Nanoethics (2020) 14:285–299
https://doi.org/10.1007/s11569-020-00374-4

ORIGINAL RESEARCH PAPER

Programming Away Human Rights and Responsibilities? "The Moral Machine Experiment" and the Need for a More "Humane" AV Future

Mrinalini Kochupillai · Christoph Lütge · Franziska Poszler

Received: 4 June 2019 / Accepted: 7 September 2020 / Published online: 25 November 2020

Dilemma situations involving the choice of which human life to save in the case of unavoidable accidents are expected to arise only rarely in the context of autonomous vehicles (AVs). Nonetheless, the scientific community has devoted significant attention to finding appropriate and (socially) acceptable automated decisions in the event that AVs or drivers of AVs were indeed to face such situations. Awad and colleagues, in their now famous paper "The Moral Machine Experiment", used a "multilingual online 'serious game' for collecting large-scale data on how citizens would want AVs to solve moral dilemmas in the context of unavoidable accidents." Awad and colleagues undoubtedly collected an impressive and philosophically useful data set of armchair intuitions. However, we argue that applying their findings to the development of "global, socially acceptable principles for machine learning" would violate basic tenets of human rights law and fundamental principles of human dignity. To make its arguments, our paper cites principles of tort law, relevant case law, provisions from the Universal Declaration of Human Rights, and rules from the German Ethics Code for Autonomous and Connected Driving.

Keywords Autonomous vehicles · Human rights · Law · Tort law · Ethics · Human values · Responsibility · Human dignity · Artificial intelligence · Moral machines

M. Kochupillai (✉)
Department of Aerospace and Geodesy, Technical University of Munich, Munich, Germany
e-mail: m.kochupillai@tum.de

M. Kochupillai
Munich Intellectual Property Law Center, Munich, Germany

M. Kochupillai · C. Lütge · F. Poszler
Institute for Ethics in Artificial Intelligence, Technical University of Munich, Munich, Germany

C. Lütge
e-mail: luetge@tum.de

F. Poszler
e-mail: franziska.poszler@tum.de

C. Lütge · F. Poszler
Chair of Business Ethics, Technical University of Munich, Munich, Germany

Introduction

Autonomous (or, as some prefer, automated) vehicles (AVs) are currently categorized based on the level of automation. Differently stated, they are categorized based on ranges of human intervention required for a vehicle to perform a desired task (e.g., getting from point A to point B), while avoiding accidents/collisions. The Society of Automotive Engineers (now called SAE International) provides a classification system based on five different levels of automation—from fully manual (level 0) to fully automated (level 5), where even the steering wheel is optional [1]. The majority of AVs on the road today are at SAE level 2, i.e., partial

automation. This means that the vehicle is equipped with systems that can control steering and acceleration/deceleration, together with sensors that can collect and provide to the driver, information or data about the driving environment. The driver can then choose when to use/act on these systems and information.

Significant amounts of research, from diverse disciplinary perspectives, are underway to make the transition of AVs incorporating higher levels of automation into regular traffic as seamlessly as possible. In the broad context of safety, researchers are currently preoccupied with questions that are relevant in a twofold mixed driving scenario—i.e., (i) a scenario where roads have AVs without drivers as well as vehicles with drivers and (ii) a scenario where there are vehicles with only level 2 of automation (and associated systems) driving together on the roads with an increasing number of vehicles with level 4 or 5 of automation.

In this twofold mixed driving scenario, highly complex techno-scientific research is already engaged with making the various levels of automation practically possible. Research is also underway to find means of making machines assist or even substitute human actors in decision-making (with the aim, inter alia, of reducing the number of human-error-based casualties on roads) in diverse situations. At the same time, part of the (social science and economics) research associated with AVs, including AV safety, continues to deal with the *acceptability* and *impact* of automation per se among human actors from diverse backgrounds. Recent studies, for example, emphasize that urban aesthetics and overall user experience and values must be considered and discussed before rolling out specific technologies, including AVs. This is necessary, inter alia, to ensure value-sensitive design, that does not (inadvertently) ignore the diversity of human values witnessed within existing "city identities"; technology, including AV technology, ought not to change these values and identities without engaging in participatory and constructive technology assessment [2, 3].

Existing research in the context of AV safety and delegation of decision-making (including in dilemma situations), either directly or indirectly, also directs attention towards the divergent rights and values that AV research must reconcile or balance out. This is necessary, inter alia, to make AVs more palatable, if not immediately acceptable, to the consuming or affected public. Indeed, the diversity of conclusions reached by the current research in this sphere indicates that there are several value conflicts related to AV technology, even within the limited context of safety.

Recent experimental research, for example, suggests that people are hesitant to delegate tasks that affect third parties, to machines, and observers judge such delegations critically [4]. Such studies are encouraging in that they testify to human eagerness to take and hold on to responsibility, especially when the lives and safety of others are dependent on their actions or inactions. This eagerness, in turn, shows a high moral fabric and the willingness to be held responsible rather than being "helpless"—important and fundamental human values, on which societies depend for their growth and evolution.[1]

Other studies, however, have shown that while people like the idea of AVs sacrificing the lives of their passengers for the "greater good," they would themselves never buy such an AV [7]. Further, while this study suggests a preference to save one's own life (and the lives of those who are closest to one), other studies, for example by Huebner and Hauser [8] and Di Nucci [9], assume that all or even a (significant) majority of "real" people would self-sacrifice in certain situations. This research is also relevant in the context of the highly distressing "trolley dilemma", which highlights situations where the driver/car has only one of two options: swerving left to avoid one set of casualties or swerving right to avoid another.

From the above, it is clear that currently, there is little consensus on the fundamental questions of desirability, acceptability and scope of delegating decision-making powers to machines (including, if not especially, in dilemma situations). Yet, a recent, highly celebrated paper, namely "The Moral Machine Experiment" by Awad et al. [10], marches forth to determine which decision the machine should be programmed to make if faced with the trolley dilemma. This article takes a critical look at the major recommendations of the paper by Awad at al. It argues that Awad et al. undoubtedly collect an impressive and philosophically useful data set

[1] In the Sufi tradition, for example, it has been said by the famous Sufi poet and saint, Rumi: "What I say now is ashamed to come here in front of you,…, but God will accept my trying as long as it is the best I can do." Rumi [5]. Similarly, in the Eastern traditions, *Karma* (right action, self-effort) that is in accordance with one's *Dharma* (duty) has been given the highest relevance, and is a pre-condition to attaining *Moksha* or *Nirvana* (self-realization/liberation). For example, in the *Yoga Vasistha* (a conversation between Lord Rama and his teacher, Sage Vasistha), it is said: "There is no power greater than right action in the present. Hence, one should take recourse to self-effort, grinding one's teeth, and one should overcome evil by good and fate by present effort." ([6], pp. 25–26)

of armchair intuitions. However, their expectation and intent that the findings based on an analysis of this data set can "contribute to developing global, socially acceptable principles for machine learning," if implemented or accepted by policy makers, would violate basic tenets of human rights law and fundamental principles of human dignity. This paper cites to principles of tort law, relevant case law, provisions from the Universal Declaration of Human Rights [11], and rules from the German Ethics Code for Automated and Connected Driving [12, 13] (to which Awad et al. [10] also refer, but override) to make its arguments.

The paper is arranged as follows: following this introduction in part 1, part 2 provides an overview of the key findings of the paper by Awad and colleagues. Part 3 then highlights the methodological, legal, and ethical concerns raised by the paper. Part 4 concludes with some recommendations for future research and innovation in the field of AVs.

The Moral Machine Experiment

Awad et al. [10], in their now famous "The Moral Machine Experiment," use a "multilingual online 'serious game' for collecting large-scale data[2] on how citizens would want AVs to solve moral dilemmas in the context of unavoidable accidents." In so doing, they presume that artificial intelligence (AI) in general, and AVs in particular, can and should be permitted to make "moral" decisions on behalf of or instead of human actors, including in situations involving human life. Indeed, if at least one of the long-term aims of AV technology is to reduce the number of deaths and injuries on roadways resulting from current limitations of human perception, observation, and response time, this presumption may indeed be legitimate. What is not quite so legitimate, however, is the extent to which Awad at al. expect their paper's findings to go. Specifically problematic, for three key reasons as discussed below, is the authors' suggestion that the paper's findings can help us "trace our path to universal machine ethics" [10] and "contribute to developing global, socially acceptable principles for machine ethics."

[2] The experiment's platform gathered "40 million decisions in ten languages from millions of people in 233 countries and territories."

Before getting into these three reasons, an overview of some of the key findings of "The Moral Machine Experiment" may be in order:

- Forty million decisions from millions of people in 233 countries and territories were collected.
- The territories were separated into three main clusters: Western (e.g., North America and many European countries), Eastern (e.g., Japan, Taiwan, Saudi Arabia), and Southern (e.g., Latin American countries of Central and South America).
- On a global basis, the strongest preferences observed are "sparing humans over animals, sparing more lives, and sparing young lives" [10].
- Results also showed broad differences in relative preferences when comparing participants in different countries: For example, the preference for sparing younger characters rather than older ones is much higher for countries in the Southern cluster compared to the Eastern cluster. "Only the (weak) preference for sparing pedestrians over passengers and the (moderate) preference for sparing the lawful over the unlawful appear to be shared to the same extent in all clusters." [10].

The "Moral Machine Experiment": Methodological, Legal, and Ethical Concerns

Methodological Concerns

The first concern is methodological: The framing of Awad et al.'s research [10], namely variations of hypothesized dilemma situations, does not represent reality [14]. Such scenarios are extremely simplistic and deterministic as a limited number of action choices and corresponding certain outcomes are assumed [15]. Furthermore, Awad et al.'s research elicits responses from a set of presumably "reasonable" (in the tort law sense) persons in a hypothetical scenario where no real human beings, nor any human identities (e.g., names of people) or human relationships are disclosed. What if the teenager, in real life, is the driver's daughter? Further, no real road circumstances or conditions were simulated. Most individuals have trouble recalling such driving incidents in real life and consequently may find it difficult to empathize with the situations at hand [16], and particularly so when asked to answer intuitively while sitting comfortably on their armchairs.

In fact, earlier studies have illustrated that mental processes are dependent on the extent to which a participant is immersed in a situation similar to driving (e.g., [17]). Therefore, the "cold" preferences that the participants reported in "The Moral Machine Experiment" may have resulted from their less ecological and more conceptual evaluation of the situation. Generally speaking, experimental philosophy may be critiqued for failing to investigate participant's true verdict about a case. Furthermore, it has been found that people's judgments about philosophical cases are influenced by contextual factors that are philosophically irrelevant, for example, by the order and presentation of cases (e.g., [18–20]). This may also explain some of the contradictory results in past literature: For example, in the study of Bonnefon et al. [21], it was found that participants in theory approved of utilitarian (and self-sacrificing) AVs but at the same time did not display willingness to consume such a vehicle for themselves. To overcome the limitations of current empirical ethics, Gigerenzer [22] has suggested the research program of ecological morality by assessing moral behavior (especially in more natural environments than laboratory settings) in addition to verbal or written indications. For example, research that utilizes virtual reality may emulate traffic incidents more realistically and hence give better insight into the true moral preferences and actions of drivers [23]. Furthermore, studies that consider and implement time pressure into the decision-making process may more realistically disclose snap judgments and intuitions of how participants would actually behave in real traffic situations. For example, it was found that under time constraints drivers decide differently, i.e., they fail to make utilitarian decisions, than when given enough time to deliberate in abstract manifestations of the trolley problem (e.g., [24]). Accordingly, responses to vague questions, such as a "jaywalking teenager" versus "three elderly ladies," elicited in entirely disconnected circumstances, are inappropriate candidates to be drawn on to construct moral aggregates which are then used to train "moral behavior" in machines and support law and policy decisions. It is relevant to recall the caveat (albeit made in the context of three famous economic models) from Elinor Ostrom in her Nobel winning work, "Governing the Commons" ([25], 6-7), where she says:

> *What makes these models so interesting and so powerful is that they capture important aspects of many different problems that occur in diverse settings in all parts of the world. What makes these model so dangerous – when they are used metaphorically as the foundation for policy – is that the constraints that are assumed to be fixed for the purposes of the analysis are taken on faith as being fixed in empirical settings, unless external authorities change them.*

Discussion on the use of experimental methods in philosophy also suggests that it is generally questionable whether such "folk intuitions" that are generated during controlled and very artificial conditions are generalizable to the real world [26] or at all have any philosophical significance as they are coming from "ordinary folks with no prior background in philosophy" [27]. Therefore, concerns have been raised about the extent to which experimental philosophy can tell us anything about (how to structure) the real world [26]. It could, accordingly, be argued that "The Moral Machine Experiment"—as an example of experimental philosophy—should not inform and direct us in determining what constitutes permitted behavior of AVs in actual dilemmas on the road.

Another methodological concern is the question whether the results could/should be considered as significant for actual technological development of AVs. Firstly, it can be assumed that at the time of the conducted experiment, individuals had limited experience with AVs. As the actual experience with self-driving cars increases, participants' attitudes and action choices may similarly change and become more prudent. If one were (at all) to allow participant's responses as legitimate measures, it would be important to not put too much weight on current attitudes (alone) but rather to constantly reassess attitudes in the future [15].

Secondly, acceptable options will be determined by technological capabilities and advances: for example, distinguishing between a criminal and a civic may currently not be (technically) feasible in reality [28]. Furthermore, making a choice on whom to spare, based merely on one's criminal record, is violative of fundamental human rights, as discussed below. Even the access to and use of such data (criminal records) for making an automated decision is questionable on legal and ethical grounds. Lastly, it needs to be questioned whether the assessment of such decision-making is necessary and expedient after all, since AVs should be designed in a way that such dilemma situations do not

arise in the first place, for example, by restructuring infrastructure and road usage, and prospectively adjusting the speed when confronted with a dangerous and uncertain situation [29] (Table 1).

Accordingly, at least the limitations of their approach should be made clearer by the authors. Indeed, the necessity to identify limitations of empirical research, including when it comes to determining "moral behavior," can be seen from the diversity of often counterintuitive conclusions reached by such research. For example, can we really assume that all or even a (significant) majority of "real" people would self-sacrifice in certain situations, as, for example, Huebner and Hauser [8] or Di Nucci [9] find in their vignette experiments? There are strong reasons and empirical evidence to doubt this.

The second reason stems from law, particularly human rights law and tort law, as discussed below.

Legal Concerns: Human Rights, Tort Law, and the German Ethics Code for AVs

Human Rights and the Non-derogable Right to Life

Fundamental human rights, enshrined under the Universal Declaration of Human Rights as well as the European Convention on Human Rights (ECHR), are broadly classified into the following three categories [30]:

1. Non-derogable, absolute rights
2. Non-derogable, non-absolute rights
3. Qualified rights

Table 1 Overview of *methodological* concerns

Methodological concerns
Unrealistic framing
• Simplified scenarios: limited number of action choices • Deterministic: certain outcomes • Disregard of human relationships • No real road circumstances or conditions simulated
Limited significance of results for implementation
• "Cold" rather than true moral preferences or intuitions • Preferences may change over time, with experience • Acceptable options are dependent on technological advances

Non-derogable, absolute rights are defined as those rights that "cannot be limited or infringed under any circumstance, not even during a declared state of emergency" [31]. They also cannot be suspended[3] [32–34]. Non-derogable, non-absolute rights are those rights whose ordinary application may be limited under *specific* circumstances [35] (e.g., the right to marry and start a family can be limited such that multiple marriages are outlawed).

Qualified rights are rights that "permit interferences subject to *various* conditions" (emphasis supplied). However, such interferences must be "in accordance with the law and necessary in a democratic state for the requirements of public order, public health or morals, national security or public safety" [36]. Thus, for example, the right to privacy is a qualified right, which can be interfered with if a search warrant has been granted by a court of law. The right to free speech can also be limited or interrupted if it encroaches upon another person's rights (defamation) or if it threatens national security by inciting violence. Similarly, the right to intellectual property, as well as the right to innovate are qualified rights, subject to several limitations, including limitations imposed in the interest of public order and morality [37].

The right to life, although not universally recognized as absolute [38], is recognized universally as a non-derogable right [39, 40]. Accordingly, its ordinary application can only be limited under *specific* circumstances. These circumstances are already defined under Article 2 of the European Convention on Human Rights, which states under paragraph (1):

> *Everyone's right to life shall be protected by law*. No one shall be deprived of his life intentionally save in the execution of a sentence of a court following his conviction of a crime for which this penalty is provided by law. (Emphasis supplied)

In addition to the limitation to the right to life affected by capital punishment (as envisaged in Article 2 paragraph (1)), the ECHR further mentions only three circumstances[4] under Article 2 paragraph (2), in which the

[3] The right to life, the right to be free from torture and other inhumane or degrading treatment or punishment, the right to be free from slavery or servitude, and the right to be free from retroactive application of penal laws are the four absolute non-derogable rights [32, 33].

[4] See the analysis in [41].

use of force resulting in the "[d]eprivation of life shall not be regarded as inflicted in contravention of Article 2", namely:

(a) In defense of any person from unlawful violence

(b) In order to effect a lawful arrest or to prevent the escape of a person lawfully detained

(c) In action lawfully taken for the purpose of quelling a riot or insurrection

Accordingly, the programming of AVs to always take the life of one category of persons, even in dilemma situations, is not envisaged as justifiable within the scope of Article 2.

More relevant in this regard are the opening words of Article 2.1 of the ECHR: "*Everyone's right to life shall be protected by law*." In this context, the provisions of Article 14 of the ECHR are relevant and must be read in conjunction:

> The enjoyment of the rights and freedoms set forth in this Convention shall be secured without discrimination on any ground such as sex, race, colour, language, religion, political or other opinion, national or social origin, association with a national minority, property, birth or other status.

Similar to the wording of Article 14 of ECHR, Article 2 of the Universal Declaration of Human Rights (UDHR) also states that every human being is entitled to *all* rights and freedoms (including the right to life) "without distinction of any kind, such as race, colour, sex, language, religion, political or other opinion, national or social origin, property, place of birth or other status (Article 2)." Further, Article 7 of the UDHR provides that "all are equal before the law and are entitled without any discrimination to equal protection of the law (Article 7)."

In other words, when we read the right to life (Article 2 of ECHR) together with the non-discrimination clause (Article 14 of ECHR and Article 2 of the UDHR), *all* human beings, irrespective of age, religions, race, gender, or nationality, are equally entitled to the right to life and the law cannot endorse any practice (even when agreed to by the majority or followed for the sake of convenience) which, either in effect or at the outset, compromises the lives of one category of persons in preference over another category of persons.

In the context of dilemma situations in AVs, the fact that every human being has an equal right to life and is entitled to have the law protect this right would require that the law does not permit AVs to be programed in such a way that a specific category of persons is preferentially compromised or spared in any situations that arise or may arise (in the future). By programming a vehicle, based on the findings of "The Moral Machine Experiment," to always strike an old person in country A or always strike a child in country B would violate the right to life of old people in country A and the right to life of children in country B. Beyond human rights law, under civilized convention as also age-old moral law, all lives are equally valuable [42].[8] Further, in natural circumstances (i.e., when an automobile faced with a dilemma situation is not programed to always hit a specific category of persons), each human being that is involved in any dilemma situation will have an equal chance of surviving. Would it be legally or ethically justifiable to permit a program to reduce this 50–50 chance of survival to a 0% chance of survival for a specific category of persons?[5]

It is in the context of human rights and the classification of the right to life as non-derogable that one must read Rule 9 of the German Ethics Code for Automated and Connected Driving. Indeed, Awad at al. refer to Rule 9 in their paper, but do not comprehensively examine its rationale and scope. Rule 9 is consistent with fundamental human rights, as well as the manner in which these rights are legitimately categorized. It states:

> In the event of unavoidable accident situations, any distinction based on personal features (age, gender, physical or mental constitution) is strictly prohibited. It is also prohibited to offset victims against one another. General programming to reduce the number of personal injuries may be justifiable. Those parties involved in the generation

[5] It is to be noted that here the authors are not making an argument in favor of AVs randomizing the decision for each time it faces a dilemma situation. Instead, and as discussed later, the effort here is to underscore the importance of AVs research focusing on means of avoiding dilemma situations altogether. Further, the aim is also to encourage empirical research in a different direction: empirical data should not be used to determine issues as fundamental and non-derogable as the right to life. Instead, empirical research can look into the possibility of gaining conditional and ephemeral access to data—data protection being a qualified right—such that the data/information can then be used to avoid the emergence of dilemma situations altogether.

of mobility risks must not sacrifice non-involved parties. [12, 13].

In effect, therefore, the recommendations of Awad et al. [10], that their paper's empirical data be used to guide AV decision-making in dilemma situations, if adopted by any original equipment manufacturer (OEM), would violate the basic principles of the UDHR and the German Ethics Code, as well as age-old moral law.

Tort Law, Empirical Data, and the "Reasonable Person" Standard

Torts, or civil wrongs, are remedied or compensated following basic principles of tort law. These principles have evolved over centuries, primarily through case law (decided court decisions). Torts are broadly classified as intentional and unintentional torts. Unintentional torts are also known as cases of "negligence." Negligence is defined as "the omission to do something which a *reasonable man* would do, or doing something which a *reasonable and prudent man* would not do." (emphasis supplied). In any case of negligence, a person (which includes a real as well as artificial person, e.g., corporations or industries) "is liable for harm that is the *foreseeable consequence* of his or her actions." ([43], 91)

When a person (the injured party) files a lawsuit (court case) for negligence, he/she must prove that (a) the alleged tortfeasor (defendant) owed a duty of care to the injured party (plaintiff); (b) the defendant breached this duty of care; (c) the plaintiff suffered injury; (d) the negligent act (or omission) caused the plaintiff's injury, and (e) the defendant's negligent act (or omission) was the proximate cause of the plaintiff's injury. ([43], 91)

In the landmark negligence case, *Donahue* v. *Stevenson* [44],[6] Lord Atkin developed the "neighbor principle" which significantly expanded the scope of the tort of negligence and the duty of care owed under it. The key finding of the court in this case was

> The rule that you are to love your neighbor becomes in law, you must not injure your neighbor; and the lawyer's question, Who is my neighbor? receives a restricted reply. The answer seems to be – persons who are so closely and directly affected by my act that I ought *reasonably* to have them in contemplation as being so affected when I am directing my mind to the acts or omissions which are called in question. (emphasis supplied)

With this tort law primer in mind, let us imagine then that a car C is programmed by a car manufacturer M, according to the recommendations of Awad et al.'s research [10]. Let us further presume that C is purchased by a person P in country B (see above, where we presume that following Awad et al.'s recommendations, country B is one where AVs are programmed to ensure that in dilemma situations, elderly persons are always spared). Now imagine that car C, while driving on the roads of country B, faces a dilemma situation and following its program diligently, hits a child and causes its death. Let us assume, as is likely to be the case, that the child's parents sue the owner of the car, P, for negligence and the car's manufacturers, M, are sued alongside. What might the decision of the court be?

It is relevant to note at the outset that, to the authors' knowledge, no court has had an opportunity so far to consider such a matter (or one with a similar fact scenario). We can safely say that there is currently a great deal of uncertainty about the manner in which the "duty of care" under tort law would be interpreted and distributed in cases involving fully automated vehicles, and more particularly, in cases involving AVs facing dilemma situations. However, applying the basic principles of tort law as described above, we expect the following will be considered by the courts in Europe, (and perhaps also courts of other common and civil law countries that follow basic principles of equity, fairness and justice.)

Coming back to our fact scenario, the first thing one may ask is: Is P liable for negligence? Probably not—P had no choice but to resign to the pre-programmed car's decision to hit the child. All (s)he could do in the circumstance was wait and watch.[7] Even if one were to argue that P had the choice of not buying an AV at all, once the described situation occurs, the court will look for the "proximate cause" (see discussion above). In our case, it is the manner in which the car was programmed and not the fact of the car's purchase, which is the

[6] 1932 S.C. (H.L.) 31

[7] We are presuming in our illustration here that all available AVs in P's country have been programed to hit a child in case of a dilemma situation involving children versus the elderly (according to the findings of Awad et al.'s paper), and that P has no say in the matter at the time of purchasing the car or thereafter.

proximate cause of the harm caused[8] [45, 46] (referring to Polemis: In re Arbitration Between Polemis and Furness, Withy & Co., Ltd. [1921] 3 K.B. 560).

Is M then liable for negligence? The answer to this question, first, and most importantly, the court will consider the duty of care and who owed this duty to whom. As stated earlier, negligence is defined as "the omission to do something which a *reasonable man* would do, or *doing* something which a *reasonable and prudent man would not do*." The question that arises, therefore, is who is a *reasonable man* and how would he/she have behaved in a similar circumstance. Can an action/inaction be deemed reasonable (or otherwise) based on empirical evidence or the opinions of many (other reasonable persons)?

The "reasonable person," or the person who has been famously described as "the man on the Clapham Omnibus," is not a real person, but rather a creature of legal fiction. Applying the "reasonable person" standard or test has been likened to applying the "Golden Rule" in both common and civil law systems [47]. A person's behavior is deemed reasonable before the eyes of the law (in a manner similar to behavior that is deemed "ethical"), when it is "fair, just or equitable. The person must be honest, moderate, sane, sensible, and tolerable" [47–50]. At the outset, it can be reiterated that no clear, universal consensus has emerged from any research so far conducted (including from Awad et al.'s research [10]) as to what choice a reasonable person would make, in *any* (dilemma) circumstance. It is also not at all clear that the decision would remain the same in *all*, including similar, (dilemma) situations, if they were to be faced multiple times.

Secondly, the terminology ("reasonable person") is rather vague, as is the term "moral person." Indeed, perhaps because the term's interpretation is heavily dependent on context, courts have warned against attempting to define what this hypothetical "reasonable person," a construct of legal fiction, would do, by attempting to lead empirical evidence in court through, for example, testimony of other "reasonable persons." In the recent case *Healthcare at Home Limited* v. *The Common Services Agency* [51], the UK Supreme Court stated as follows:

> It follows from the nature of the reasonable man, as a means of describing a standard applied by the court, that it would be misconceived for a party to seek to lead evidence from actual passengers on the Clapham omnibus as to how they would have acted in a given situation or what they would have foreseen, in order to establish how the reasonable man would have acted or what he would have foreseen. Even if the party offered to prove that his witnesses were reasonable men, the evidence would be beside the point. The behavior of the reasonable man is not established by the evidence of witnesses, but by the application of a legal standard by the court. The court may require to be informed by *evidence of circumstances* which bear on its application of the standard of the reasonable man in *any particular case*; but it is then for the court to determine the outcome, in those circumstances, of applying that impersonal standard. [51].

Court decisions are valuable guiding/check posts, including in the currently under-regulated AV domain [52], not least because cases of human injury, whether caused by human or machine actors, will eventually boil down to questions of legal liability. Because legal liability in tort law, in both civil and common law countries, is based on the "reasonable person" standard, it may be safe to presume, based on the above court decision, that using empirical data to determine "socially acceptable principles" may be highly problematic in a court of law, especially when dealing with questions of legal liability.

The above court ruling, therefore, calls to question research (as done by Awad et al. [10]) aimed at recommending the programming of "reasonable" and even "moral" behavior in machines based on empirical survey data. Pre-determining "moral" or "reasonable" behavior based on such statistical or empirical data would also give out uncomfortable, inaccurate, and illegal signals as to how much specific categories of life are valued by a region or country, potentially creating a

[8] Courts have come up with several tests to determine "proximity." Two well-known tests are the "foreseeability test" [45] and the test of directness [46]. Under the foreseeability test, the court considers whether the damage or harm that occurred, foreseeable by the defendant's act (or omission). In our hypothetical, by programming AVs in such a way, the harm is not only foreseeable, but inevitable once a dilemma situation arises. In the test for directness, the court asks whether there is direct connection or link between the action (or inaction) of the defendant and the damage or harm that occurred to the plaintiff. Here again, in our hypothetical, the action of programming the AV to always strike the child in the country B is directly responsible for and connected to the harm caused to the plaintiff.

chain of other undesirable socio-cultural consequences. Policy and codified law, including computer codes, which increasingly play the role of law, must seek to implement higher moral goals rather than perpetuating aggregated individual preferences (which, in the worse of cases, may be reflective of individual prejudices), of any statistical majority.

Pre-determining the outcome of any accident by programming AVs in a specific way based entirely on what a large number of presumably reasonable persons have said in response to an empirical or experimental survey/study, is, therefore, most likely to be considered a violation of duty of care under torts. Such programming would also not accurately track, and would perhaps also disincentivize, manufacturers' overall duties vis-à-vis ensuring vehicular safety and readiness to face complex driving scenarios.

Once a duty of care and its violation are established, the other elements of negligence clearly follow in our hypothetical trolley dilemma situation. Because of the way the AV was programmed, the child (or the elderly) was injured/sacrificed. If not *solely* responsible, the programming was at least *foreseeably and directly* responsible for causing the injury. Had it not been for the programming, the person would have had at least a 50% chance of survival. Undoubtedly, programming an AV in this manner would result in placing all liability in the hands of the car manufacturer—several perplexing questions associated with liability in case of accidents involving AVs may then disappear [53]. In fact, it has been opined that individuals may then *prefer* AVs at level 5 automation, in order to be completely absolved from any responsibility and, therefore, also from any liability [53]. This is problematic for several ethical reasons, as discussed in the following section.

Such empirical data-based programming may even result in a finding of strict liability (for M—the equipment manufacturer), because such programming arguably results in the AV becoming "inherently dangerous" for specific categories of persons, for example, for children in country B. It would also be a willful violation of the non-derogable right to life of children (or of the elderly, or even of those with criminal records), as discussed above. This would be the case even if the majority of the surveyed people in country B were to find the programming and its resulting casualty "reasonable" or "moral."

From a legal perspective, it is also noteworthy that in a trolley dilemma situation, no matter what split second decision a human driver makes, (s)he is likely to *not* be held liable either for negligence or under principles of strict liability (unless mala fides or recklessness is proved, in which case, we would go out of the purview of negligence under tort law and towards criminal liability).

If, however, the recommendations of Awad et al. [10]. are used to design AV policy and regulation that permit car manufacturers (OEMs) to program AVs in the above manner (e.g., in country B), the express legislation would override principles of tort law that have evolved over centuries in close compliance with fundamental principles of equity, fairness, and responsibility. Such a legislation would result in making no one liable (neither the OEM nor the car owner) and leave the injured party, and indeed the entire class of persons affected by the policy-based programming, without remedy.

In this context, it is also relevant to highlight the paper's statement that "In summary, the individual variations that we observe are theoretically important, but not essential information for policymakers [10]." Imposing an alleged "Universal ethic" on individuals whose personal (and very justifiable) moral conscience would urge them to act differently in dilemma situations is also a violation of the individual's right to freedom of conscience (Article 10 of the European Charter on Fundamental Rights) and the right to conscientious objection. The armchair "conscience" of an alleged "majority" cannot be used to impose allegedly conscientious choices on the whole. Not everyone who buys or uses AVs programed in such a way would consider them as "moral machines."

In the context of the human right to "equality" and "liberty," it has been said that:

> *This understanding of the equality of all human beings leads "naturally" to a political emphasis on autonomy. Personal liberty, especially the liberty to choose and pursue one's own life, clearly is entailed by the idea of equal respect. For the state to interfere in matters of personal morality would be to treat the life plans and values of some as superior to others.*
>
> […]
> *Autonomy (liberty) and equality are less a pair of guiding principles – let alone competing principles – than different manifestations of the central commitment to the equal worth and dignity of*

> *each and every person, whatever her social utility…. Equal and autonomous rights-bearing individuals…. have no right to force on one another ideas of what is right and proper, because to do so would treat those others as less than equal moral agents. Regardless of who they are or where they stand, individuals have an inherent dignity and worth for which the state must demonstrate an active and equal concern. And everyone is* ***entitled*** *to this equal concern and respect…* ([54], 63)

Awad et al. themselves recognize that their sample is close to the internet-connected, tech savvy population that is interested in driverless car technology [10]. While the authors seem to suggest that the opinions of this group are rather (more) important as they will be the early adopters of the technology, it must also be borne in mind that in implementing the choice of the majority in the manner recommended (e.g., all old people in country A or all young people in country B), the morality of the tech savvy will be imposed on the population as a whole—especially the weak and the poor who, while not being inside the AV, may be the victims of its generalized programming. Legislation based on such statistical findings would be scary at the very least, and would be in violation of the right to equal treatment under the law (Article 7, UDHR). Here again, Rule 9 of the German Ethics Code, which states that "Those parties involved in the generation of mobility risks must not sacrifice non-involved parties," is centrally relevant and must be borne in mind.

Empirical Research, Data Protection, and AV Innovation

Are we then suggesting that all empirical and experimental research be abandoned? Not at all! Unlike the right to life, the right to personal data and the right to privacy are qualified rights [39, 55]. In the EU, Recital 4 of the General Data Protection Regulation (GDPR) also states that "The right to protection of personal data is not an absolute right; it must be considered in relation to its functionality in society and be balanced against other fundamental rights, in accordance with the principle of proportionality." [56] This recital paves the way for the use of data, including personal data, in AVs (for example, for the creation of "ethical black boxes" [57] or for constructing necessary V2x systems [58]), especially when necessary to secure the lives of persons inside and outside AVs.

Research into the instinctive willingness of human agents to permit the use of non-personal, and in some (limited and controllable) instances, personal data, in order to enhance the safety of persons inside and outside AVs would be an example of where experimental or even empirical data collected using the methodology of Awad et al., may more legitimately "contribute to developing global, socially acceptable principles for machine ethics" [10, 42]. Indeed, studying the willingness of human agents to share their data, especially when limited by time and purpose, and for securing lives of persons inside and outside AVs, is not only necessary for constructing efficient V2x systems, but may become critically necessary to avoid the emergence of dilemma situations in the first place [58]. The data so collected can then be used for the organization of road infrastructure, including IoT devices, which can give signals of danger to the AV well before the humans inside or outside the vehicle become aware of the existence of danger.

Undoubtedly, to arrive at this juncture, several questions and issues need to be addressed [53]. Further, several divergent human rights and responsibilities need to be examined and balanced out—but this is a topic that requires undivided attention in a separate paper.

Ethical Concerns

Similar to previously stated laws such as the equal right to life and human dignity, "[c]lassical liberalism acknowledges that while people are factually unequal, people ought to be treated as political equals by the government. Under state authority no individuals or classes of people should be assigned greater or lesser worth compared to others" [54, 59]. Programming machines to make "moral" decisions based on personal physical features such as age and gender is accordingly also ethically discarded and prohibited by Rule 9 of the German Ethics Code for Automated and Connected Driving, (as well as by the UDHR and tort law's duty of care).

Furthermore, there have been calls for the recognition and declaration of a universally applicable set of human values [60–62] that all of humanity shares in common. These include fairness, honesty, integrity, responsibility, compassion, friendliness, and several others. Yet, the individual, sincere manifestations or expressions of these ethical values may legitimately vary—indeed,

they *should* be permitted to vary, from person to person and situation to situation. For example, in a dilemma situation, it may be equally justifiable, from an individual moral (as well as legal) standpoint, for a driver to spare either the young or the old. The higher ethical value to be safeguarded here is the right and the responsibility of each driver (or car owner) to choose and take some concrete action (see also in this context endnotes 1, 8 and accompanying text above; [5, 6])—at least as long as the driver officially constitutes the moral agent.

Some scholars have proposed that drivers should be allowed to select an ethical program or weighting from a menu of acceptable options (e.g., [62, 63]). They also argue that it *should* be prohibited to modify one's settings in such a selfish way that results in an unjust distribution of harms [63]. Gogoll and Müller [64] argue that this, however, is exactly what would happen due to a prisoner's dilemma if no single universal ethics setting for AVs was implemented. This argument, however, does not take into account the functioning of and factors determining liability in negligence cases (as discussed above. See also quotes from Elinor Ostrom above and below). Liability in cases of negligence (unintended harm) is determined based on the standards of a reasonable person—if the individual person's settings (and behavior leading up to the dilemma situation) are reasonable, then on a case by case basis, the individual would not be held liable and would, therefore, not be in a prisoner's dilemma situation at all. Although individual settings (especially if these are applicable also when the driver is in the car) may still be problematic from a human rights and responsibilities perspective, they will at least have fewer implications on the democratic setup of any country and the equal treatment, mandated by law, of all individuals. It will also ensure that human being (and not machines) need to take responsibility for (and face the consequences of) their actions and setting choices—not just in a legal liability sense (which, as stated, may be limited), but also in the context of a personal moral conscience.

This brings us to another reason for our discomfort with the recommendations of Awad et al.'s [10] paper, which is also based on one of the fundamental principles of ethics, namely, responsibility. On the one hand, the presented scenarios in "The Moral Machine Experiment" omit the disclosure and attribution of responsibility. For example, no information is provided on the underlying reasons why the vehicle arrived in the particular dilemma: Was the driver previously inattentive or driving too fast? This ignorance is mainly the result of the methodological concerns (such as the unrealistic/undetailed framing) mentioned in "Methodological Concerns." Therefore, it can be argued that it remains intransparent as to who was responsible in the first place and thus maybe should shoulder (more) responsibility [65].

In this regard, scholars have advocated a "responsibility-sensitive safety model" that takes into consideration factors such as safe distance or proper response executed by the driver when determining reasonable care and accountability on the road [66]. The paper "The Moral Machine Experiment" does not provide information of such behavior and decisions. Its recommendations are also for this reason not usable in real world scenarios, including for attribution or distribution of responsibility and liability.

Even if complete information about the behavior and decisions that led to the dilemma were disclosed in the experiment, the internal limitations of models used in experimental philosophy would prevent participants from exercising their "free will" and assume responsibility in an unfettered manner—as they would be able to in the real world [67]. In fact, some experimental philosophers question whether free will and moral responsibility can at all exist if determinism is assumed[9] [27]. In other words, if experimental methods in philosophy were framed in a deterministic way, free will and moral responsibility could not prevail. Indeed, one could argue that the scenarios in "The Moral Machine Experiment" represent an extremely deterministic universe, not in the sense that cause and effect is bound by physical law but rather that the actions and outcomes available to the participants are limited, clear, and fixed (e.g., keep lane and consequently kill five individuals or steer left and consequently kill one individual). Thus, if policy makers were to adopt recommendations emerging from such

[9] It may be relevant to note in this context that even moral philosophies across the globe do not (all) deny such a "deterministic" nature of the universe. See footnote 1 and the quote from the *Yoga Vasistha* ([6], pp. 25–26) above. The *Yoga Vasistha* further states (p. 26), for example, "If you see that the present self-effort is sometimes thwarted by fate (divine will), you should understand that the present self-effort is weak. A weak and dull-witted man sees the hand of providence when he is confronted by a strong and powerful adversary and succumbs to him." Further (p. 27–28), "Fate or divine dispensation is merely a convention which has come to be regarded as truth by being repeatedly declared to be true. If this god or fate is truly the ordainer of everything in this world, of what meaning is any action…, and whom should one teach at all? No. In this world, except a corpse, everything is active and such activity yields its appropriate result…. Hence, renounce fatalism and apply yourself to self-effort." [6].

experimental methods, they would, in effect, be excluding the possibility of human beings exercising free will and assuming responsibility. Yet, such free and responsible actions are exactly the ones that should be accounted for in real traffic accidents. In this context, it is relevant to turn, once again, to what Elinor Ostrom said in the context of the prisoner's dilemma:

> *The prisoners in the famous dilemma cannot change the constraints imposed on them by the district attorney; they are in jail.... As long as individuals are viewed as prisoners, policy prescriptions will address this metaphor. I would rather address the question of how to enhance the capabilities of those involved to change the constraining rules of the game to lead to outcomes other than remorseless tragedies.*

Programming machines to take decisions in *all* situations, especially situations involving human life when there is a human decision-maker also available, could lead to an inclination to fatalistically accept the allegedly "moral" decision of the machine. This could, over time, have a significant impact on human values associated with the eagerness to assume responsibility and have a say in critical decision-making. Some trends in this direction have already been noted [53, 68]. As of now, as mentioned above, going beyond armchair intuitions, experimental ethics research has found that people are hesitant to delegate tasks that affect third parties, to machines [4]. These findings are not counterintuitive and are in line with broader concepts of human dignity and liberty, as well as philosophical foundations of morality in diverse cultures ([5, 6], endnotes 1 and 8), that give fundamental importance to human willingness to choose and to take responsibility, or at least be actively involved in critical decision-making using free will and reason [69]. When taking these rights out of the hands of human beings, it is of indispensable importance to ensure that the basis on which these rights are re-delegated to machines are not in themselves contrary to fundamental and universal human rights and values (as discussed above) (Table 2).

Conclusion

Keeping in mind a not-so-distant world where complete vehicular automation is inevitable and perhaps also desirable, research on how best to arrange (physically, as well as from a legal and ethical standpoint) technology such as sensors, software as well as infrastructure in a way that best secures *all* human lives, should be a priority. The "dilemmas" that would then need to be resolved would likely involve much more mundane decisions than the ones envisaged by "The Moral Machine experiment." Such decisions can be (more) legitimately assigned to machines. Nevertheless, in such decisions also, several important trade-offs may have to be made, and human values and aesthetics will have to be considered [2]. Each such decision for programming AI software may, therefore, also need to be examined from ethical and legal (in addition to participatory) perspectives.

While innovation may well be an end in itself while it is limited to laboratory and garage experiments, when it starts impeaching the rights of specific categories of persons, we, as a society, must beware of a tendency to rely on numbers and statistics for a quick solution.

Undoubtedly, Awad and colleagues conducted a legitimate, highly informative and worthwhile philosophical exercise that highlights the intuitions of individual persons from diverse nationalities and walks of life when it comes to moral decision-making. However,

Table 2 Overview of legal and ethical concerns

Legal concerns	Ethical concerns
Disregard of equal right to life, human dignity	Disregard of equal moral worth
Discrimination	
Avoidance of legal standards: • Reasonable person standard • Duty of care and foreseeability of harm • Proximate cause • Strict liability	Inability to attribute responsibility
Limited right to freedom of conscience, right to conscientious objection	Limited self-determination

while engaging an impressive set of experts, bringing together insights from various disciplines, the final recommendations have ignored insights from two centrally significant disciplines, namely, law and ethics. As discussed above, fundamental principles and rules from these disciplines must be borne in mind by any AI program in order for it to be capable of safeguarding human values, human dignity, and basic human responsibility in the long run. Indeed, popular acceptance and adoption of the technology also depend on this.

Acknowledgments The authors would like to thank Dr. Matthias Uhl, Technical University of Munich, for his comments on a previous version of this article. We also acknowledge the research assistance of Ms. Julia Köninger.

Funding Open Access funding enabled and organized by Projekt DEAL.

References

1. Kyriakidis M, de Winter JCF, Stanton N, Bellet T, van Arem B, Brookhuis K, Martens MH, Bengler K, Andersson J, Merat N, Reed N, Flament M, Hagenzieker M, Happee R (2019) A human factors perspective on automated driving. Theor Issues Ergon Sci 20(3):223–249
2. Mladenović MN, Lehtinen S, Soh E, Martens K (2019) Emerging urban mobility technologies through the lens of everyday urban aesthetics: case of self-driving vehicle. Essays Philos 20(2):1526–0569
3. Lehtinen ST, Vihanninjoki VJ (forthcoming) Aesthetic perspectives to urban technologies: conceptualizing and evaluating the technology-driven changes in the urban everyday experience. In: Nagenborg M, Stone T, González Woge M, Vermaas PE (eds.) Technology and the city: towards a philosophy of urban technologies. Springer
4. Gogoll J, Uhl M (2018) Rage against the machine: automation in the moral domain. Journal of Behavioral and Experimental Economics 74:97–103
5. Barks C (1995) The essential Rumi: new expanded edition. HarperCollins, New York
6. Venkatesananda S (1993) Visistha's yoga. State University of New York Press, Albany
7. Bonnefon J-F, Shariff A, Rahwan I (2016) The social dilemma of autonomous vehicles. Science. 352(6293):1573–1576
8. Huebner B, Hauser MD (2011) Moral judgments about altruistic self-sacrifice: when philosophical and folk intuitions clash. Philos Psychol 24(1):73–94
9. Di Nucci E (2013) Self-sacrifice and the trolley problem. Philos Psychol 26(5):662–672
10. Awad E, Dsouza S, Kim R, Schulz J, Henrich J, Shariff A, Bonnefon JF, Rahwan I (2018) The moral machine experiment. Nature 563:59–64. https://doi.org/10.1038/s41586-018-0637-6
11. Assembly UG (1948) Universal declaration of human rights. UN General Assembly 302(2)
12. Di Fabio U, Broy M, Brüngger RJ (2017) Ethics commission automated and connected driving. Federal Ministry of Transport and Digital Infrastructure of the Federal Republic of Germany
13. Lütge C (2017) The German ethics code for automated and connected driving. Philos Technol 30(4):547–558
14. Himmelreich J (2018) Never mind the trolley: the ethics of autonomous vehicles in mundane situations. Ethical Theory and Moral Pract 21:669–684. https://doi.org/10.1007/s10677-018-9896-4
15. Nyholm S (2018) The ethics of crashes with self-driving cars: a roadmap, I. Philos Compass 13:e12507. https://doi.org/10.1111/phc3.12507
16. Goodall NJ (2016) Away from trolley problems and toward risk management. Appl Artif Intell 30:810–821. https://doi.org/10.1080/08839514.2016.1229922
17. Madary M, Metzinger TK (2016) Real virtuality: a code of ethical conduct. Recommendations for good scientific practice and the consumers of VR-technology. Front Robot AI 3:3. https://doi.org/10.3389/frobt.2016.00003
18. Petrinovich L, O'Neill P (1996) Influence of wording and framing effects on moral intuitions. Ethol and Sociobiol 17(3):145–171. https://doi.org/10.1016/0162-3095(96)00041-6
19. Swain S, Alexander J, Weinberg JM (2008) The instability of philosophical intuitions: running hot and cold on truetemp. Philos and Phenomenol Res 76(1):138–155. https://doi.org/10.1111/j.1933-1592.2007.00118.x
20. Wright JC (2010) On intuitional stability: the clear, the strong, and the paradigmatic. Cognit 115(3):491–503. https://doi.org/10.1016/j.cognition.2010.02.003
21. Bonnefon JF, Shariff a RI (2016) the social dilemma of autonomous vehicles. Sci 352:1573–1576. https://doi.org/10.1126/science.aaf2654
22. Gigerenzer G (2010) Moral satisficing: rethinking moral behavior as bounded rationality. Top in Cogn Sci 2(3):528–554. https://doi.org/10.1111/j.1756-8765.2010.01094.x
23. Grasso GM, Lucifora C, Perconti P, Plebe A (2020) Integrating human acceptable morality in autonomous vehicles. In: Ahram T et al (eds) Intelligent human systems integration 2020. Springer, Cham, pp 41–45
24. Samuel S, Yahoodik S, Yamani Y, Valluru K, Fisher DL (2020) Ethical decision making behind the wheel – a driving simulator study. Transp Res Interdiscip Perspect 5:100147. https://doi.org/10.1016/j.trip.2020.100147

25. Ostrom E (2011) Governing the commons, Cambridge University Press, Cambridge (29th print.) [orig. 1990]
26. Rusch H (2014) Philosophy as the behaviorist views it? Historical parallels to the discussion of the use of experimental methods in philosophy. In: Lütge C et al (eds) Experimental ethics. Palgrave Macmillan, London, pp 264–282
27. Knobe J, Nichols S (2017) Experimental philosophy. The Stanford encyclopedia of philosophy. Available from: https://plato.stanford.edu/entries/experimental-philosophy/
28. Davnall R (2020) Solving the single-vehicle self-driving car trolley problem using risk theory and vehicle dynamics. Sci and Eng Eth 26:431–449. https://doi.org/10.1007/s11948-019-00102-6
29. Goodall N (2019) More than trolleys: plausible, ethically ambiguous scenarios likely to be encountered by automated vehicles. Transfers 9:45–58. https://doi.org/10.3167/TRANS.2019.090204
30. Mavronicola N (2017) Is the prohibition against torture and cruel, inhuman and degrading treatment absolute in international human rights law? A reply to Steven Greer. Hum Rights Law Rev 17(3):479–498
31. Charter of Fundamental Rights of the European Union (1950) Note:2. Available from: https://ec.europa.eu/home-affairs/content/fundamental-rights_en
32. Absolute Rights, Available from: https://www.ag.gov.au/RightsAndProtections/HumanRights/Human-rights-scrutiny/PublicSectorGuidanceSheets/Pages/Absoluterights.aspx
33. UNTERM The United Nations Terminology Database: Non-derogable right [cited Cited. Available from: https://unterm.un.org/UNTERM/portal/welcome
34. Farer T (1992) The hierarchy of human rights. Am UJ Int'l L & Pol'y 8:115
35. Charter of Fundamental Rights of the European Union (1950) Note:3
36. Charter of Fundamental Rights of the European Union (1950) Note:4
37. World Trade Organization (1995) Part II — Standards concerning the availability, scope and use of intellectual property rights. Available from: https://www.wto.org/english/docs_e/legal_e/27-trips_04c_e.htm
38. Australian Government, Attorney-General's Department (n.y.) Absolute rights. Public sector guidance sheet. Available at: https://www.ag.gov.au/rights-and-protections/human-rights-and-anti-discrimination/human-rights-scrutiny/public-sector-guidance-sheets/absolute-rights
39. Koji T (2001) Emerging hierarchy in international human rights and beyond: from the perspective of non-derogable rights. European Journal of International Law 12:917–941
40. ECHR (195), Article 2(1) Right to life
41. Mendelsen D, Bagaric M (2013) Assisted suicide through the prism of the right to life. Int J Law Psychiatry 36(5–6):406–418
42. European Union (2000) Charter of fundamental rights of the European Union, Art. 21(1) Non-discrimination. European Union 2012/C 326/02
43. Cheeseman H (2013) Business law (8th edn.). Pearson, New York
44. Donahue v. Stevenson (1932) S.C. (H.L.) 31
45. Overseas Tankship (UK), Ltd. v. Morts Dock & Eng'g Co. (The Wagon Mound No. 1) [1961] 1 All E.R. 404 (Privy Council Austl)
46. Zipursky BJ (2016) Proximate cause in the law of torts. Available at https://lawexplores.com/proximate-cause-in-the-law-of-torts-2/
47. Joachim W (1992) The "reasonable man" in United States and German commercial law. Comparative Law Yearbook of International Business 15:341–362
48. Black HC, Garner BA, BR MD, Schultz DW, Company WP (1999) Black's law dictionary. West Group, St. Paul
49. Thatcher VS, McQueen A (1971) The new Webster encyclopedic dictionary of the English language. Consolidated Book Publishers, Chicago
50. Webster N, McKechnie N (1975) Webster's new twentieth century dictionary of the English language unabridged (2nd edn.). Collins World, New York
51. Healthcare at Home Limited v. The Common Services Agency (2014) UKSC:49. Available from: https://www.supremecourt.uk/cases/uksc-2013-0108.html
52. Claybrook J, Kildare S (2018) Autonomous vehicles: no driver… no regulation? Science 361:36–37
53. Ryan M (2019) The future of transportation: ethical, legal, social and economic impacts of self-driving vehicles in the year 2025. Sci Eng Ethics:1–24
54. Donnelly J (2013) Universal human rights (3rd edn.). Cornell University Press
55. Bergelson V (2003) It's personal but is it mine-toward property rights in personal information. UC Davis L Rev 37:379
56. European Union (2018) General Data Protection Regulation (GDPR). European Union
57. Winfield AF, Jirotka M (2017) The case for an ethical black box. In: Annual Conference Towards Autonomous Robotic Systems. Springer, Cham, pp 262–273
58. Schmidt T, Philipsen R, Ziefle M (2016) Dos and don'ts of datasharing in V2X-technology. In: Helfert M et al. (eds) Smart Cities, Green Technologies, and Intelligent Transport Systems. Springer, Cham, pp 257–274
59. Gentzel M (2019) Classical liberalism, discrimination, and the problem of autonomous cars. Sci and Eng Eth 26:931–946. https://doi.org/10.1007/s11948-019-00155-7
60. InterAction Council (1997) A universal declaration of human responsibilities. Proposed by the InterAction Council. InterAction Council, Tokyo
61. Solomon A (1998) Draft declaration of human values. Humanist in Canada 31(3)
62. Shankar SSR (2007) Universal declaration of human values (proposal). Available from: https://www.iahv.org/us-en/wp-content/themes/IAHV/PDF/Universal-Declaration-of-Human-Values.pdf
63. Jenkins R (2016) Autonomous vehicles ethics & law. New America. Available from https://d1y8sb8igg2f8e.cloudfront.net/documents/AV-Ethics-Law.pdf. Accessed 12 May 2020
64. Gogoll J, Müller JF (2017) Autonomous cars: in favor of a mandatory ethics setting. Sci and Eng Eth 23:681–700. https://doi.org/10.1007/s11948-016-9806-x
65. Mirnig AG, Meschtscherjakov A (2019) Trolled by the trolley problem: on what matters for ethical decision making in automated vehicles. In: Proc of the 2019 CHI Conf on hum factors in computing Syst, pp 1–10. https://doi.org/10.1145/3290605.3300739

66. Shalev-Shwartz S, Shammah S, Shashua A (2017) On a formal model of safe and scalable self-driving cars. arXiv preprint arXiv:1708.06374
67. Brandão M (2018) Moral autonomy and equality of opportunity for algorithms in autonomous vehicles. In: Coeckelbergh M et al (eds) Envisioning robots in society – power, politics, and public space. IOS Press BV, Amsterdam, pp 302–310
68. CNIL (2017) How can humans keep the upper hand? The ethical matters raised by algorithms and arti cial intelligence. Report on the public debate led by the French Data Protection Authority (CNIL) as part of the ethical discussio assignment set by the Digital Republic bill, December 2017, https://www.cnil.fr/sites/default/les/atoms/les/cnil_rapport_ai_gb_web.pdf
69. Kant I (2018) Fundamental principles of the metaphysic of morals (transl. by Abbott TK). Litres, Moscow [orig. in German 1785]