examples of techno-regulation proper. The prevailing speed limit for trucks can be hardcoded in the truck's system, preventing the driver from speeding.

It is important to realise that regulation by design covers the entire spectrum of techniques outlined above. It concerns not only *intentional* choices made to regulate users' behaviour, but also the unintentional ways in which artefacts elicit human behaviour.[55] This potentially leads to a complicated interplay between intentional and unintentional

48 eg Amos Tversky and Daniel Kahneman, 'The Framing of Decisions and the Psychology of Choice' (1982) 211 *Science* 453.

49 Bibi Van den Berg and Ronald Leenes, 'Keeping Up Appearances: Audience Segregation in Social Network Sites' in Serge Gutwirth, Yves Poullet, Paul de Hert and Ronald Leenes (eds), *Computers, Privacy and Data Protection: An Element of Choice* (Springer, 2011) ch 10, 211.

50 In fact, Volvo's lane departure warning system is activated by default when the car is started, according to this Volvo instruction video: www.youtube.com/watch?v=dvXhpxrKvb0 (accessed on 22 July 2014). It can be turned off manually.

51 James J Gibson, 'The Theory of Affordances' in Robert Shaw and John Bransford (eds), *Perceiving, Acting, and Knowing* (Lawrence Erlbaum Associates, 1977) 67.

52 Norman (n 41).

53 The world would be a better place if affordance were taught at design academy and schools of engineering.

54 See Leenes (n 23) 149.

55 Bibi Van den Berg, 'Robots as Tools for Techno-Regulation' (2013) 3 *Law, Innovation and Technology* 317.

elicited behaviour of technology users in which humans are both recipients and actors. Speaking of law of robots, for robots, in robots or through robots cannot be interpreted unless account is taken of the interaction between humans and robots.[56] The centrality of human–machine interaction implies that technology regulation should concern not only agents in the legal field, but also the environment in which the interaction takes place. Regulation by design should therefore take into account the fact that shaping the world by design means shaping not only products, but also environments (space). Pagallo thus shifts the focus to the ways in which the environment of human–robot interaction can be designed to encourage and change social behaviour. In a similar vein, Van den Berg writes that design influences users' capacity to interact and to establish an emotional relationship with carebots. 'The central aim is to generate discussion on the values we embed into machines, and the effects that such embedding may have in the settings in which they will be deployed.'[57]

Regulation of users' behaviour is imperative in automated cars. Automated car developers acknowledge that these vehicles put the human driver in a very unusual situation, wherein their role shifts from 'manual controller' to 'supervisor'.[58] This requires an adaptation of driving behaviour, which is important to monitor and eventually regulate (or nudge), in order to predict possible consequences due to the driver's lack of awareness or vigilance.[59] Manufacturers and projects have addressed the issue of human–machine interfaces (HMIs) in different ways.[60] In some cases a display, activated by haptic feedback, provides information and warning (BMW); in others an LED-matrix applied beneath the windshield is combined with an acoustic warning and feedback on the steering wheel (Daimler). These different designs aim to find a balance between keeping the driver aware of the driving process and avoiding an overload of information that may make the experience of driving unpleasant and stressful. However, technical specifications and functionalities differ and ultimately provide different experiences for the driver.

The choice of HMI is important because the way information is available to drivers affects the way they will feel and behave in the car, deciding to delegate decisions to the system or taking over. Explicit regulation may be inscribed in the design choices of HMIs. For example, if Article 8.5 of the Vienna Convention—'Every driver shall at all times be able to control his vehicle or to guide his animals'—remains unaltered, HMI

56 See Pagallo (n 13) 17.

57 Van den Berg (n 55) 332; Aimee Van Wynsberghe, 'Designing Robots for Care: Care Centred Value-Sensitive Design' (2013) 19 *Science and Engineering Ethics* 407.

58 Raymond Hoogendoorn, Bart van Arem and Serge Hoogendoorn, 'Automated Driving, Traffic Flow Efficiency, and Human Factors: Literature Review', Transportation Research Board Annual Meeting, Washington, DC, January 2014.

59 See also EU projects working on 'human factors' in automated driving: http://adaptation-itn.eu and https://sites.google.com/site/itnhfauto.

60 For impressions see Margriet van Schijndel-de Nooij *et al*, *Definition of Necessary Vehicle and Infrastructure Systems for Automated Driving*, SMART 2010/0064, Study Report, European Commission, Brussels, 2011.

interfaces will have to embed functions that monitor the user's attention to the street and act in case the user is perceived as distracted. This form of techno-regulation governs the robot user's behaviour.

3. Regulating the Behaviour of Robots through Law

The third type of regulation concerns the effects of robot behaviour. The distinction in law between legal acts and acts with legal consequences is relevant here.[61] Legal acts aim at having a particular legal effect, such as marriage or contract. Marriage does not exist without legislation enacting the institution of marriage. Similarly, the rules of contract constitute who or what can enter into contracts and which forms and consequences of contracts exist. Legal acts aim at establishing a legal state. Simply put, they change the law. Acts with legal consequences aim at establishing a change in the world and the law happens to associate legal consequences (as well) to this change. Tort, for example, is (usually) an unintended effect of someone's action. People appropriating the property of others do not intend to receive the label 'thief'—that is merely an effect of criminal law; instead they intend to obtain new goods at zero cost. In the field of law and robots, this distinction also comes to the fore. Regarding legal acts, there is a longstanding debate over whether autonomous agents should have legal personality or at least whether they can enter into contracts in another form.[62] But given the fact that robots are not legal subjects, they do not have legal rights and legal duties and they cannot perform legal acts. Robots are, from a legal point of view, treated as tools and it is always a human being that is legally responsible for the robot's actions and hence responsible for ensuring that they operate within the boundaries of the law. In respect of acts with legal consequences, discussions relate, for instance, to the question whether tort law, especially liability for things, is equipped to cope with highly autonomous robots.[63] The acts of robots may have effects that are, or need to be, dealt with by the law. When a robot causes damages, the plaintiff will seek compensation, and then the question is who (or what) is liable for the damage. Robots are currently treated as tools, no different from hammers and fridges,[64] and thus the owner will generally be liable.

This is not to say that the existing regulation is fully satisfactory;[65] it may, for instance, place liability on the wrong entity. If manufacturers are always liable for damage caused by correctly functioning robots that decide to cause certain damage in order to prevent more severe damage from happening (eg crashing into the cupboard in order

61 eg Dick WP Ruiter, *De vorm behouden: Verslag van een levenswerk* (PrintPartners Ipskamp, 2008) 14.

62 See eg Solum (n 18) 1231; Pagallo (n 13) 95.

63 For a reflection on this topic regarding automated cars see Maurice Schellekens, *Automated Cars: A Legal Analysis in Guidelines for RoboLaw*, internal report, RoboLaw project, 2014.

64 Luciano Floridi, 'Artificial Companions and their Philosophical Challenges' (2009) 19 *Dialogue and Universalism* 31.

65 Pagallo (n 13) 100.

to catch the falling baby), then this may inhibit the development of such robots. Higher levels of autonomous action of robots may challenge the assumptions on which existing legal domains are predicated. In those cases, the law must and will change—just as slaves, women and corporations have achieved legal subjecthood in the past.

The need to readjust the existing liability regime in order to account for the characteristics of automated vehicles is indeed acknowledged by actors in this area. With automated vehicles, responsibility for an accident or damage is not only with the driver because, depending on the level of automation, some or most controlling and sensing functions are delegated to the system. Thus if something goes wrong, liability is likely attributable to the manufacturer of the car. It has been argued that manufacturers' increased risk of liability and the consequent costs and damage to reputation may create a 'chilling effect' in the automated car industry that might delay the entry of these systems onto the market.[66] Automated cars may present new risks of injury and damage, due for example to imprecise sensing or control. However, they also increase safety by eliminating risks due to human error (because of tiredness or distraction, for example). According to Schellekens, in order not to hamper innovation and to allow society to benefit from automated cars, existing liability regulation needs to be readjusted and standardised across different countries. Traffic liability models attributing liability to the driver following an accident (as in Germany) may not be the most appropriate options in the case of automated cars. The Swedish model, where one's own insurance company pays one's own damages, offers a promising alternative, because it limits the number of tort liability cases. This model should however be integrated with specific measures to regulate insurance companies' interests.

4. Regulating the Behaviour of Robots through Code/Design

Design and technical instruments can be used to influence and determine robot behaviour in order to make it compliant with social and legal norms that are relevant in the environment in which the robot operates. This is quintessential Asimov law. As far-fetched and Sci-Fi as they sound to people with knowledge and expertise regarding the current state of the art in robotics, Asimov's laws capture an issue that is highly relevant for studies of law and robotics, namely the fact that robots, to behave properly and acceptably, have to comply with some norms. Life for humans is full of social and legal norms and it might be reasonable to require robots to comply with at least a minimal set of these if they are to enter our everyday lives and have complex interactions with humans. Such robots will not be mere tools that are under the constant control of their owners and can be switched on and off at any point, but will perform all sorts of tasks unsupervised and thus with some level of autonomy. If robots do not have to observe any of the rules that structure our lives, humans constantly have to pay attention when-

66 Schellekens (n 63).

ever an autonomous robot is near, which implies a significant burden that to some extent defeats the benefits of outsourcing tasks to robots. Hence, it may be desirable to have robots comply with some norms and have them behave like law-abiding citizens. For instance, it seems obvious that driverless vehicles should observe traffic laws[67] if they are to intermingle with human driven cars. It can be required that social robots should do their work quietly instead of drawing unnecessary attention by sirens and flashlights. Robot helpers in hospitals might need to clear the way when people approach, rather than move on in the middle of the hallway. Enforcing compliance with these social and legal norms will have to be included in the design of robots, sometimes through very technical solutions.

What is at stake here is that as (semi-)autonomous agents, robots are expected to act in the world, interact with human beings, and make decisions with some legal relevance. This may not be sufficient to attribute moral agency to them, but it raises questions concerning their legal agency.[68] Without getting into the question of whether we should attribute legal responsibility[69] and legal personhood[70] to robots, the question of regulating robot behaviour and ensuring that they comply with legal norms is quite relevant because robots will act with a high level of autonomy with limited or no human intervention.

In order to ensure that urban and social robots observe relevant regulation, the rules will have to be embedded in their 'positronic brain', to use Asimov's terminology. Or to use contemporary terms, the robot's behaviour will have to be regulated by design (including, or mainly, by code). However, if Asimov refers to laws *specific to robots* and embedded in their positronic brain, what we refer to here concerns robots able to comply with existing *human* laws and norms. In the case of autonomous vehicles, for example, it is acknowledged that this type of robot 'also needs to be aware of its surroundings, in order to obey local traffic laws and interact safely with other vehicles and pedestrians'.[71] Obeying traffic laws is a crucial point, but this seems only to receive passing mention and is not fully explored in discourses on the legal challenges of automated cars. Whereas technological designers use accurate and explicit models of physics, they seem to rely on implicit models of relevant traffic regulation.[72]

67 UK Parliamentary Office of Science & Technology note on Autonomous Road Vehicles, www.parliament.uk/briefing-papers/post-pn-443.pdf.

68 Luciano Floridi and JA Sanders, 'On the Morality of Artificial Agents' (2004) 14 *Minds and Machines* 349.

69 Andreas Matthias, 'The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata' (2004) 6 *Ethics and Information Technology* 175; Dante Marino and Guglielmo Tamburrini, 'Learning Robots and Human Responsibility' (2006) 6 *International Review of Information Ethics* 46.

70 David J Calverley, 'Imagining a Non-Biological Machine as a Legal Person' (2008) 22 *AI & Society* 523.

71 See Schellekens (n 63).

72 See eg Urmson *et al* (n 4) for an extensive account of the design and technology behind the Google car that participated in the 2006 DARPA challenge. The algorithms and technical constraints are discussed at length, whereas the traffic law related material is simplified into 'driving events' that the car should successfully master. The driving events seem to be based on designers' understanding of the traffic code without any reference to the underlying legal provisions.

This focus on the technological side of designing artefacts that are supposed to enter everyday life is quite common and systemic in technological design processes that tend to abstract from a rich model of the target socio-legal context. This is also understandable given the technical challenges; after all, what is the point of figuring out how the car should behave in urban traffic if the vehicle cannot even stay on the road? In designing autonomous vehicles able to navigate urban environments, researchers focus on several functions such as 'structured driving' on a road that requires the vehicle to 'robustly follow the desired lane', 'avoid[ance of] static and dynamic obstacles', as well as 'unstructured driving' in a parking lot that presents fewer navigation constraints but still requires the perception of obstacles and other vehicles.[73] Distance keeping, merge planning, localisation and intersection handling are other important functionalities that autonomous cars need to navigate real environments. Despite the need for abstraction, design processes acknowledge that these vehicles will operate in complex geographical and social environments.

Preparing highly autonomous vehicles for life on the road is commonly done by going through 'play books' that contain scenarios the car must be able to properly handle. The Google Boss team that participated in the DARPA challenge adopted a master playbook of 250 such 'driving events'.[74] These events describe concrete situations (eg passing a slow moving vehicle), its pre-conditions (eg vehicle to pass (V1) drives at less than half the speed of test vehicle) and the requirements for success (eg determine correct speed of V1 and pass without stopping). The driving events represent a set of mundane (correctly stopping at a stop sign) or challenging (successfully navigating a jammed intersection) situations.[75] They are meant to test the complex interplay of the various system components, rather than compliance with traffic regulation. The scenarios do assume a working model of traffic regulation though. For instance, the scenario regarding passing a slow driving vehicle presumes overtaking on the left and that a check is performed to ascertain whether V1 may be overtaken,[76] without making explicit what the underlying (road law) rules are and how these are modelled/represented in the system. It is therefore unclear whether the car is programmed to detect the appropriate scenario and act accordingly, or whether it actually operates on the basis of a more abstract model of the traffic regulation.

Mapping the intricate socio-legal environment into a set of scenarios that can be used to draft technical requirements and serve as a test suite contributes to managing the difficulty of developing complex systems such as highly automated vehicles, but also car-

73 See eg Urmson *et al* (n 4); S Kolski, D Ferguson, M Bellino and R Siegwart, 'Autonomous Driving in Structured and Unstructured Environments', IEEE Intelligent Vehicles Symposium (2006), 558; P Falcone, F Borrelli, J Asgari, HE Tseng and D Hrovat, 'Predictive Active Steering Control for Autonomous Vehicle Systems' (2007) 15(3) *IEEE Transactions on Control Systems Technology* 566.

74 See Urmson *et al* (n 4) 457.

75 *Ibid*.

76 The diagram in Urmson *et al* (n 4) 457 contains an arrow depicting the intended path the vehicle takes, which shows overtaking vehicle V1 on the left, crossing a dotted line.

ries the risks inherent in all modelling. It is difficult to assess whether the scenarios fully capture the intended behaviour and the models may turn out to be over-simplifications. In an environment where legal rules exist, the question arises as to whether cases (scenarios) should be used as an intermediate step to implement the intended behaviour into the system, or whether instead the legal rules themselves should serve as the foundation. We will return to this point in the following section. In any case, as pointed out in the DAVI white paper, 'the relationship between traffic flow efficiency and automation of vehicles is far more complex than assumed in these theoretical considerations and modelling studies'.[77] This shows that the developers are aware of the limitations of their current approach.

The fact that the automated vehicle will have to obey traffic laws is a consideration that should be taken into account at an earlier rather than later stage of technological development. An incremental approach can be adopted in which development moves from the mundane to the challenging and from frequent to rare events, but this may create path dependencies that are particularly hard to fix later on and that may require significant software re-engineering.

Traffic regulation contains a limited set of rules that aim for the most part at promoting traffic safety and to a lesser extent at civilised behaviour. The basic rules, such as 'adhere to speed limits', 'overtake on the left', 'do not block intersection' etc, seem fairly straightforward, but there are also many subtle exceptions to rules that require situational awareness and may require (value) judgements. Implementing these into highly autonomous vehicles requires appropriate sensors, but also appropriate algorithms and reasoning systems. Below we will discuss some examples of the kinds of rules and their implications that driverless vehicles should comply with.

Consider, for example, Article 16 of the Dutch Road Traffic Signs and Regulations Act 1990 (RVV1990), which provides that 'Road users must not cut across military columns and motorised funeral processions'. For humans, this is not a complex rule. Most people will not have a problem recognising one of these special convoys when they see one, irrespective of whether they carry the appropriate flags (in the case of military columns at least). But if we look at Boss, the automated vehicle that won the 2007 DARPA Urban Challenge, this rule would be quite a challenge because in the system 'no explicit vehicle classification is performed. The tracking system provides information only about the movement state of object hypotheses.'[78] Boss would simply not recognise an individual military vehicle, let alone a military convoy. While including vehicle classification might not have been necessary to complete the Urban Challenge tasks, it would have contributed to making the system's behaviour more compliant with existing traffic regulation. Complying with Article 16 RVV 1990 is not an insurmountable obstacle,[79] but

77 DAVI white paper (n 33) 3.

78 See Urmson *et al* (n 4) 434.

79 Vehicle type detection can be implemented in different ways, including vehicle-to-vehicle communication, which involves designated vehicles in special columns signalling their status to neighbouring cars.

taking such rules into account in systems design requires sensors, software constructs, rules, algorithms, etc.

Observing traffic regulation not only requires situational awareness, that is, the ability to detect vehicles, obstacles and humans, but also explicit or implicit knowledge about the traffic code itself. For instance, when multiple vehicles approach an intersection, being aware of the various vehicles is only one aspect. Determining which one takes precedence is another. Priority is determined by traffic law, and hence the vehicle must possess the knowledge to determine when it is its turn to cross the junction. Priority is determined by general rules (Article 15 para 1 RVV 1990: 'At road junctions, drivers must give priority to traffic approaching from the right') and exceptions (such as Article 15 para 2: 'a. drivers on unpaved roads must give priority to drivers on paved roads; b. all drivers must give priority to tram drivers'). Giving priority at real-world road junctions will have to be based on traffic law and must therefore be embedded in the vehicle's system. This can be done through some form of explicit representation of (traffic) rules, which can then be applied by some reasoning engine that determines the applicable rules and within this set resolves potential rule conflicts (see below) and makes decisions based on the (hopefully one) remaining rule. Other methods of implementation which do not involve reasoning with rules are conceivable as well, such as hardcoding the rules in such a way that potential rule conflicts are resolved during software design, rather than on the fly.

Given the fact that traffic laws occasionally change and differ between jurisdictions, ensuring that the vehicle system can adapt to different rules is a design requirement that also needs to be accounted for at an early design stage. If cars are to be able to cross borders, they have to be able to cope with different rule sets. The differences between left-hand driving countries (such as the UK and Japan) and right-hand driving countries clearly show that some adaptation in behaviour is required. Needless to say that simply programming vehicles to remain on the right hand side of what is detected as a road does not lead to the desired behaviour in all jurisdictions.

Another issue in law concerns rule conflicts. The law does not constitute a logically coherent and consistent set of rules amenable to machine execution.[80] The rules may give rise to conflicting duties in concrete cases. Consider, for instance, Article 14 RVV 1990, 'Drivers must not block road junctions', alongside Article 82 para 1, 'Road users are required to follow instructions given verbally or by means of gestures by: authorised officials who are identifiable as such …'. These two rules can cause problems for a vehicle that enters a junction and then is made to stop by a police officer. The vehicle cannot

80 See eg Bert-Jaap Koops and Ronald Leenes, 'Privacy Regulation Cannot be Hardcoded: A Critical Comment on the "Privacy By Design" Provision in Data-Protection Law' (2014) 28 *International Review of Law, Computers & Technology* 159; Bert-Jaap Koops, 'The (In)flexibility of Techno-Regulation and the Case of Purpose-Binding' (2012) 5 *Legisprudence* 171. See also Ugo Pagallo, 'On the Principle of Privacy by Design and its Limits: Technology, Ethics and the Rule of Law' in Serge Gutwirth, Ronald Leenes, Paul de Hert and Yves Poullet (eds), *European Data Protection: In Good Health?* (Springer, 2012) on the implementation of data protection regulation in machine executable form.

comply with both rules at the same time. In this case, the conflict is easily resolved by one of the conflict resolution rules in the RVV 1990, which states that signals by authorised officials take priority over signals and rules. But more subtle conflicts of duty may also arise. In law, rule conflicts are resolved by various implicit/general (*lex specialis derogat legi generali*) and explicit conflict resolution rules. Humans generally cope reasonably well with these rule conflicts, but things are different for computers.[81]

Many situations in ordinary traffic require deliberation and value judgements, which will be based on an interpretation of the meaning and purpose of traffic regulation. For instance, Article 19 RVV 1990 states: '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.' This provision clearly aims at promoting road safety by determining that drivers should at all times be capable of stopping their vehicle without crashing into anything in front of them. As such the provision outlines a condition that needs to be taken into account: visibility. Drivers should adapt their speed to the visibility conditions. Doing so may overrule the prevailing speed limit at a certain location, making the maximum speed conditional on other factors, such as visibility. In foggy conditions, cars are not supposed to drive at maximum speed. But there are many other situations that warrant the driver lowering their speed. For a highly automated vehicle this rule needs to be operationalised in such a manner that sensors, actuators and software give effect to its intended meaning.[82]

Decision-making based on social cues may also be required. Handling traffic queues is a clear example. The rules for merging into a queue are relatively simple, but in actual practice

> human driving is a social activity consisting of many subtle and some not-so-subtle cues. Drivers will indicate their willingness for other vehicles to change lanes by varying their speed and the gap between themselves and another vehicle, by small amounts. At other times it is necessary to interpret hand gestures and eye contact in situations when the normal rules of the road are violated or need to be violated for traffic to flow smoothly and efficiently. For autonomous vehicles to seamlessly integrate into our society, they would need to be able to interpret these gestures.[83]

Maybe this is where the fully autonomous car in mixed driverless/driven car scenarios reaches its limits. Getting the fully autonomous vehicle to correctly interpret[84] these subtle social cues may be extremely hard. This may have inspired Urmson *et al* to con-

81 Much work in the field of Artificial Intelligence and law has focused on logics to deal with such rule conflicts and non-monotonic reasoning. See Bench-Capon *et al* (n 11); Jaap Hage, *Reasoning with Rules: An Essay on Legal Reasoning and its Underlying Logic* (Kluwer, 1997).

82 An interesting complication in terms of man–machine symbiosis is that safe distance and speed are different for a highly automated vehicle and a human driven car. The automated vehicle may be equipped with infra-red sensors and have a reaction time well below that of human drivers. How will this affect the perception of human drivers?

83 See Urmson *et al* (n 4) 465.

84 The numerous fights that appear to take place between human drivers shows that humans also have problems interpreting signals.

clude that 'perhaps it will be sufficient and easier to assume that we humans will adapt to robotic conventions of driving rather than the other way around'.[85] That, however, may be undesirable. If the promise of automated cars has to be met, we should be realistic and not expect robots to act as slaves, but also not to expect humans to blindly adapt to robots. When sensors and automation are not the best means to guide robot behaviour and rule compliance, the human controller may provide assistance in order to reduce the system's complexity. This requires a holistic approach wherein designers conceive of cooperating human–robot systems wherein responsibilities for rule-compliance are shared.[86]

V. LEARNING THE ROPES: NORMS AND LAWS THROUGH DESIGN

This analytical distinction between four types of robot regulation serves the objectives of (1) taking seriously the idea of 'law of robotics', showing its concrete and timely aspects that cannot be dismissed as purely Sci-Fi dreams, (2) offering a systematic way of dealing with issues of robot regulation without clustering them together, and (3) highlighting the fact that robots' compliance with normative rules is crucial to their functioning in complex social environments and does not necessarily require articulate moral reasoning.

As shown in section IV, the developers of automated vehicles deal with regulatory issues of the four types we have distinguished. Respect for safety standards, human–machine interaction and regulation of the driver's behaviour, and the regulation of the effects of robots in human transactions, such as in the case of liability, are key issues that have not yet been resolved. The fourth type of robot regulation concerns regulation of a robot's behaviour according to various existing legal norms through technical design. This is an issue that has already been addressed by technology developers, but only to a certain extent. Giving a car the ability to respect the speed limit and to sense existing infrastructures or speed limits is not difficult given the state of the art.[87] Things become more complicated when lateral control systems are involved, as in the case of traffic merging, overtaking and lane changing. As often happens in robot development, difficulties arise with the increasing complexity of the infrastructures, human behaviour and social/legal norms in which the machine has to operate. How can the vehicle determine priorities at a roundabout or junction? How can we 'design in' the ability to determine when it is desirable to break a norm (eg exceeding the speed limit) in a situation of danger (eg an obstacle on the road during overtaking)?

85 See Urmson *et al* (n 4) 465.

86 Peter Asaro, 'What Should We Want From a Robotic Ethic?' (2006) 6 *International Review of Information Ethics* 9.

87 Several cars, such as the Volvo V40, include technology that informs the user about the prevailing speed limit based on interpreting traffic signs (Road Sign Information system). This information could easily be used to actually control the speed of the vehicle.

Designing a robot that complies with existing normative and legal frameworks is not easy for technology developers, who may rather opt for re-creating an experimental set-up in real-world situations and shifting the burden onto human users. The case of DustBot is telling in this respect. The DustBot project,[88] the aim of which was to develop an automated machine for urban trash collection, included an experimental phase, which included use in a real operative scenario in a small medieval village in Tuscany. The laboratory experimental environment was re-created in the real-life trial. In the village of Peccioli a priority lane for the robot was created and citizens were notified of the presence of the robot. They had to pay attention and stop if the robot was coming. As yet, however, there has been no attempt to implement the ability to respect the traffic code in the robot. Although traffic rules can be expected to be a fixed and stable element of the environment in which the robot will operate, they are not addressed by engineers who are working on demonstrating the 'general feasibility, scientific/technical effectiveness and social/legal plausibility and acceptability by end-users' of these robotic services.[89] When asked,[90] engineers said that this is something to be addressed in a further experimental phase. They also envisage possible integrated systems of sensors that allow, for example, the robot to communicate with a smart traffic light that instructs it to stop when required. Embedding priority rules at a crossroads is however considered more challenging from an engineering point of view because it involves some hierarchical rule reasoning of the machine.

To summarise, our fourfold analysis has raised the following points. First, regulating robots is not only a matter of reflecting on the regulatory challenges of a robot's use and production; it also necessitates exploring how robot behaviour can comply with existing legal (and social) norms that currently regulate human behaviour in the space in which the robots operate.

Secondly, the types of legal challenges that must be met by robots in everyday life are often under-determined and require a thorough study of the legal landscape, as we have seen in the case of automated cars. It is particularly important to take the legal landscape into account in the early phases of robot development and design in order to avoid having to incorporate fundamental design changes later on. As we saw in the case of embedding recognition of different kinds of vehicles, a closer look at the existing legal provisions in traffic regulation provides researchers and manufacturers with a much richer set of requirements that ultimately need to be met and that are often dismissed in scenario-based tests and simulations.

88 eg Gabriele Ferri, Alessandro Manzi, Pericle Salvini, Barbara Mazzolai, Cecilia Laschi and Paolo Dario, 'DustCart, an Autonomous Robot for Door-to-Door Garbage Collection: From DustBot Project to the Experimentation in the Small Town of Peccioli' [2011] *IEEE International Conference on Robotics and Automation* 655.

89 F Cavallo, M Aquilano, MC Carozza and P Dario, 'Implementation of 3D Services for "Ageing Well" Applications: Robot-Era Project', Forfitaal: Ambient Assisted Living IV Forum Italiano, Parma, 18–20 June 2012.

90 Personal communication between one of the authors and the developers.

Thirdly, not every rule in legal codes can easily be translated into technical solutions. Legal provisions require interpretation in view of their purposes and other factors, such as case law. As pointed out above, Article 19 of the Dutch RVV 1990 states that drivers should be able to stop their vehicle. For the purposes of this rule, it is crucial to determine what 'safety' is. Human drivers are not given a definition and they have to come up with their interpretation of what is a safe situation in a certain context. They make mistakes sometimes and of course a machine can exhibit more standard behaviour. But to what extent is it better? It depends not only on contextual factors, but also on the interpretation of safety, which is encoded in the design. A discussion about different interpretations of normative values is crucial, in order to ensure that robots display desirable behaviours. Obviously there are also legal provisions that should not be implemented because they are meaningless in the robot. For instance, Article 61a RVV 1990 states that 'A person driving a motor vehicle, moped, motor-assisted bicycle or disabled person's vehicle equipped with an engine may not hold a mobile phone while driving'. In the case of driverless cars the norm no longer has an addressee. In the case of highly autonomous vehicles, where a human is (still) considered to be in control of the vehicle (for emergencies), whereas the car in fact has control, the 'driver' is no different from the passenger, so why should they not be permitted to use their mobile phone?

Fourthly, the interpretive flexibility of legal norms also emerges in cases where some norms are conflicting. As the case of rules on behaviour at traffic junctions shows, in some situations, we are supposed to leave the junction clear, while in others it is desirable to stop for safety reasons. The assessment of context is pivotal in order to determine the correct behaviour according to the rule. Automating this flexibility in rule compliance is difficult, just as it is difficult to automate social activities and non-written rules that are embedded in drivers' practices—as in the case of establishing eye contact with other drivers at a junction.[91]

Fifthly, in regulating robots through code, it is important to note that social and normative practices and technology co-evolve. Not only should social practices in driving behaviour be taken into account in regulating robot behaviour through code; designers should also take into account the fact that driving practices are shaped by technology. This mutual shaping between technology (which should adapt to human behaviour) and the user (who adapts her practices to the technology) may have to be included in the robot's design. Rather than focusing on automating behaviour and communicating this to the user, it is important to regard the user as part of the system. Regulation is therefore regulation of a hybrid system comprising a robot and a human being in which responsibilities are distributed.

Finally, it is important to note that technologies are not only shaped by the current normative and legal framework, they also affect it. As we discussed in the case of rules

91 This point resonates with the arguments expressed by ethicist Patrick Lin concerning the need for a human in the control loop who can make difficult decisions that imply moral reasoning. See Lin (n 3).

that oblige drivers to pay attention to the road (by forbidding them to use a mobile phone while driving, for example), automated cars will not only have to be regulated, they will also change existing regulation. More generally, robotic technologies may require societies to change the law.

VI. CONCEPTUAL, TECHNICAL AND NORMATIVE ISSUES

The regulation of robot behaviour through design raises a number of conceptual, technical and normative issues. First of all, it is legitimate to ask whether designing robots able to comply with legal norms is a matter of regulation or a matter of functional design. After all, stopping at a traffic light is quintessential to the functioning of an urban robot as much as its ability to cover a distance or navigate the environment. How does regulation come into play? In order to address this question it is helpful to engage in some conceptual distinctions.[92] Technological artefacts have a 'utility value', that is, 'the value of being useful for a certain end, whether that end is good or bad'.[93] The quality of the artefact (whether it is a good/poor hammer, for example) depends on how well the artefact fulfils the specific function that it is supposed to achieve. During the design process, the artefact's functional requirements 'are translated into technical specifications which are embodied in a certain physical structure'.[94] Functional design requirements are based on the intended utility value of the artefact. As Van de Poel explains, in the case of a pencil, a functional requirement may be that it should be easy to hold. This functional requirement is translated into technical specifications, like the length of the pencil, its thickness and its shape. The designer plays a role in translating functional requirements into concrete technical specifications. Additional design requirements are not based on the function but on other factors, such as the preferences of different stakeholders with their individual desires, needs and interests. Additional design requirements, also known as non-functional requirements,[95] can also be dictated by larger socio-technical systems in which the artefact is supposed to be embedded (the legal system represents part of these socio-technical systems).[96] They can also depend on moral considerations. Moral values that play a role in determining technical specifications include safety, human health, privacy, justice and sustainability. These values are often too vague to be used directly in the design process and need to be translated. In designing and building artefacts such as robots, a heterogeneous set of requirements needs to be taken into account.

92 See Van de Poel (n 6) 6.

93 *Ibid*.

94 See Van de Poel (n 6) 15.

95 eg Martin Glinz, 'On Non-Functional Requirements' [2007] *15th IEEE International Requirements Engineering Conference* 21; Daniel Gross and Eric Yu, 'From Non-Functional Requirements to Design through Patterns' (2001) 6 *Requirements Engineering* 18.

96 Glinz (n 96) 25 shares these under the heading of 'specific quality requirements', which are requirements that pertain to a quality concern other than the quality of meeting the functional requirements.

Engineers, by their nature and expertise, tend to focus on functional requirements, which seem to be at the core of the artefact. The distinction between functional and non-functional requirements, however, is less sharp than it seems at first glance and depends on how the requirement is expressed.[97] The example of an urban robot obeying the requirement to stop at a red traffic light is a case in point here. The robot may not maintain its structural integrity for long if it ignores red traffic lights, and hence this requirement could be considered a functional one. But one can equally argue that stopping at the red light is simply a legal requirement. In this case the goals of both functional and legal requirements coincide, making it difficult to classify the requirement of stopping at a red light. It should however be treated as a legal requirement because the regulation instituting the red light as a traffic signal may change, requiring that the robot behave differently. Because of the (potentially) dynamic nature of legal requirements it does make sense to treat them as a distinct category. Unless there is specific and explicit attention to the full set of legal requirements, we expect designers to take into account in their designs only those legal requirements that have a direct effect on the functioning of the robot. Taking into account all relevant requirements requires a specific type of expertise. Just as implementing privacy by design requires a good understanding of the 'state-of-the-art research in security and privacy technologies, legal frameworks and the current privacy and surveillance discourses',[98] engineering in the domain of robots requires knowledge about the specifics of the technology, relevant legal frameworks and the socio-technical context in which the robots will be deployed. Our experience in large European projects[99] has taught us that taking into account socio-legal-ethical requirements is far from standard practice in complex state-of-the-art engineering projects.[100] And the same applies to the interaction between the various requirements—as Seda Gürses et al note, 'The interaction between legal and technical requirements remains an under-researched field.'[101] For robots to achieve their full utility, they will have to implement the full set of heterogeneous functional, safety, legal, social and ethical requirements.

A second issue concerns the very idea of embedding (legal or social) rules in machine executable code. As Bert-Jaap Koops and Ronald Leenes note in relation to attempts to inscribe privacy regulations in software, 'the idea of encoding legal norms at the start of information processing systems is at odds with the dynamic and fluid nature of many legal norms, which need a breathing space that is typically not something that can be

97 *Ibid*, 24. This is apparent from the two ways we framed the obligation of the urban robot to stop at the red traffic light.

98 Seda Gürses, Carmen Troncoso and Claudia Diaz, 'Engineering Privacy by Design', International Conference on Privacy and Data Protection, Brussels, January 2011, https://www.cosic.esat.kuleuven.be/publications/article-1542.pdf.

99 Both as researchers and as evaluators.

100 It usually takes scholars outside the technical disciplines to point out this omission.

101 Seda Gürses, Magali Seguran and Nicola Zannone, 'Requirements Engineering within a Large-Scale Security-Oriented Research Project: Lessons Learned' (2013) 18 *Requirements Engineering* 43.

embedded in software'.[102] As Koops suggests elsewhere,[103] there is a clash between the inflexibility of code and flexibility and openness to interpretation of the law, which often requires creative reasoning. Can rules embedded in machines cover several situations and be interpreted flexibly in light of changes in circumstances? In examining a case study of implementing the purpose specification principle (for data collection) in software, he points out the difficulty involved in

1. defining the norm which is often ambiguous and subject to interpretation;
2. transforming the legal norm into a technical framework; and
3. deploying the system because of conflicting interests among stakeholders.[104]

Koops therefore concludes that 'if the norm is representationally complex, be it due to openness, fuzziness, contextual complexity or to regulatory turbulence', techno-regulation is difficult to implement and should be avoided. Koops discusses the issues of implementing regulation in computer code in the context of personal data processing and having the system enforce the purpose limitation principle. In the case he discusses, there is no strict need to implement the regulation in code, because data protection compliance can be achieved in other ways. The systems in question are part of a larger socio-organisational system where humans and machines together form the 'data processing system'. Humans determine how the data in the system will be used (which may be implemented in software, but also in operational procedures and manual processes) and are therefore in the loop, even though the IT system may function without human intervention after design (for instance, there will be little human involvement in an online store that distributes digital content). In case of highly autonomous robots, the system will have to be able to guarantee legal (and ethical) compliance on its own. We therefore, unlike in the cases discussed by Koops and Leenes, have a Catch-22 scenario: we cannot do without programmed legal compliance and we cannot (easily) implement programmed legal compliance. Somehow this dilemma has to be overcome to achieve full utility of robots in our everyday lives. Difficult though this may be, keeping track of where and how legal requirements are implemented in the robot, certainly in the case of driverless and highly autonomous vehicles, is important.[105] This is not only because the

102 See Koops and Leenes (n 80) 167.

103 See Koops (n 80) 171.

104 *Ibid*.

105 One approach is to represent the relevant legal knowledge in a formal language that can be processed by the robot. Depending on the level of congruency between represented legal knowledge and legal source, this may take the form of isomorphic representation (Trevor Bench-Capon and Frans Coenen, 'Isomorphism in Legal Knowledge Based Systems' (1992) 1 *Artificial Intelligence and Law* 65) or transparent representation (Travis Breaux, 'Legal Requirements Acquisition for the Specification of Legally Compliant Information Systems', unpublished PhD thesis, 2009, 12–13).

underlying legal provisions may change, but also because the robot may have to operate in different jurisdictions and therefore will need to be able to apply the correct rules.

A third issue concerns the desirability of this kind of robot regulation. In the automated vehicles case (as well as the DustBot example), there is a strong interaction between a robot's ability to act in a social environment according to various legal/social norms and the construction of smart environments that intentionally regulate robot behaviour and unintentionally elicit human behaviour. Robots' sensing systems can be improved by designing smart environments that send specific information and retrieve relevant data from the physical surroundings. For example, road infrastructures can retrieve information about current traffic and send it to vehicles in order for them to adjust their behaviour. This raises new legal questions and ethical questions concerning the desirability of such highly morally designed environments.[106] David Smith,[107] for instance, has pointed out the potential moralising and de-moralising effect of design-based instruments. As an example, consider the waist-high ticket barriers at the entrances to the London Underground. These barriers do not make passage without a valid ticket impossible; people can jump over a barrier relatively easily, but doing so is a flagrant transgression of the norm. Jumping over a barrier dramatises the choice between morality and deviance. Other ticket barriers, such as some in Paris and Brussels, make entering the Metro without a valid ticket impossible by means of a tall tourniquet. Smith argues that a ticket system that is hard or impossible to bypass and removes all personal choice may lead to weakening self-controls and have a de-moralising effect. Brownsword takes this further and argues that in a moral community people act in line with norms not only because they have a moral obligation to do so (or are obliged by the architecture design), but also (especially) because they subscribe to the norm: ideal-typically, the fully moral action will involve an agent doing the right things (as a matter of act morality) for the right reasons (as a matter of agent morality).[108] A driverless car undermines this moral reflection. The car will always strictly observe the rules. Brownsword fears that this lack of moral reflection and moral choice will ultimately undermine the moral community and the legitimacy of the rules. This concern can also be observed in Koops's account of the acceptability of normative technology: '[T]he reduction of *ought* or *ought not* to *can* or *cannot* threatens the flexibility and human interpretation of norms that are fundamental elements of law in practice.'[109] These concerns are all very valid from the

106 See eg Mireille Hildebrandt, 'Law at a Crossroads: Losing the Thread of Regaining Control? The Collapse of Distance in Real Time Computing' in Morag Goodwin, Bert-Jaap Koops and Ronald Leenes (eds), *Dimensions of Technology Regulation* (Wolf Legal Publishers, 2010) 165, pointing to the decline of the rule of law in case of far-reaching techno-regulation.

107 David J Smith, 'Changing Situations and Changing People' in Andrew Von Hirsch, David Garland and Alison Wakefield (eds), *Ethical and Social Perspectives on Situational Crime Prevention* (Hart Publishing, 2000).

108 Brownsword, 'Code, Control, and Choice' (n 42) 17.

109 Bert-Jaap Koops, 'Criteria for Normative Technology: The Acceptability of Code as Law in Light of Democratic and Constitutional Values' in Brownsword and Yeung (n 16) 158.

perspective of maintaining a moral community and the rule of law, but again, if we want to achieve full utility of highly autonomous robots and vehicles we will have to shift legal compliance towards these artefacts, which does require programming them to observe legal and social norms.

VII. CONCLUSION

In distinguishing between different meanings of 'robo-regulation', this article has proposed and analysed a technically and socially plausible meaning of Asimov's idea of 'embedding rules in the positronic brain of robots'. Beyond Sci-Fi dreams, we have discussed existing attempts and failures to embed legal norm compliance into the design of robots. In this respect, dealing with regulatory issues raised by robots also encompasses the techno-regulation of robot behaviour by inscribing particular rules in their design.

Asimov's second law of robotics states that 'robots must obey any order given to them by human beings'. However, we should ask whether we really want to give orders to robots all the time. Norms can be considered to be part of the environment in which the robot will operate. Why, in experimentation with robots outside the lab, are simple legal norms such as those in traffic codes not considered a static element of the model to be implemented in the robot? This paper brings the relevance of the question to the fore.

What is clear is that robots will embed values and challenge the existing ones, but they will also have to be designed in such a way that they respect existing values and comply with a set of norms. As pointed out above, there are many technical hurdles to be faced when attempting to implement legal and social norms in robots. This is especially true if we take the broader ethical considerations into account. As Mireille Hildebrandt writes in relation to Ambient Intelligence:

> To sustain the distantiation inherent in modern law legal protection will need to be re-articulated into the digital infrastructures that will soon regulate our daily life, our critical infrastructures, our educational and professional settings, healthcare, defence etc. To ensure the distance needed for hesitation and contestation as preconditions for the legal protection of privacy, non-discrimination and due process this distance will have to be introduced into the emerging smart, real time environments. To succeed, the distance will have to be a constitutive part of the smart infrastructure, without losing the thread between law and the digitally native citizens it aims to protect. As discussed in other work, designing legal protection into the Ambient Intelligent environment (Ambient Law) will require creative cooperation between lawyers, legislators, computer scientists and citizens.[110]

This call applies equally to developing and managing the robots that enter our lives.

110 Hildebrandt (n 106) 183.