Ethics and Self-driving Cars: A White Paper on Responsible Innovation in Automated Driving Systems

Santoni De Sio, Filippo

Publication date
2016

Citation (APA)
Santoni De Sio, F. (2016). *Ethics and Self-driving Cars: A White Paper on Responsible Innovation in Automated Driving Systems*.

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

ETHICS AND SELF-DRIVING CARS
A WHITE PAPER ON RESPONSIBLE INNOVATION IN AUTOMATED DRIVING SYSTEMS

Filippo Santoni de Sio*

* Department Values, Technology and Innovation, section Philosophy/Ethics of Technology, Delft University of Technology. Some ideas contained in this paper have been presented and discussed at a colloquium on the ethics of Automated Driving Systems organized by The *Dutch Automated Vehicle Initiative* (DAVI) at Connekt in Delft on March 15th 2016. The author is grateful to the professionals who took part in the meeting for their insights and suggestions.

EXECUTIVE SUMMARY

In the context of the knowledge agenda automated driving (knowledgeagenda.connekt.nl/engels/),Rijkswaterstaat commissioned TU Delft to write a white paper on ethical issues in automated driving to provide a basis for discussion and some recommendations on how to take into account this subject when deploying automated vehicles.

In this paper I present, discuss, and offer some recommendations on some major ethical issues presented by the introduction on the public road of automated driving systems (ADS), aka self-driving cars. The recommended methodology is that of Responsible Innovation and Value-Sensitive Design. The concept of “meaningful human control” is introduced and proposed as a basis for a policy approach which prevents morally unacceptable risks for human safety, and anticipates issues of moral and legal responsibility for accidents. The importance of the individual rights to safety, access to mobility and privacy is highlighted too.

CONTENTS

INTRODUCTION

THE NECESSITY OF AN INTEGRATED ANALYSIS

In the recent "Nota" titled "Bezinning op ethische vraagstukken rondom de zelfrijdende auto" prepared by the Dutch Ministry of Infrastructure and the Environment, the ethical issues raised by the introduction of Automated Driving Systems (henceforth ADS) were divided into **three levels**: a) **operational**: how *automated vehicles* should (be programmed to) behave under different circumstances (for instance in the event of an accident); b) **tactical**: how the *road traffic* should be regulated given the presence of automated vehicles (for instance the interaction between automated and traditional vehicles); c) **strategic**: how should the broader societal impact of automated driving system (for instance on individual autonomy, privacy, distributive justice) be anticipated, guided, and/or regulated.

By endorsing and following up on that Nota, in this paper I map, present and discuss ethical issues pertaining to all of these three levels. However, I argue that **this division of levels may *not* be used as a sequential timeline** for the societal and political action: we should *not* first analyze the ethical dimension of the vehicle, *then* the traffic implication and finally the broader ethical and societal impact. The three dimensions are practically intertwined and have to be studied at the same time. In fact, the gradual switch to autonomous driving may represent a radical change in transport systems and in society more broadly. Therefore, as many ethical implications as possible of this switch should be assessed from the beginning: and this comprehensive ethical assessment should in turn form the basis of a consistent policy covering the three level: operational, tactical and strategic. If we are not clear about the ethical values, opportunities and risks involved in this switch in general, we can hardly properly assess the desirability of any specific technical or institutional solution, even at the "simple" operational level.

METHODOLOGY: RESPONSIBLE INNOVATION AND VALUE-SENSITIVE DESIGN

The "**Responsible Innovation**" approach in ethics of technology (van den Hoven 2013) relies on the idea that **ethics needs to be pro-active** to make a societal difference. We must prevent a situation where there is a disconnect between abstract moral discussions and the real world of engineering and policy. The basic idea is that ethical constraints and aims should become the very shapers of innovations at the relevant point in time and a place, where they still can make a difference, instead of fuelling political and academic discussions after the fact. Therefore in order to realize the ideal of Responsible Innovation two important requirements have to be met: a) **interdisciplinarity**: engineers, philosophers, policy-makers and stakeholders should actively work together to guide the future development of technology, and b) **anticipatory analysis**: this collaboration should happen from the early stages of the technological design process, when design options are still open, and the development of technology can still be steered.

Connected to the ideal of Responsible Innovation is the concept of **Value Sensitive Design** or Design for Values (Friedman 1996, van den Hoven 2007). If ethics wants to make a difference in the real world of technology it needs to take a *design stance* and use moral considerations and **values as requirements for the design of technologies**. In order to achieve this goal, next to the two requirement mentioned above (interdisciplinarity and anticipatory analysis), value-sensitive design contains two additional challenges: a) **systematic ethical analysis**: we have to clearly identifying as many relevant values as possible connected to the development of a given technology, and b) we have to **embed values in design**: translating these values into concrete design guidelines, where the term design includes both technical design (engineering) and institutional design (law and politics) (in what follows the term "socio-technical system" will be used to cover the technical and institutional levels). One advantage of this methodology is that by proceeding in this design oriented and proactive way, value conflicts and tensions may be anticipated and some of these may even be overcome and reconciled by means of design. It is often argued that this is precisely what genuine innovation is about: the introduction of new solutions that

accommodate conflicting or competing values that we find difficult to satisfy jointly (van den Hoven, Lokhorst, and van de Poel 2012) .

BEYOND ETHICAL DILEMMAS

This methodology also aims to go beyond one particular aspect of the current debate on the ethics of ADS: the "thought experiments" on **tragic moral dilemmas for future ADS**. In a series of academic and media articles (see e.g. Lin 2015), the reader is asked to imagine a future scenario in which a fully autonomous vehicle (without any human driving control onboard or remote supervision) faces an emergency situation in which a crash is unavoidable and the only choice open is one between hitting two or more different "targets"; according to different variations of the story, the autonomous vehicle has the "choice" to hit an old person or a young one; a motorist with a helmet and one without helmet; some innocent bystanders or an object which will kill the passengers of the vehicle, etc. How should the programmer instruct the vehicle to behave in these and similar circumstances? And who should take the decision (the programmer, the owner, the manufacturer)? The declared goal of one of the proponents of these cases is to show that "**ethics matter for self-driving vehicles**" (Lin 2015), and if one looks at the interest that these hypothetical cases have raised in the media and in the public debate, it can be said that this general goal has been achieved.

However, one problem with **this approach** is that while raising awareness on the ethical dimension of autonomous driving it **doesn't offer any clear methodology** to develop an ethics of automated driving systems. The reason for this is twofold: first, the approach is based on one particular **futuristic scenario**, in which fully autonomous vehicles are present in society, they operate in mixed traffic, one (individual) agent has the chance to freely decide on the programming, etc.; secondly, it addresses **only one specific controversial ethical issue** that may emerge in that particular scenario, that of the moral programming of the behaviour of autonomous vehicles in tragic, and dilemmatic emergency circumstances where a forced choice between the damaging of different human persons is requested.

In contrast with this approach, and on the basis of the principles of Responsible Innovation and Value-sensitive Design presented above, **I recommend**: a) **to not take for granted any specific future scenario** but rather exploring many different options until they are still open (we might decide that some specific future scenarios should not even realize); b) **to broaden the scope of the ethical analysis** as to include more societal and normative issues (not only specific and controversial ethical dilemmas); c) to take a **pro-active attitude** which tries to anticipate and avoid by design the realization of tragic moral dilemmas in the first place.

Overview of the ethical issues of ADS

Safety

Current vehicular traffic, it is often argued, is not sustainable, often inefficient and dangerous. **Safety** is particularly relevant from an ethical point of view. It has been convincingly argued that it is simply morally unacceptable to die while using the transport systems and that the system designers have the moral responsibility to prevent the realization of (fatal) accidents (Nihlén Fahlquist 2006, 2009). Insofar as the introduction of automated driving systems promises to reduce the number of accidents caused by human error, there may be **a moral obligation for policy makers** to promote the introduction of these systems in society.

However, reducing the impact of human drivers' errors is not the only obligation we have as a society. We also have an obligation to guide innovation without introducing **new unpredictable risks for human life.**

Meaningful human control and responsibility

In order to avoid the introduction of unwanted new risks for human safety, the introduction of ADS – I claim – should be done in such a way that **"meaningful human control"** over the behaviour of the system is always preserved. This can be done either by an appropriate design of a *partial automation* system, or by an appropriate design of a *supervised automation* system. Meaningful human control might in

principle be maintained also in a *fully automated* system, but achieving this combination may be problematic in the current state of art of technology and legislation.

Moreover, the introduction of ADS should be done in parallel with the introduction of new schemes for legal responsibility that prevent **"responsibility gaps"**, both in criminal and in tort law.

RIGHTS

The introduction of ADS may also take into account from the very beginning some basic individual **rights** and interests, including: the right to physical integrity (the **"new victims" problem**), the right to have **access to basic transport service** (especially underserved groups), and the right to **privacy.**

A PRO-ACTIVE ETHICAL APPROACH

From the proposed pro-active, "design" perspective, **the general overarching ethical question** about the future of ADS can be formulated in the new, following way: how can designers and policy makers comply with the moral obligation to make traffic safer, more sustainable and efficient via ADS *while at the same* time avoiding unreasonable risks, maintaining human accountability and respecting competing vital interests and individual rights like the right to life and physical integrity and the right to privacy.

THE TRANSITION TOWARDS AUTOMATION[2]

Worldwide automated driving is subject to intensive R&D, driven by the expectation that automated driving may deliver a breakthrough in making road traffic safer, more efficient and more sustainable. Much attention has been devoted to the futuristic scenario of cities swarmed with driverless automated vehicles (***full automation***). I think

[2] The content of this section is based on various conversations with Prof Bart van Arem from the department Transport and Planning at the Faculty of Civil Engineering at TU Delft. The responsibility for the content remains the sole author's.

that given the current state of art of technology and legal regulation this **is not a realistic and desirable scenario to pursue in the short run** (more on this below). I therefore focus on two emerging transitions toward automation that appear to be realistic in the coming 10-15 years.

The **first transition** starts from current road vehicles towards vehicles increasingly equipped with automated driving function, potentially leading to ***dual mode vehicles*** that can be driven either manually or in automated mode (*partial automation*). In automated mode the responsibility for steering, acceleration, deceleration, environmental monitoring and back up performance resides in an automated system. The automated mode may be restricted to special conditions such as well-maintained, well-mapped and well-marked motorways.

A **second transition** takes place starting from *single mode 'pods'* that are fully computer controlled and drive at low speeds on special tracks separated from other traffic. The transition takes place through the expansion of the network to which the vehicles are admitted, by allowing operation in mixed traffic and increasing the operational speed. While operating in automated mode, the **'pods'** are continuously **monitored by a (possibly remote) human supervisor** (*supervised automation*).

Although automated driving may be enjoying the use of colossal progress in information, sensing and computing research, I expect that the stepping stones for the coming 10-15 years will continue to involve **a human driver or supervisor in the loop in some way** and be restricted to specific operating conditions. One of the major challenges in the responsible introduction of automated driving we are facing is (1) which design of *partial or supervised automation* are desirable (2) under which operating and societal conditions.

ADS, Ethical issues and design challenges

Lessons from the autonomous weapon systems debate

The perspective of future development of fully autonomous robotic systems (operating without *any* human supervision) has already raised deep **ethical, societal and legal concerns**; the biggest concerns have been put forward in relation to a possible development and use of autonomous weapon systems in warfare (AWS, sometimes called **"killer robots"**) (Human Rights Watch 2012). Whereas unlike AWS, **civilian Autonomous Driving systems (ADS) are not designed to injure or kill, still they may accidentally do so**. Therefore, some of the ethical and legal concerns raised in relation to the deployment of fully autonomous (weapon) systems in warfare may also apply to ADS (Lin 2015).

Two main ethical concerns raised in relation to military robots arguably apply also to ADS:
a) it may be **morally wrong to give a technical system full control** over dangerous and potentially lethal activities (Asaro 2012, Wagner 2014), at least insofar as the system cannot replicate the relevant subtle ways of human (moral) thinking;
b) the use of ADS may create undesired **gaps in responsibility** attributions for damages caused by a reckless behaviour of the system (Matthias 2004).

"Meaningful human control"

In order to address these concerns, in the recent political debate on autonomous weapon systems the concept of **"*meaningful human control*"** has been put center stage (Horowitz and Scharre 2015). The notion of "meaningful human control" tries to captures three ideas.

Firstly, simple human presence or **"being in the loop" is not** a **sufficient** condition for being in control of an activity. It is not sufficient because one can be present in the sense of being able to influence some parts of the system by causal intervention, while (a) *not being able to influence other parts of the causal chains* that could

be seen as even more relevant than the parts one can influence or (b) *without having enough information* or options to influence the process, for instance if the human task consists in "merely pushing a button reflexively when a light goes on" (Horowitz and Scharre 2015).

Secondly, controlling in the sense of **being in the position of causally influencing the process** and/or the outcome of a (military) activity through one's intentional actions might **not** be a **sufficient** condition for meaningful control either, for instance if one does *not have the capacity* or the motivation to direct the activity in the (morally) right way.

Thirdly and relatedly, whereas some forms of legal responsibility (tort liability, strict liability) require only that the agent have relatively simple forms of causal control over events, other forms of ***legal responsibility*** (typically criminal responsibility) usually **require stricter control conditions** of knowledge, intention, capacity and opportunity; therefore no matter how strong the political will to keep some human responsible or accountable for the behaviour of autonomous systems, attributions of (criminal) responsibility that are not grounded in the relevant control conditions may turn out to be not only morally unfair but also ultimately difficult to enforce in tribunals. From this perspective, meaningful human control is required to make sure that every time that a potentially wrong (criminal) action is performed, for instance an injury or killing due to the reckless or negligent behaviour of a driving system some human agent is morally and legally liable (be it the programmer, the manufacturer, the driver, the traffic supervisor, etc.).

"MEANINGFUL" DOES NOT MEAN DIRECT

It is important to note that **"meaningful human control" does not require as a necessary condition that a human *directly* controls** every aspect of the operation of the system. Meaningful human control may be seen as close to Andy Clark's concept of **"ecological control"** in the philosophy of mind and human action. Ecological control is "the kind of top-level control that does not micro-manage every detail, but rather encourages substantial devolvement of power and responsibility" (Clark 2007:101). Clark sees the distribution of tasks –

to sub-personal unconscious mechanisms, or even to external tools – as a crucial factor even in everyday human control of one own's action. As applied to the behaviour of ADS the idea of ecological control entails that: **a human agent A may be in control of an action performed by an autonomous system** S provided that S is part of a system Y whose general functioning is guided by the moral and practical reasons of A; in particular, the *distribution of tasks* to different parts of the system is such as to allow the system to be responsive to the relevant moral and practical aims of A (cf. Di Nucci and Santoni de Sio 2014, Santoni de Sio and van den Hoven manuscript).

In this perspective **some kinds of delegation may preserve meaningful human control**, namely those in which the system as a whole (in the case of ADS: vehicle + technical infrastructure + social/legal institutions) is designed in such a way as to properly "respond" to the relevant moral and legal reasons of the human designers and users. As with military robots, also with civilian ADS, the major ethical challenge is that of elaborating a detailed theory of what "meaningful human control" exactly *means* (UNIDIR 2014, Horowitz and Scharre 2015); and to try to translate this into specific *legal regulations* and *design guidelines* for policy-makers and technical designers.

In what follows, I present what I see as the three main options open for the future of ADS: partial automation, supervised automation, and full automation, and I try to assess under which technical and social/legal conditions these forms of automation can remain under "meaningful human control".

PARTIAL AUTOMATION

PARTIAL AUTOMATION UNDER MEANINGFUL HUMAN CONTROL

Equipping vehicles with partial automation means assisting the human driver with more and more automated functions while at the same time expecting her to still perform certain driving activities. The

Tesla autopilot discussed in the Appendix to this paper is one example, but other companies have similar assisting driving technologies.
From the point of view of meaningful human control the ideal scenario is probably one in which there is **a clear division of the tasks**: the automated mode takes over only in special conditions such as well-maintained, well-mapped and well-marked motorways, whereas the human driver remains in charge in all other scenarios. This scenario looks safe from the point of view of meaningful human control because **in both modes the vehicles clearly respond to the motivations and intentions of some humans**: to the driver's when she is driving, to the reasons of the vehicle designer and the road planner when the vehicle is in the automated mode; this is true assuming that the interaction between vehicle and infrastructure has been properly designed and sufficiently tested. Also a fair application of moral and legal responsibility seems to be possible in this case: when the driver is (legitimately) in charge of the vehicle she is responsible for it; when she (legitimately) hands over the control to the vehicle-system, then designers of the vehicle and/or road system are morally responsible for mistakes.

PARTIAL AUTOMATION OUT OF MEANINGFUL HUMAN CONTROL

A more problematic scenario of partial automation is one in which the **human driver and the technical system share the control of the vehicle**, for instance the human driver may decide to leave the automated system the control of some dynamic operations like steering and braking, provided that she remains ready to intervene to perform some specific performance, for instance in the event of an unexpected problematic situation. One issue raised by this scenario is that of ***trust calibration***: on the one hand, **humans may overestimate the capacity of the vehicle** and not intervening even when required (overdelegation); on the other hand humans may underestimate the capacity of the system (and/or overestimate their own capacities) and keeping or taking control when it would be safer to leave the vehicle operate autonomously (underdelegation). The system might here not be under meaningful human control, insofar as, even though the distribution of tasks in the socio-technical system is properly designed

to guarantee an appropriate response of the vehicle to the external circumstances, still **the human driver is not equipped with the motivational capacity to realize this distribution of tasks** (for instance the driver tends to use the "autopilot" mode also when not recommended).

Another risk is that of a bad design of the transition between machine and human control. This may happen, for instance, when though the **human driver** is rightly motivated to take and shift control according to the design requirements, she **may not have psychological capacity to properly drive the vehicle once in control**, for instance because of the too short time available to regain awareness of the circumstances.

PARTIAL AUTOMATION AND MEANINGFUL HUMAN CONTROL: TECHNICAL AND NORMATIVE CHALLENGES

The design challenge of preserving meaningful human control in partial automation is complex. We need to ***understand*** **what the ideal distribution of tasks** between human driver and automated system would be both a) from a technical point of view: who can do what better? and b) from a social and psychological point of view: **how to make sure that the human will be able and motivated to do his part when requested**? Then we need to *implement* this distribution in an appropriate design of a socio-technical system, including new systems of training and licensing for users.

One technical challenge to be addressed with partial automation would be to design **an interface that is as transparent and "accountable"** as possible to the human driver, such that he is able to have a clear picture of what circumstances are, what he is supposed to do and how (de Greef 2016).

One **normative issue** would be to decide how much ***freedom*** should the driver be left under different circumstance to comply with the indication of the system. Assuming that she cannot be *forced* to leave the control of the vehicle, as this would be in violation of the principle of individual freedom, it is reasonable to assume that she may not just be neutrally *informed* that she should take/leave control, but she may also be *nudged* to do so.

SUPERVISED AUTOMATION

SUPERVISED AUTOMATION AND MEANINGFUL HUMAN CONTROL

Currently, *single mode 'pods'* already exist that are fully computer controlled and drive at low speeds on special tracks separated from other traffic (e.g. the "Wepods" currently tested in the Netherlands). We can imagine that in the near future, the network to which these vehicles are admitted is expanded and eventually these are allowed some operation in mixed traffic, possibly with an increased operational speed. While operating in automated mode, the **'pods'** are **continuously monitored by a (possibly remote) human supervisor** (*supervised automation*). This scenario has the following advantages. First, whereas the **control of the system is in a way shared** between humans and machine, as no human driver is present onboard, **the human tasks are discharged by trained professionals** who sit in a remote control room, and therefore do not operate under time or emotional pressure.

Meaningful human control may be achieved here by a combination of design of smart infrastructure and appropriate (remote) human supervision. In order to extend this model as far as becoming largely available in cities, **a substantive redesign of the urban landscape** is in order. But this is not necessarily a negative aspect: one important drive behind the ADS "revolution" is arguably the idea of going beyond the current model of urban traffic centred on private, human-driven vehicles.

INSIGHTS FROM CIVIL AVIATION?

One possible methodological suggestion in order to develop a model of supervised automation would be to **look at the problems and solutions of control in civil aviation**, where airplanes are flown and controlled via a combination of human control onboard, automated pilot but also complex mechanisms of remote traffic control.

FULL AUTOMATION

FULL AUTOMATION AND "MEANINGFUL HUMAN CONTROL"?

Finally, removing the human driver from the loop of the driving system as proposed, for instance, by Google vehicles does not *necessarily* mean losing meaningful human control. However, maintaining this control would be very difficult and it is **an open question** how long it may take for technology and society to be ready **to introduce "fully autonomous" vehicles which do not risk to go out of meaningful human control**; that is, they will be sufficiently responsive to the moral and legal reasons of humans who design, use and interact with these vehicles (Hearing Self-Driving Cars 2016).

THE ETHICS OF TRIAL-AND-ERROR

A first problem with a full automation like the one proposed and advertised by Google, is that its efficiency is based on learning through a huge acquisition and elaboration of data: this means that in order to reach a sufficient level of responsiveness to a fair number of different circumstances a full autonomous vehicle has to "learn" to drive on the real roads for a very long time. In the section on the "Responsible Innovation and special testing zones" below, I argue that such a **trial-and-error procedure** is **not socially responsible**: no matter how good the end-point result might be, we should not allow for systematic learning-by-mistakes trials of not-yet-proven safe technology in the real world, any more than we would allow for trial-and-error introduction of life-saving medicines on the market. We'd rather have to make first controlled tests to actively prevent fatal effects to occur.

MACHINE LEARNING AND BAD MASTERS

Moreover, as it would simply be impossible to test all potential autonomous vehicles behaviours under all possible circumstances, due to the dynamic nature of the real-life environments in which they will operate, unpredictable outcomes are in principle always possible.

One recent example of a Artificial Intelligence fail is the Microsoft "chatbot" Tay, a robot Twitter user switched off after few days because it has learnt from users to make racist and discriminatory comments. This has been a relatively innocuous accident, which still sheds some light on a twofold risk of adopting **a machine learning approach** in open environments: **machines could learn from anything and anyone**; in the case of ADS from the rule-abiding and reasonable human driver, from the naughty but still reasonable one but also from the reckless and dangerous ones; in addition, by knowing that such open systems are circulating ill-intentioned persons may intentionally trick the them into behaving in undesirable and dangerous way.

Mixing intelligences

Thirdly and relatedly, fully autonomous vehicles will have to **interact with other road users**, and even assuming that they will be able to comply with all traffic rules, they will have to coordinate with different cognitive and behavioural styles especially in the interpretation and application of these rules. One big challenge of this scenario is the interaction between **different kinds of intelligence**.

On the one hand, we may want ADS to strictly and rigidly abide by the traffic rules; but this may create **a gap in the interaction with human users** who typically adopt a more reasonable, flexible, experience-based, sometimes loose and even naughty interpretation of the rules of traffic. On the other hand, assuming that we want automated vehicles to learn to reason like humans, that is to make case-by-case evaluations of circumstances rather than rigidly applying top-down rules and procedures we would again need to allow for a massive use of learning machines on the road, with the risks for control described above.

Non-verbal communication between human and robotic road users

Finally, fully autonomous vehicles operating in mixed traffic will face the issue of the **communication with humans**. A good part of everyday communication between humans happens through non-verbal communication (**body language or sign language**); this is even

more true in road traffic condition, where verbal communication is most of the time impossible. Humans communicate their intentions and attitudes to each other via conventional signs, gestures, etc. Entering this arena might be particularly problematic for a robotic vehicle. In her recent hearing at the US Senate, Duke researcher Mary Cummings pointed out that current autonomous vehicles would be for instance unable to follow the indication of a policeman ("Hearing Self-Driving Cars | Video | C-SPAN.org" 2016).

The design challenge is here twofold: on the one hand, new automated **vehicles may be equipped with systems from interpreting and conveying simple conventional message** (for instance, leds blinking to "tell" a pedestrian: "I have seen you, I will wait for you to cross the road"). On the other hand **human drivers might have to learn** to understand, anticipate and smoothly interact with automated driving systems, and this would require: a) the creation of a new repertoire of signs and conventions; b) an introduction of this capacity in the licensing procedures; c) a system to prevent, discourage and penalise the intentional gaming or tricking of autonomous vehicles by human users.

Responsible Innovation and "special testing zones"

From a broader societal perspective, ADS developers who want to design for complex socio-technical values like meaningful human control may face the following dilemma: they need to make reliable tests on the interaction between complex systems and real people in a real environment: a real city. However, **doing tests with real people in real cities is morally prohibited** until the technology has proven to be safe enough. This may lead to **a stalemate in the progress of innovation**. In order to address this problem, the Japanese cities of Fukuoka, Osaka, Gifu, Kanagawa and Tsukuba have created **"special zones" for the testing of robotic technologies**, under the incentive of a special legislation of the Japanese government (Weng et al. 2015).

The "special zone" solves this problem by creating a controlled space within the real society where general regulations on the use of robots

are adjusted in order to allow for the presence of some test robots, which have already been proven to be safe in laboratory; on the other hand, special precautions are taken in order to prevent serious accidents and undesired outcomes, for instance persons entering the zone receive specific information, there are specific signs, and possibly specific insurance schemes to cover unexpected damages.

The general goal of a "special zone" is that of **acquiring reliable information** about the precise nature and scope of specific robots' impact in real cities, **while avoiding** the risks of a hasty and **irresponsible introduction**. This will allow to achieve two specific goals: a) boosting the research and design of robots which guarantee high levels of autonomy while at the same time guaranteeing safety and human responsibility; b) helping institutions and policy-makers to develop well-informed policies and legal regulations for the responsible introduction and use of robots in the cities of the future.

"Responsibility gaps"

Crimes

Another great concern raised in the ethical literature on autonomous robots is that their use may lead to unacceptable **"responsibility gaps"** (Matthias 2004, Sparrow 2007, Di Nucci and Santoni de Sio 2016), circumstances in which a serious accident happens and nobody can be reasonably held responsible or accountable due to the unpredictability or opaqueness of the process leading to the accident. Again, this concern is particularly serious in warfare, where serious crimes can be committed, triggering major moral and legal obligations which only humans can fulfill: to account for crimes, to offer moral remedy and economic compensation to the victims, to have someone publicly held blameworthy and punished etc (Saxon 2016, Meloni 2016).

However, similar concerns may arise also with ADS: as the American judge LJ Hale once wrote, **cars are "potentially dangerous weapons"** (*Eagle v Chambers*), whose reckless or **negligent use sometimes amounts to a crime**. Also in relation to road traffic, it is therefore

necessary to design the system in such a way that dangerous behaviour by ADS are prevented, and when this happens it is possible to have some human agent accountable, liable and punishable for that. Given the different distributions of control presented above, this should not necessarily be the "driver", but it might be a designer, a programmer or a controller. Here again, a comparison with the legal distribution of (criminal) responsibility in complex partially automated transport like train or aviation systems may be helpful to develop a legal policy.

TORTS

Road traffic also raises important issues of **tort liability**: **who should pay** for the costs of the accident. From a legal perspective the introduction of ADS presents at least three new issues (Pagallo 2013: 110): first, **the law has so far seen robots and autonomous systems merely as tools** and not as agents and doesn't seem equipped to cope with the presence of non-human intelligent systems (see also Calo 2016); secondly and relatedly, when systems equipped with complex artificial intelligence are used **the driver/owner may not always be responsible** for the behaviour of the system: sometimes others should, other times nobody may, for instance in the event of a malfunctioning that no reasonable person could have predicted; in this respect things can be even more complex in the case of a shared vehicle, where owner and user do not coincide (see section below on ownership); third, unlike what happens for instance with robo-traders, liability for **road accidents concern also "extra-contractual" third parties**, that is parties not bound by any contractual relationship with the owner/driver (a typical example here would be an unknown pedestrian).

INSIGHTS FROM TORT LAW

According to current tort law an agent can be held liable to pay for a damage either based on ***fault liability***, when the damage has been caused by a breach of a standard of reasonable behaviour, or based on ***strict liability***, that is independently from any fault. Varieties of

strict liability would be: vicarious responsibility, when the defendant is held liable for the damage caused by another agent connected to the defendant by a special legal relationship, i.e. employer-employee or parent-son/daughter; and product liability, when the defendant, typically a company, is held liable for a damage caused by a defective product.

If the law sticks to its current attitude of considering robotics systems as mere products or tools rather than agents, the **only available applicable laws** in the case of an accident involving ADS will be **fault liability and product liability**: that is, a compensation will be due to the plaintiff in either of the following two case: a) the driver, the manufacturer or the programmer have breached a recognized standard of precaution in the production/use of the ADS (fault liability); b) the manufacturer is held strictly liable for the malfunctioning of the ADS (independently from her fault). **Fault liability** has the advantage of being fair to the human actors, as it applies only to negligent behaviour; its downside is that it will arguably not apply to many accidents where the malfunctioning of the system was fairly unpredictable, so that **some damages may remain not covered**, to the disappointment of the plaintiffs. **Strict liability** has the advantage of always guaranteeing a compensation to the plaintiffs, while putting a high burden on programmers and manufacturers. In a nutshell, fault liability may create legitimate public discontent, strict liability may **discourage innovation**.

One alternative solution would be for the law to start considering ADS as agents, and applying some form of vicarious responsibility to their "parents" or "employers". However, it is not clear who should count as a parent or employer here (the manufacturer, the programmer, the owner, the user, etc.) and any of the party burdened with this strict liability may criticize the fairness of such a scheme.

Insights from Roman law

In order to strike a fair balance between the different interests involved in these scenarios, some academic lawyers have suggested to look at one institution typical of the Roman law: the peculium. In *The Human Use of Human Beings* (1950), the father of cybernetics,

Norbert Wiener, wrote that **"the automatic machine … is the precise equivalent of slave labor."** (cited by Pagallo 2013: 102). This comparison may be appropriate also from a legal point of view, since in ancient Rome slaves were considered as things that still could perform some social activities. The ***peculium*** was a limited amount of money assigned to these slaves to do some activities on behalf of their master. It aimed to strike a balance between the claim of the masters not to be dilapidated by their slaves' businesses and commercial activities (as the slave responsibility was limited by the amount of their available peculium) and the interest of the slaves' counterparties to safely transact with them (Pagallo 2013: 103). It has been suggested that a **"digital peculium"** may be created to give robotic agents, including ADS, a limited capacity to pay for damages that cannot be traced to any fault or liability of their human "masters". The question remains, who should fund the digital peculium, and how to determine its value. One promising solution would be to create a sort of insurance scheme to fund the digital peculium.

"Designing" and "Testing" tort liability

Designing legal constructions to cover for damages of automated driving systems is another challenging "design" task. Unlike criminal law, which is based on the principle of the rule of law, namely on the necessity of having clear (prohibitive) rules set in advance, civil law also strongly rely on the societal experience: standards of reasonable behaviour in different areas of life are elaborated over time, risks and damages are assigned a monetary value through the accumulation of experience of similar cases, and so forth. With the introduction of ADS, **lawyers and legislators** are facing a problem similar to that faced by technical designers and programmers: the **lack of "data" to determine the relevant "parameters"**. Assuming, for instance, that they want to introduce a system like the "digital peculium" funded by an insurance scheme (see the previous section), it would not be easy to determine the values of the peculium and of the insurance premiums, in the absence of any record or precedent. From this perspective the idea of "special testing zones" presented above may reveal decisive also from a legal point of view. They may be used not only to make controlled tests on technical aspects of ADS but also to

start **acquiring data about legal and economic aspects** of the interaction between humans and ADS.

RIGHTS

Ethical research has highlighted some important **limits of so-called "cost-benefit analyses"** in ex-ante policy analyses, and of the "utilitarian" approach on which it is grounded (van Wee and Roeser 2013). The most important limits concern: CBA not taking into sufficient consideration "deontological constrains", i.e. the moral duties and rights of individuals; CBA focusing more on the "aggregate" results of a policy, and not being sufficiently sensitive to the issue of a fair distribution of benefits over different groups or individual in society. A responsible introduction of ADS shouls take into account also these non-utilitarian elements.

"NEW VICTIMS": RIGHT, RESPONSIBILITIES, AND LONG-TERM EFFECTS OF POLICIES

While considering the potential impact of the introduction of ADS on the safety of transport, one important consideration is certainly whether this introduction will lead to an overall reduction of (fatal) accidents. However, I have argued above that there are ethical constrains to the process through which we try to get to desired reduction of accidents (see the section "The ethics of trial-and-error" above). Another important issue is whether and how this introduction will change the ***distribution* of the risks** of losses and damages across different (groups of) people in society. Let's assume for instance, that the introduction of ADS will considerably reduce the number of losses and damages among vehicle drivers and passengers while at the same time also (slightly) increasing the number of losses among pedestrians. Would such a scenario be morally acceptable? Whereas, based on a simple utilitarian interpretation, such a scenario may look acceptable (the overall death toll would be lower), two objections may be advanced, one based on the **normative positions of the actors involved** (their rights and responsibilities), the other based on the **long-term effect of a policy**.

THE RIGHT NOT TO BE INTENTIONALLY INJURED OR KILL AND THE ROLE OF PUBLIC AUTHORITY

The legal-philosophical literature on the "doctrine of necessity" (Dennis 2008) shows that our legal systems set quite **stringent limits to the use of intentional force on innocent persons**, even when some other, possibly greater, evils can be avoided by the use of force. For instance, it is usually legally forbidden to intentionally injure or kill an innocent person P in order to save other people A, B, C from a threat which is not related to the behaviour of P. This prohibition is grounded in **the individual right to life and physical integrity**. So, it may be thought that this prohibition applies also to the developers of a policy on road traffic (Santoni de Sio, manuscript): a reduction of the losses of motor vehicle drivers may not be achieved, for instance, via an expected increase in the losses among pedestrians or cyclists. On the other hand, unlike private citizens, the public authority *does* have the legitimate power to decide to penalize one specific group of citizens in order to pursue what it considers to be the "public good", so that looking at individual rights may not be enough to decide on this issue (Christie 1999).

SPECIAL RESPONSIBILITIES OF (PRIVATE) MOTOR VEHICLES USERS

Another relevant aspect to consider here is the normative relationship between (private) motor vehicle users and other road users. Current law puts a **high duty of care on motor vehicle drivers**, based on the fact that they handle "potentially dangerous weapons" (LJ Hale In *Eagle v Chambers*). It is reasonable to think that a similar or even higher burden should be put on designers and/or users of (private) ADS. In fact, even assuming that future AVs will be safer than current ones in the sense of causing lesser accidents due to the elimination of the impact of human drivers errors, ADS will still be potentially more dangerous than bicycles or pedestrians in the sense of having **a higher potential for causing serious losses and damages to third parties** in the event of a crash. In addition, according to current tort law the duty of care of drivers towards pedestrians extends as far to cover most of the damages that could be prevented by the driver's diligent behaviour, no matter how negligent the behaviour of pedestrians

might be. Therefore, **the enhanced ability for crash avoidance brought by artificial intelligence may put an even higher duty of care** on those who want to introduce new vehicles on the public road.

"Spirit" and long-term effects of ADS policies

An additional reason to maintain a stringent obligation to not harm pedestrians, cyclists and other road users on the part of motor vehicles designers and users is that this is in **the deep and long-term interest of a consistent policy oriented at creating a safer, more sustainable and efficient traffic**. We should avoid the paradoxical scenario in which the presence of ADS on the road is perceived as threatening the safety of those who do not use motor vehicles, as this scenario might eventually create an incentive to use motor vehicles also for those who would be able and willing to not do so; and so reducing (or even eliminating) one or more of the beneficial goals of the policy. In other words, and here comes another "design challenge", we should strive to both increase safety, sustainability, efficiency of vehicular traffic *and* maintaining cities as pedestrian- and cyclist-friendly as possible.

Access to mobility

Another related problem of a simple "cost-benefit" approach to a policy on ADS is that it may lead to neglect the ***distribution* of benefits** across different (groups of) citizens. It is often stated that ADS may be particularly beneficial for minority groups who cannot drive traditional vehicles, for instance due to age, physical and psychological conditions, and the like; and that this is a moral reason to support the introduction of ADS. However, whether the introduction of ADS will in fact benefit these groups depends not only on the availability of the technology but also, and crucially, on its ***accessibility***: **economic** and **technical**. Different socio-technical design choices may produce different outcomes in terms of accessibility. Relevant factors include: the cost of the ADS vehicles; whether ADS vehicles will be mainly private or also utilized in public services; whether they will be utilized in car-sharing schemes; whether the access to ADS public or shared services will require

specific technical (use of digital technologies) or economic capabilities (costly subscriptions), and whether there will be appropriate programs aimed to provide underserved groups with these capabilities.

PRIVACY AND DATA PROTECTION

From a technical point of view, many of the valuable goals listed and discussed so far: safety, accountability, accessibility may be achieved by a huge acquisition, storage and elaboration of data. Both partial automation and supervised automation systems may need to acquire **a huge amount of data** about the behaviour of the vehicles and their drivers or passengers via sensing and communication technologies. Moreover, by being equipped with sensors and cameras, the vehicles are also likely to incidentally acquire many data on other road users interacting with them.

Two ethical and societal risks highlighted in the ethical literature on **privacy and data protection** (e.g. van Den Hoven 2008) are clearly present also in the case of ADS. Firstly, ADS may be the target of cyber attacks or hacking. Secondly, the massive acquisition and storage of personal data about road users may threaten their moral autonomy in two ways: a) by creating an **information asymmetry**: a huge quantity of *information about individual persons* may become available to those who own or control transport infrastructures. This information may be used to benefit citizens, but there is also a risk that it will be collected and used against the interests of minorities or even the majority of people and in violation of their rights; b) by creating an **imbalance of power**: similarly, a dramatic increase in the information capabilities of governments or other agencies may enhance their capacity to promote the citizen's safety and well-being, but this capacity may also be used to *control, coerce, exploit, discriminate* and even *oppress* people.

Serious ethical issues of privacy and data protection have emerged in relation to the current use of digital technologies. We now have an advantage: ADS are not yet on the road, and we have the chance to regulate the acquisition, storage and sharing of personal data in

advance, based on the experience and mistakes from these other sectors.

APPENDIX: THE TESLA AUTOPILOT FATAL ACCIDENT OF MAY 7TH 2016

During the preparation of this paper the first fatal accident involving a vehicle equipped with an automated driving system – a Tesla Model S – has been reported.

THE ACCIDENT

On Saturday, May 7th, 2016, a 2015 Tesla Model S, traveling on US Highway 27A, west of Williston, Florida, struck and passed beneath a 2014 Freightliner Cascadia truck-tractor in combination with a 53-foot semitrailer. The 40-year-old driver and sole occupant of the Tesla died as a result of the crash. System performance data revealed that the driver was operating the car using the advanced driver assistance features Traffic-Aware Cruise Control and Autosteer lane keeping assistance ("Highway Preliminary Report HWY16FH018" 2016).

WHAT DOES THE TESLA ACCIDENT SAY

As the investigations on the accident are still in progress I won't express any opinion about the causes of this particular accidents or the responsibility of the parties involved in it. However, based on the general description of the accident made available by the authorities, I will quickly address the question: What, if anything, does the actual occurrence of a fatal accident involving an autonomous driving system say about the ethical approach presented in this paper. My answer, in a nutshell, is that the occurrence of such an accident confirms the soundness of the proposed ethical approach. In fact:

1) according to the proposed **pro-active ethical approach** (value-sensitive design), it is the responsibility of designers and policy-makers to see to it that the introduction of ADS does not create new serious and unpredictable risks for human safety; from this perspective, it does not make much difference whether an

accident like the one of May 7th was caused by a technical fault or by an inappropriate use of the technology by the human driver; what matters is whether designers and policy-makers could have prevented the accident through a better design of the technical system (the vehicle) or the socio-technical system (vehicle + road + regulations + behaviour of the human driver);

2) Maintaining driving systems with partial autonomy like the Tesla autopilot under "**meaningful human control**" is more challenging than some think; from a technical perspective, the system should be able to respond in the right way to as many circumstances as a human driver would be; from a socio-technical perspective, the shift of control between the technical and the human agent should be designed by keeping into account not only the respective technical capabilities, but also the human psychological and motivational capabilities (see the section "Partial automation out of meaningful human control" above);
3) The task of assessing the **moral and legal responsibilities** of the actors involved in accidents like that of May 7th may be difficult and may lead to morally unsatisfactory results if a new appropriate system of liability for this kind of accidents is not designed in due time, ideally before such accidents happen (see the sections on "Responsibility gaps" above);
4) Whereas a zero-risk approach to transport technology would probably be not recommendable or feasible, at least the risks that fatal accidents like that of May 7th happen during the trial stage should be avoided; a more robust policy about **"special zones"** (see section "Responsible Innovation and special testing zones" above) may provide a more acceptable balance between safety and innovation.

CONCLUSION AND RECOMMENDATIONS

In conclusion, based on the previous discussion, the following recommendations can be formulated.

From a **methodological** point of view, we should:

- ✓ embrace the methodology of **Responsible Innovation and Value-Sensitive Design**, and create the conditions for interdisciplinary, anticipatory analyses, aimed at embedding ethical values into future socio-technical automated driving systems
- ✓ adopt **a more comprehensive ethical approach**, aiming to:
  a) improve safety by reducing of accidents caused by human error *but also* by avoiding the introduction of new risks with a potential negative impact on human safety (new kinds of fatal accidents)
  b) enhance human moral and legal responsibility
  c) respect individual rights

From a **practical** point of view, we should:

- ✓ promote a gradual introduction of automated driving systems, one which starts from partial and supervised automation and gradually moves towards higher levels of automation
- ✓ maintain ADS within "meaningful human control" in order to:
  a) prevent new kinds of accidents with potential negative impact on human safety
  b) avoid moral and legal "responsibility gaps"
- ✓ create the conditions for a systematic testing of ADS, for instance by a strong policy of "special zones"
- ✓ develop appropriate legal constructions to regulate the legal liabilities of different actors involved in the design and use of ADS
- ✓ promoting ADS that do not penalize vulnerable road users, and are accessible to underserved groups (technologically and financially)
- ✓ promoting ADS that respect privacy by design

References

Asaro, Peter. 2012. “On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-Making.” *International Review of the Red Cross*. https://www.icrc.org/eng/resources/documents/article/review-2012/irrc-886-asaro.htm.

Calo, Ryan. 2016. “Robots in American Law.” *University of Washington School of Law Research Paper No. 2016-04*.

Christie, George C. 1999. “The Defense of Necessity Considered from the Legal and Moral Points of View.” *Duke Law Journal* 48 (5): 975–1042.

Clark, Andy. 2007. “Soft Selves and Ecological Control.” In *Distributed Cognition and the Will: Individual Volition and Social Context*, by Don Ross, David Spurret, Harold Kincaid, and G. Lynn Stephens. MIT Press.

De Greef, Tjerk. 2016. “Delegation and Responsibility: A Human-Machine Perspective.” In *Drones and Responsibility: Legal, Philosophical and Socio-Technical Perspectives on Remotely Controlled Weapons*, by Ezio Di Nucci and Filippo Santoni de Sio. Routledge.

Dennis, Ian Howard. 2008. “On Necessity as a Defence to Crime: Possibilities, Problems and the Limits of Justification and Excuse.” *Criminal Law and Philosophy* 3 (1): 29–49. doi:10.1007/s11572-008-9062-5.

Di Nucci, Ezio, and Filippo Santoni de Sio. 2014. “Who’s Afraid of Robots? Fear of Automation and the Ideal of Direct Control.” In *Roboethics in Film*, by Fiorella Battaglia and Natalie Weidenfeld. Pisa University Press.

———. , eds. 2016. *Drones and Responsibility: Legal, Philosophical and Socio-Technical Perspectives on Remotely Controlled Weapons*. Emerging Technologies, Ethics and International Affairs. Routledge. https://books.google.nl/books/about/Drones_and_Responsibility.html?hl=it&id=Pi2TDAAAQBAJ.

Friedman, Batya. 1996. “Value-Sensitive Design.” *Interactions* 3 (6): 16–23. doi:10.1145/242485.242493.

“Hearing Self-Driving Cars | Video | C-SPAN.org.” 2016. https://www.c-span.org/video/?406555-1/hearing-selfdriving-cars.

“Highway Preliminary Report HWY16FH018.” 2016. Accessed August 12. http://www.ntsb.gov/investigations/AccidentReports/Pages/HWY16FH018-preliminary.aspx.

Horowitz, Michael, and Paul Scharre. 2015. “Meaningful Human Control in Weapon Systems: A Primer.” Center for a New American Security. http://www.cnas.org/human-control-in-weapon-systems.

Human Rights Watch. 2012. “Losing Humanity.” *Human Rights Watch*. November 19. https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots.

Lin, Patrick. 2015. “Why Ethics Matters for Autonomous Cars.” In *Autonomes Fahren*, edited by Markus Maurer, J. Christian Gerdes, Barbara Lenz, and Hermann Winner, 69–85. http://link.springer.com/chapter/10.1007/978-3-662-45854-9_4.

Matthias, Andreas. 2004. “The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata.” *Ethics and Information Technology* 6 (3): 175–83.

Meloni, Chantal. 2016. “State and Individual Responsibility for Targeted Killings by Drones.” In *Drones and Responsibility: Legal, Philosophical and Socio-Technical Perspectives on Remotely Controlled Weapons*, by Ezio Di Nucci and Filippo Santoni de Sio. Routledge.

Nihlén Fahlquist, Jessica. 2006. “Responsibility Ascriptions and Vision Zero.” *Accident Analysis & Prevention* 38 (6): 1113–18. doi:10.1016/j.aap.2006.04.020.

———. 2009. “Saving Lives in Road Traffic—ethical Aspects.” *Zeitschrift Fur Gesundheitswissenschaften* 17 (6): 385. doi:10.1007/s10389-009-0264-7.

Pagallo, Ugo. 2013. *The Laws of Robots: Crimes, Contracts, and Torts*. Springer. http://www.springer.com/gp/book/9789400765634.

Santoni de Sio, Filippo. manuscript. “Killing by Autonomous Vehicles and the Legal Doctrine of Necessity.” *Manuscript under Review*

Santoni de Sio, Filippo, and Jeroen van den Hoven. manuscript. “Meaningful Human Control over Autonomous Systems: A Philosophical Analysis.” *Manuscript under Review*

Saxon, Dan. 2016. “Autonomous Drones and Individual Criminal Responsibility.” In *Drones and Responsibility: Legal, Philosophical and Socio-Technical Perspectives on Remotely Controlled Weapons*, by Ezio Di Nucci and Filippo Santoni de Sio. Routledge.

Sparrow, Robert. 2007. “Killer Robots.” *Journal of Applied Philosophy* 24 (1): 62–77. doi:10.1111/j.1468-5930.2007.00346.x.

UNIDIR. 2014. “The Weaponization of Increasingly Autonomous Technologies: Considering How Meaningful Human Control Might Move the Discussion Forward.” UNIDIR (United Nations Institute for Disarmament Research). http://www.unidir.org/files/publications/pdfs/considering-how-meaningful-human-control-might-move-the-discussion-forward-en-615.pdf.

Van den Hoven, Jeroen. 2007. “ICT and Value Sensitive Design.” In *The Information Society: Innovation, Legitimacy, Ethics and Democracy In Honor of Professor Jacques Berleur S.j.*, edited by Philippe Goujon, Sylvian Lavelle, Penny Duquenoy, Kai Kimppa, and Véronique Laurent, 67–72. IFIP International Federation for Information Processing 233. Springer US. http://link.springer.com/chapter/10.1007/978-0-387-72381-5_8.

———. 2008. *Information Technology and Moral Philosophy: Information Technology, Privacy, and the Protection of Personal Data*. /chapter.jsf?bid=CBO9780511498725A022&cid=CBO9780511498725A022.

———. 2013. “Value Sensitive Design and Responsible Innovation.” In *Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society*, 75–83. John Wiley & Sons, Ltd. http://onlinelibrary.wiley.com/doi/10.1002/9781118551424.ch4/summary.

Van den Hoven, Jeroen, Gert-Jan Lokhorst, and Ibo Van de Poel. 2012. “Engineering and the Problem of Moral Overload.” *Science and Engineering Ethics* 18 (1): 143–55. doi:10.1007/s11948-011-9277-z.

Van Wee, Bert, and Sabine Roeser. 2013. “Ethical Theories and the Cost–Benefit Analysis-Based Ex Ante Evaluation of Transport Policies and Plans.” *Transport Reviews* 33 (6): 743–60. doi:10.1080/01441647.2013.854281.

Wagner, Markus. 2014. “The Dehumanization of International Humanitarian Law: Legal, Ethical, and Political Implications of Autonomous Weapon Systems.” *Vanderbilt Journal of Transnational Law* 47. https://wp0.its.vanderbilt.edu/jotl/2014/12/the-dehumanization-of-international-humanitarian-law-legal-ethical-and-political-implications-of-autonomous-weapon-systems-2/.

Weng, Yueh-Hsuan, Yusuke Sugahara, Kenji Hashimoto, and Atsuo Takanishi. 2015. “Intersection of ‘Tokku’ Special Zone, Robots, and the Law: A Case Study on Legal Impacts to Humanoid Robots.” *International Journal of Social Robotics* 7 (5): 841–57. doi:10.1007/s12369-015-0287-x.

Eagle v Chambers. 2004. RTR 9