Never Mind the Trolley: The Ethics of Autonomous Vehicles in Mundane Situations

Johannes Himmelreich

Trolley cases are widely considered central to the ethics of autonomous vehicles. We caution against this by identifying four problems. (1) Trolley cases, given technical limitations, rest on assumptions that are in tension with one another. Furthermore, (2) trolley cases illuminate only a limited range of ethical issues insofar as they cohere with a certain design framework. Furthermore, (3) trolley cases seem to demand a moral answer when a political answer is called for. Finally, (4) trolley cases might be epistemically problematic in several ways. To put forward a positive proposal, we illustrate how ethical challenges arise from mundane driving situations. We argue that mundane situations are relevant because of the specificity they require and the scale they exhibit. We then illustrate some of the ethical challenges arising from optimizing for safety, balancing safety with other values such as mobility, and adjusting to incentives of legal frameworks.

1 Introduction

Imagine an autonomous vehicle is approaching a tunnel when a person suddenly appears in the car's path.[1] A collision is unavoidable and there are only two options. Either the car runs over the person, killing her, or the car swerves into the wall of the tunnel, killing the passenger. What should the car do? Such situations are known as *trolley cases*. Two trolley cases can be speciously similar and yet lead to conflicting intuitions about what to do. How should we make sense of these conflicting intuitions? This further question that arises from considering two or more trolley cases together is known as the *trolley problem* (Foot 1967; Thomson 1976, 1985). Although trolley cases originated from the trolley problem, in recent years, trolley cases have found a life of their own.[2]

Trolley cases are now widely taken to pose a central challenge in the ethics of autonomous vehicles. What looked like a purely hypothetical dilemma situation is

[1] We define an "autonomous vehicle" as a motorized ground vehicle with the capability of highly or fully automated driving, what is sometimes called automation level 4 and 5 (SAE International 2016).

[2] Many contributions do not distinguish as strictly as we do between trolley cases and trolley problems. However, we think this distinction is important. We are grateful to anonymous reviewers for their encouragement to make this distinction clear upfront.

about to become reality.[3] Before long, autonomous vehicles may have to decide about the distribution of harms. Accordingly, trolley cases have animated op-eds (Marcus 2012; Achenbach 2015; Shariff, Bonnefon, and Rahwan 2016), policy documents (Luetge 2017), and research papers (Bonnefon, Shariff, and Rahwan 2016; Nyholm and Smids 2016; Gogoll and Müller 2017; Fleetwood 2017; Millar 2017; Santoni de Sio 2017).

With this paper, we join a growing number of authors who caution against the view that trolley cases pose a central challenge in the ethics of autonomous vehicles (cf. Goodall 2016; Nyholm and Smids 2016). To offer a positive proposal, we argue that mundane situations, such as pedestrian crosswalks or left-turns at intersections, give rise to important ethical challenges.

We proceed in three steps. First, we present the case in favor of trolley cases. We identify four ways in which trolley cases can be useful in investigating the ethics of autonomous vehicles. Second, we make our negative case. We argue that the usefulness of trolley cases in investigations of the ethics of autonomous vehicles is limited. We discuss four objections to trolley cases. Some of these objections have been raised before (Goodall 2016; Nyholm and Smids 2016).[4] Other objections that we discuss have been put forth against the trolley problem more generally and have not yet been applied to the use of trolley cases in the ethics of autonomous vehicles (Elster 2011; Fried 2012; Wood 2013; Kagan 2015). Many objections that we discuss are, to our knowledge, novel.[5] Third, we return to a positive outlook. We argue for the ethical relevance of mundane traffic situations.

---

[3] We do not endorse this claim. For many authors this claim motivates trolley cases as relevant to the ethics of autonomous vehicles (see Nyholm and Smids (2016) for an overview).

[4] The objections that we discuss in this paper largely supplement objections discussed by Goodall (2016) and Nyholm and Smids (2016). Trolley cases, according to Goodall (2016), are problematic in that they (1) pose a false dilemma (in fact, there are more than two options), (2) assume certainty over outcomes, (3) assume certainty over the environment, and (4) are in fact rare. Nyholm and Smids (2016) argue that trolley cases are not perfectly analogous to the situations of autonomous vehicles. This is because (5) the decision problem is different (e.g. with respect to when decision is taken, and the numbers of agents involved), (6) the issues of moral and legal responsibility are in fact relevant but neglected by trolley cases, and because (7) decisions in fact need to be made under uncertainty. We take on board the points about uncertainty, that is, point (2), (3), and (7) in our fourth objection. We also agree with points (5) and (6) as raised by Nyholm and Smids (2016) but we do not pursue these points in our paper in this way (but see note 5). Our positive proposal on mundane situations incorporates the proposal made by Goodall (2016) on the importance of risk-management but it also extends this proposal in that we highlight considerations beyond risk and safety.

[5] Specifically, we are not aware of a full discussion elsewhere of our first objection (that given technical restrictions, trolley cases rest on assumptions that are in tension with one another) and our second objection (that trolley cases cohere with a certain design framework). Our third objection (that trolley cases look for a moral answer when a political answer is called for), can be seen as a version of point (5) made by Nyholm and Smids (2016). However, we instead focus on a specific instance of their point highlighting a difference between moral and political philosophy. Our fourth objection (that trolley cases might be epistemically problematic) combines objections made by many others (Elster 2011; Fried 2012; Wood 2013; Kagan 2015).

Our overall conclusion is comparative and cautious. We recognize that there is an important place for trolley cases. In light of the current state of the debate on autonomous vehicles, however, we caution against overstating the importance of trolley cases vis-à-vis mundane situations. Unlike for trolley cases, the ethical relevance of mundane situations is easily overlooked.

2 The Case for Trolley Cases

Trolley cases are idealized situations in which an agent has to decide between two actions that lead to different distributions of unavoidable harms. Specifically, for something to count as a trolley case, at least three conditions must be met. First, in trolley cases a collision is imminent and unavoidable.[6] Second, the agent is able to choose how to distribute the harms that ensue as a result of this collision. Third, the decision situation is one of certainty. Actions carry no risk so that the agent can choose between outcomes.[7]

Trolley cases are useful in at least four ways. They can be used (1) in trolley problems, (2) in experimental paradigms, and (3) as didactical tools. Furthermore, (4) trolley cases illustrate a social dilemma. We briefly discuss each of these points in turn.

First, trolley cases, of course, are used in trolley problems. In trolley problems, individuals ask themselves how they would decide in two or more trolley cases that are seemingly similar. Despite a specious structural similarity between cases, individuals often come to opposite intuitive judgments about what to do (Foot 1967; Thomson 1976). The *trolley problem* is the systematic search for a principled answer to the question of "[w]hat difference […] explains the moral difference between [two cases]?" (Thomson 1985, 1395). In other words, the trolley problem is a process of reflection on multiple trolley cases that are contrasted against one another, in order to bring to light deep distinctions and to inform moral principles. What is the moral difference between doing and allowing? Should you save the greater number or avoid treating people as a means? With this method of contrast, trolley cases have been used extraordinarily fruitfully and have led to the development of subtle and intricate normative theories (Kamm 2008, 2016).[8]

Second, trolley cases can be used in an experimental paradigm. In this way, trolley cases are used as a way of systematically eliciting intuitions about individual situations. Individuals are asked what they would do in one or more trolley cases. Just as with trolley problems, when used in an experimental paradigm, trolley cases are

[6] We restrict the discussion to collisions, given the context of autonomous vehicles. A more general definition would instead be formulated in terms of distributions of harms and benefits.
[7] Nyholm and Smids (2016) argue that each of these three assumptions is not met in the reality faced by autonomous vehicles and that trolley cases are therefore not a good analogy.
[8] For a helpful overview see Nyholm and Smids (2016, 1280).

thought experiments. As thought experiments, they can help us examine closely a small set of relevant considerations.[9] Many facts, such as age and identity of the people involved, and how the situation came about, are abstracted away. But the use of trolley cases in an experimental paradigm differs from the use of trolley cases in the trolley problem. For example, used in an experimental paradigm, trolley cases are generally considered in separation. An experimental paradigm aims to elicit intuitions and use them as data or evidence, not to reflect on moral differences between cases. The trolley problem, by contrast, is not primarily about intuitions in individual cases but instead about differences in intuitions between apparently similar cases.

When using trolley cases in an experimental paradigm, experimenters can systematically gather intuitions on a large number of perturbations of trolley cases to better understand how judgments about the relative importance of values may depend on contextual variations. Most notably, trolley cases have been employed in an experimental paradigm in moral psychology (e.g. Greene et al. 2001, 2009). Similarly, in discussions on the ethics of autonomous vehicles, trolley cases are used as separable decision situations that require individuals to make intuitive judgments about what to do. In this vein trolley cases are also used by the Moral Machine project.[10]

Third, trolley cases are an effective didactical device. They command a captivating fascination, they pose a tragic choice, and they instill a sense of moral urgency.[11] For this reason, it seems, many contributions to the broader discussion about autonomous vehicles open with a description of a trolley case. Any teacher who has given an introductory course in ethics can attest to the usefulness of trolley cases to engage students and foster a lively discussion. Trolley cases are also useful in engaging non-philosophers in a conversation about morality and to illustrate distinctions, such as that between doing and allowing. In short, at least in some circumstances, trolley cases succeed in motivating reflection on ethical issues.

Finally, in the context of autonomous vehicles, trolley cases give rise to an important social dilemma. Trolley cases reveal how individuals' ethical views and their strategic incentives conflict. A majority of individuals would be unwilling to use an autonomous vehicle that makes decisions in line with what they themselves would agree is ethically preferable. People want cars to be moral, except if they drive in them (Bonnefon, Shariff, and Rahwan 2016). This is a social dilemma in that it might hinder the technological transformation towards autonomous vehicles and achieving safety

[9] In the literature on autonomous vehicles, Lin (2014) argues that trolley problems are "meant to simplify the issues in order to isolate and study certain variables."
[10] See http://moralmachine.mit.edu. For another example see Frison, Wintersberger, and Riener (2016).
[11] Trolley cases in some ways resemble the party game of "would you rather" questions. Some of the reasons for which "would you rather" questions exert a certain attraction might also explain why trolley cases are captivating.

improvements that would come with it. With this social dilemma, trolley cases are useful because they illustrate an important issue for policymaking.[12]

3 Four Worries about Trolley Cases

These points on the usefulness of trolley cases notwithstanding, we identify four worries that caution against relying centrally on trolley cases to investigate the ethics of autonomous vehicles. First, trolley cases rest on assumptions that are in tension with one another, given technical restrictions. Second, trolley cases cohere with a certain design framework. Third, trolley cases tend to prompt for the wrong kind of answers. Trolley cases look for a moral answer when a political answer is called for. Finally, various reasons militate against the usefulness of trolley cases as a method for gathering intuitions.

Admittedly, one might object that trolley cases were never meant to do the things that we argue they fail to do. We acknowledge this point. However, it should be recalled that trolley cases were originally meant to be used in trolley problems, the methodology of which has been critically investigated. By contrast, in the context of autonomous vehicles, trolley cases have a different methodology. They seem to be used as a model to help investigate a relevant ethical challenge. We criticize the adequacy of this model. Our view, consonant with other authors, is that the model falls short in several ways (cf. Goodall 2016; Nyholm and Smids 2016). The upshot is not that the model must be abandoned. Rather, we take the upshot to be that in awareness of the limitations of this model, the model should be relied on with caution and only for a limited range of uses.

3.1 Consistency

The first worry we want to discuss is that basic assumptions of trolley cases might be in tension with one another. This claim needs some explanation. First of all, by "in tension" we mean that the assumption might be inconsistent. We write "might" because whether they are rest on (empirical) questions that we cannot answer here. Second, this claim about inconsistency rests on a semantic and not on a syntactic notion of consistency. We understand consistency as compossibility. A set of assumptions is compossible when there is a possibility, which we understand as a possible world, in which each of the assumptions is true. We argue that trolley cases might not be compossible, given certain restrictions on what is possible. However, we agree that trolley cases are not contradictory. That is, trolley cases are not committed both to an assumption and its negation.[13] We also concede that trolley cases are

[12] However, it stands to reason to what extent this situation – a paradigmatic instance of a collective action problem in which individual incentives lead to an outcome that is overall worse – is actually typical of politics.

[13] That trolley cases are not contradictory in this way is supported by the fact that they are clearly conceivable. Their conceivability suggests that trolley cases are epistemically possible.

compossible *simpliciter*. That is, absent restrictions on what is possible, there *are* epistemically possible worlds in which trolley cases occur. Our claim is, by contrast, that trolley cases might not be possible in a restricted sense of possibility, which we can call "technical possibility." We can think about technical possibilities as that subset of all epistemically possible worlds that satisfy restrictions concerning engineering design and traffic circumstances that we describe below. Within the set of those epistemically possible worlds that satisfy these restrictions, there are no worlds in which each of the different assumptions on which trolley cases rest are true.

Recall two of the assumptions necessary for a situation to qualify as a trolley case. The first assumption is that a collision must be imminent and unavoidable. The second assumption is that the agent in the situation must, nevertheless, have a choice over the distribution of unavoidable harms. In short, something is a trolley case only if a collision is unavoidable, but a meaningful choice is nevertheless possible. Call these the assumptions of *unavoidability* and *control*, respectively.

We worry that these two assumptions cannot be simultaneously satisfied given plausible constraints. First, under certain assumptions concerning the engineering design of autonomous vehicles and its failure modes, if one of these two assumptions is true, the other one seems to be false. We assume here that the failure modes in self-driving vehicles are correlated. When one part of the system fails, another is likely to fail too. There might be a meaningful choice, but then the collision might also be avoidable. Or a collision might be unavoidable, but then the agent does not have a choice about how to distribute harms because the vehicle is already out of control. There seems to be an inherent tension between these two basic assumptions about trolley cases.

Consider what would happen leading up to a trolley case. One option, in which a collision becomes unavoidable, is that the vehicle undergoes a total systems failure. In this case, a collision might be unavoidable, but the vehicle automation would at the same time not be able to make a meaningful choice. This illustrates that the two assumptions of control and unavoidability are in tension with each other if we assume a correlation of failure modes.

Of course, a trolley case might come about without a total systems failure. In fact, failure modes seem unlikely to be perfectly correlated. Yet, given plausible traffic conditions, the worry about trolley cases' inconsistency remains. Instead of assuming a correlation of failure modes, suppose instead that the vehicle is fully functional and that a pedestrian appears in the path of the vehicle unexpectedly. In this case, there are two options. Call them the low-speed (or long-distance) and the high-speed (or close-distance) scenario. In the low-speed scenario, there is enough time for a meaningful choice to be possible. Suppose also that a collision is unavoidable. In this way, both assumptions – control and unavoidability – are true. However, the low-speed scenario is not a usual trolley case in that the harms that accrue to different parties would not

be equal in kind. Instead, if hit, the pedestrian might die but, when driving into a wall, the car passenger might only be injured. Unlike what is usually expected of trolley cases, such scenarios vary several factors at the same time instead of holding almost all things constant. This may undermine the usefulness of this scenario as a trolley case.[14]

In a high-speed scenario, it is plausible that the harms are equal in kind. The pedestrian and the vehicle passenger would die as a result of the collision. This scenario would meet the assumption of a collision being unavoidable. But this scenario risks being inconsistent with the assumption of control. The speed needed to "ensure" that the passenger dies is likely so high that there is not sufficient means to avoid a collision. This, again, suggests that there is an inherent tension between the assumptions of a meaningful choice and the unavoidability of a crash.

In sum, by assuming a correlation of failure modes, or by thinking through plausible ways in which trolley cases come about, there seems to be a tension between having a meaningful choice and a collision being unavoidable. This undermines the relevance of trolley cases for the ethics of autonomous vehicles in a practical way. If our argument is correct, trolley cases hardly represent situations that might in fact occur. Control over a vehicle and unavoidability of collisions do not travel well together. Engineering constraints seem to leave little room for trolley cases. This worry, of course, leaves various theoretical benefits that are to be had by reflection on trolley cases unaffected. We provide this consideration about consistence as an improved version of an argument based on low frequency of trolley cases, which we do not think is plausible.[15]

We should be clear that we chose our terms in this conclusion – "might be inconsistent," "tension," and "threatens to undermine" – with deliberate caution. We realize that the question of whether trolley cases are in fact inconsistent in the way we describe depends on issues that we cannot settle here. Our limited aim is only to raise this potential inconsistency as a worry about the practical applicability of trolley cases and their usefulness as a model of ethical challenges in the context of autonomous vehicles.

[14] Because this scenario raises not only the question of to whom the harms accrue but also the question of whether harms should be minimized, it fails to isolate two values. An intuition is hence no clear indication about relative importance of two values.

[15] Some argue against trolley cases on the basis that they are rare (Goodall 2016). We do not pursue this objection. Even if the situations that give rise to trolley cases are rare, they will occur with certainty over the long run. Moreover, regardless of whether these situations in fact occur, autonomous vehicles still need to be programmed to behave in one way or another to prepare for the eventuality of unavoidable collisions. In short, the low frequency of trolley cases is, as such, not yet an argument against their relevance for the ethics of autonomous vehicles.

3.2 Limitations of Design

Another limitation on the range of issues on which trolley cases can shed light is that trolley cases lend themselves naturally to a specific design approach. This leads to two limitations. First, the ethical differences between different design approaches are not illuminated by trolley cases. Second, investigations of trolley cases might be discontinuous with actual engineering practice. Engineers might follow the opposite approach from that to which trolley cases most naturally relate.

Trolley cases assume what is known as a top-down approach to automated decision making or artificial intelligence. This approach is similar to a deliberative, conscious decision-making process. The development of artificial intelligence following this top-down approach aims at implementing principles to directly steer a given process. This top-down design approach contrasts with a bottom-up approach in which behavior is "learned," for example with neural networks, which resembles rather intuitive and unconscious decision-making. The reasons for which a choice is made in a bottom-up process are often inscrutable and hard to explain because the choice is not based on an explicit decision rule or principle.

Trolley cases naturally lend themselves to the top-down design approach. This top-down approach manifests in trolley cases in that trolley cases assume that an agent makes a decision explicitly, perhaps by way of drawing on a general principle (Wallach and Allen 2008; Nyholm and Smids 2016). This top-down approach has certain welcome features. For example, it allows, to some extent, to change the decision of what to do in a trolley case at any later point in time.

Of course, following a top-down approach gives rise to ethical questions of its own. For example: Given that the decision of how to behave in trolley cases can be changed at any point in time, should passengers have a say in these decisions? Issues like these cannot be illuminated by answers of what to do in trolley cases. Yet, for most challenges that autonomous vehicles will face on the engineering front, design decisions need to be made. Those decisions are likely to have serious repercussions for how autonomous vehicles perform and how society will accept this technology. Some go so far as to suggest that the "morally most important decision … is made at the planning stage when it is decided how the autonomous vehicles are going to be programmed" (Nyholm and Smids 2016, 1280).[16] By their inability to illuminate questions of design, trolley cases are importantly limited. This cautions against focusing narrowly on trolley cases in investigations of ethical questions of autonomous vehicles. Furthermore, given the current prominence of the bottom-up approach in artificial intelligence in the form of neural networks, there is a risk of a discontinuity

[16] It should be noted that Nyholm and Smids (2016) discuss decision-making situations – such as the number of agents involved, and the information available. Nyholm and Smids do not discuss different design approaches in artificial intelligence.

of approaches between ethics and engineering. Engineers might follow the approach opposite to the approach to which trolley cases most naturally relate.[17]

3.3 Moral and Political Problems

Furthermore, we worry that trolley cases demand the wrong kind of solution. We take it that a solution to a trolley case consists in, first, a choice of actions, and second, a justification of this choice, which often takes the form of an abstract normative theory (Rakowski 2016; Bonnefon, Shariff, and Rahwan 2016).[18] In short, trolley cases are taken to be an issue of morality. But we think that this locates the problem on the wrong level. Instead, solutions are called for on the level of politics. Whereas moral philosophy is a reflection on individual conduct, political philosophy is a reflection on social arrangements before the backdrop of substantive disagreement.[19]

It seems unlikely that solutions to trolley cases would find broad societal acceptance. What counts as the right choice in such dilemma situations is essentially contested.[20] The disagreement runs deep.

How should we deal with widespread and deep disagreement about issues of morality?[21] This question raises a political problem. Political philosophers reflect on different models of governing our common life in the face of moral disagreement and value pluralism. A political approach takes as its starting point the diversity of views and values that are the predicament of any political community (e.g. Rawls 1993). This political approach contrasts with the approach of moral philosophy that is taken by trolley cases. Trolley cases make no room for such pluralism by aiming to elicit an individual's decision. A trolley case prompts us to make an individual choice when what we in fact face is a social choice. What seems needed is a kind of compromise to overcome disagreements over issues of value. Insofar as we value the moral diversity of our political community, it should be recognized that autonomous vehicles pose primarily a political problem, not a moral one.[22]

---

[17] Despite these limitations, trolley cases here play their role as a didactical device.

[18] In this way the methodology of trolley cases differs significantly from that of trolley problems which aims at the formulation of moral principles. We thank an anonymous referee for pressing us to make this clear.

[19] Nyholm and Smids (2016, 1282) make a similar point in that they identify as a disanalogy between autonomous vehicles and the trolley problem the fact that the former is a decision-situation involving "multiple stakeholders" whereas the in latter "the morally relevant decision-making is done by a single agent." However, their objection is much more general. They do not highlight this distinction between moral and political philosophical approaches.

[20] Judith Jarvis Thomson reminds us that "we should be troubled by the fact that so many people have tried, for so many years—well over a quarter of a century by now—and come up wanting." (2008)

[21] This question is the starting point of Gogoll and Müller (2017) for their discussion of whether ethics settings should be mandatory or personal.

[22] Nevertheless, trolley cases can play a useful role as a didactical device by illustrating the issue of the ethics of user settings (Millar 2014; Gogoll and Müller 2017; Millar 2017).

3.4 Uncertainty and Evidential Value

For the sake of argument, suppose we were to find a solution to trolley cases; and suppose further that almost everybody agreed on this solution. Nevertheless, in determining the ethics settings of autonomous vehicles, this solution would still only be of limited help. This is because there are various epistemic problems with trolley cases and trolley problems more generally (Fried 2012; Wood 2013; Kagan 2015; Nyholm and Smids 2016).

First of all, the solution to one trolley case might not carry over to other, novel situations. What seems like the right choice in one situation might turn out to be the wrong choice in the next. A trolley case often stipulates that all individuals involved are identical with respect to personal characteristics. But small changes might matter. Depending on the age of those involved, their relation to us, their responsibilities in a given traffic situation, we might want to choose differently (cf. Wood 2013). More generally, whether or not a given moral choice or principle extends from one situation to a different situation is itself a moral question.

Moreover, a solution to a trolley case might not extend to decisions under uncertainty (Fried 2012; Goodall 2016; Nyholm and Smids 2016). All decisions that we make – and any decision that an autonomous vehicle would have to make – are appropriately represented only in terms of probability. In contrast, trolley cases assume certainty in two ways. First, they assume that the agent facing the decision is certain that a collision is unavoidable. But recognizing whether or not a collision is unavoidable is not trivial but is instead a matter of probability (Fraichard and Asama 2004). Hence, there is a problem of *situation-uncertainty*. Second, trolley cases stipulate as an idealization that the outcomes of our choices are certain. But in fact, what outcomes result from a choice is also a matter of probability. Hence trolley cases do not represent *decision-uncertainty*. For either of these kinds of uncertainty, the following problem arises. It is not clear whether and how principles for decisions under certainty extend to decisions under uncertainty (cf. Jackson and Smith 2006; Hansson 2013, chap. 2). "Decisions that are easy to make under certainty can become much more difficult and morally fraught under uncertainty" (Bjorndahl, London, and Zollman 2017). Uncertainty raises deep problems for an entire range of normative theories. This insight has led to a sustained debate in the recent literature (Jackson and Smith 2016; Bjorndahl, London, and Zollman 2017; Lazar 2018; Lazar and Lee-Stronach 2017; Tenenbaum 2017). However, in addressing these problems of uncertainty, reflection on trolley cases is only of limited help. Fried (2012, 506) goes as far as to conclude: "Of the various moral principles that have emerged from the now four-decades-long preoccupation with trolley problems, none can handle the problem of garden-variety risk."

Finally, the evidential value of intuitive judgments made in the context of highly abstract decision situations is questionable (Elster 2011; Fried 2012; Kagan 2015;

Etzioni and Etzioni 2017). Our competence in making moral judgments in imaginary cases is diminished insofar as these cases are highly idealized and abstract and thereby very different from the environment we are familiar with. A similar point is raised by Kagan's (2015) worry that "our intuitions about trolley problems respond to factors that simply do not have any genuine moral significance." This line of argument casts doubts on the role of trolley cases as a way of acquiring evidence to inform ethical decisions in the context of autonomous vehicles.

In summary, we think that investigating trolley cases is not central to addressing ethical issues surrounding autonomous vehicles. We have identified four worries. First, trolley cases might be inconsistent, which might lead to a practical limitation on their usefulness. Second, trolley cases cohere best with a top-down design approach. They thereby leave important design decisions out of view and they might be discontinuous with currently prominent approaches in artificial intelligence. Third, trolley cases look for a moral solution when a political solution is called for. Fourth, and finally, even if we found a solution that is acceptable to all, such a solution would only be of limited help for the following reasons. (1) Whether a solution for one kind of trolley case will carry over to another situation is unclear. (2) Specifically, investigations based on trolley cases might not be able to inform issues concerning situation-uncertainty and decision-uncertainty. (3) Given that our intuitions are a good guide only in familiar environments, the idealizations of trolley cases might undermine the evidential value of intuitions. Each of these points illustrates the limitations of trolley cases for investigations concerning the ethics of autonomous vehicles.

4 The Ethical Challenge of Mundane Situations

We offer a positive vision for the ethics of autonomous vehicles. We argue that mundane traffic situations, such as approaching a crosswalk with limited visibility, making a left turn with oncoming traffic, and navigating through busy intersections, raise important ethical questions for autonomous vehicles. The range of ethical issues relating to such mundane situations is inclusive and encompasses techniques of risk management (Goodall 2016), issues of social justice (Mladenovic and McPherson 2016), as well as challenges arising on the level of the traffic system as a whole (Borenstein, Herkert, and Miller 2017). We here focus on questions of driving strategy or yielding behavior because in this respect mundane situations might not look as if they would give rise to ethical issues at all. That mundane situations pose ethical questions for driving behavior might seem surprising. After all, what can be so hard about, for example, approaching a crosswalk? When unsure as to whether a pedestrian is about to cross, one might argue, you just have to slow down.

For autonomous vehicles, driving strategies in mundane situations are challenging for two reasons. First, we can make decisions intuitively, whereas machines cannot. This is the *challenge of specificity* and it is an instance of what is known as Moravec's

paradox: What is easy for us is hard for automated systems. We decide intuitively how carefully we need to proceed. Yet this intuitive feel or know-how does not easily translate into an algorithm. We understand the meaning of the imperative "slow down" but its meaning is hard to make precise because it depends on various contextual and environmental parameters. Is there another car close behind that might not expect us to slow down? How likely will there be pedestrians on the street in this neighborhood at this time of the day? The difficulty of spelling out driving strategies explicitly and precisely suggests that implementing vehicle behavior in such mundane situations in a top-down approach is unlikely to succeed.

The challenge of specificity cannot easily be avoided by taking a bottom-up approach that inductively replicates actual behavior by human drivers to inform vehicle behavior strategies. Data on actual behavior is biased in problematic ways. Actual yielding culture is subject to notable geographic differences (Schneider and Sanders 2015) and tends to be discriminatory with respect to age (Rosenbloom, Nemrodov, and Eliyahu 2006), race (Goddard, Kahn, and Adkins 2015), and income (Coughenour et al. 2017). Furthermore, since pedestrians' behavior towards autonomous vehicles is likely to differ from their behavior towards human-driven cars, replicating actual behavior inadequately addresses the problem that autonomous vehicles could be strategically exploited by pedestrians (Millard-Ball 2016).

Next to the challenge of specificity, there is the *challenge of scale*. For human drivers, it is not worth dwelling on the question of how to drive best in mundane traffic situations. This is not only because mundane situations are easy to handle intuitively, but also because how each of us drives – as long as we drive somewhat reasonably – does not make a significant difference overall. With autonomous vehicles, by contrast, driving behavior in mundane situations becomes a general policy. The decision of how an autonomous vehicle approaches a crosswalk affects not the behavior of only one car but the behavior of all vehicles programmed this way at all crosswalks. Instead of a small-scale problem, mundane situations now pose a large-scale problem. Small differences about driving behavior will make a big difference in the aggregate. Because mundane situations occur so often, the resulting statistical injuries and deaths are likely to be sizeable.

The challenges of specificity and scale make mundane situations ethically relevant. Human drivers can be diverse in style and intuitive in their decisions. But autonomous vehicles will be uniform in style and have to be specific in their approach. Before this backdrop, we identify three ways in which mundane situations raise ethically relevant questions. First, issues arise from the optimization problem of making autonomous vehicles as safe as possible. Second, there is a trade-off between safety and other values, such as mobility, environmental protection, and affordability. Third, problems arise as to how the existing legal framework can accommodate legal questions arising from autonomous vehicles. Specifically, we illustrate this point with

the argument that the existing legal framework produces objectionable incentives in mundane situations.

4.1 Optimizing for Safety

Autonomous vehicles can potentially make driving significantly safer than it is today. Yet, the extent to which potential safety improvements are realized depends on engineering and policy decisions. Take the mundane situation of approaching a crosswalk with limited visibility. Some manufacturers might decide to follow a broadly top-down approach, other manufacturers might rely on a bottom-up approach and let the car learn from human drivers and supervisors. It is likely that some approaches will be safer than others.

This creates two challenges. First, if coordination over different technical approaches is needed to improve overall safety, how is such a technological coordination facilitated in a competitive environment? Similar questions arise concerning interoperability standards concerning, for example, protocols for vehicle-to-vehicle communication, or the operation of a centralized management of intersections for vehicle-to-infrastructure communication (Borenstein, Herkert, and Miller 2017). Second, with their respective solutions different manufacturers each might find local safety optima. How can the solutions be combined to escape local optima to reach a feasible global optimum? These are substantive ethical questions leading to conflicts with intellectual property rights, for example (Crane, Logue, and Pilz 2017). To illustrate the ethical relevance, consider what available answers to these questions might look like.

Different measures are conceivable to encourage breaking out of local safety optima and facilitate coordination on technical approaches. One measure is based on regulation. Only the safest autonomous vehicles, or those meeting certain minimal standards, would be allowed on the road. In a regulatory approach, economic freedoms are given up in order to prevent accidents. Another measure would be to require a knowledge-transfer, or data-sharing between manufacturers. The overall safety performance of autonomous vehicles can be improved by identifying successful solutions publicly such that competing manufacturers can adopt strategies from each other. However, the privacy of user data might be at risk (Borenstein, Herkert, and Miller 2017). Furthermore, intellectual property rights and the potential for product differentiation are given up in order to promote safety. A third measure would be to leave questions of safety optimization in the hands of consumers. However, since consumers will choose between products based on various reasons distinct from safety, this last approach is likely to result in a situation that would be overall less safe compared to measures based on exchange or regulation. Each of these three measures would help to optimize the safety of autonomous vehicles, but which measure, or mix thereof, is preferable is an ethical question.

In short, given the challenges of specificity and scale, engineering how autonomous vehicles behave in mundane situations is a complex challenge and, given that human health and lives are at stake, the choice of policy approach is ethically relevant. Although similar governance questions – improving safety through regulation, exchange, or though the market – have been raised for other products, the situation of autonomous vehicles poses unique problems given the challenges of specificity and scale.

4.2 Balancing Different Values

Apart from considerations based on safety, other values might be relevant in designing driving and yielding behavior in mundane situations. We consider three further values that will have to be traded off against safety: mobility, environmental impact, and urban design.[23]

By "mobility" we understand a measure of traffic efficiency, such as the average speed of traffic flow. If a vehicle decelerates strongly in approaching a crosswalk with limited visibility, this will increase safety at the expense of reduced traffic flow. Individuals make these trade-offs between efficient mobility and safety intuitively. But with autonomous vehicles, these trade-offs can be made on a systemic level, which gives rise to the question of what the right balance between safety and mobility is. Instead of opting for a general balance that covers everyone, another option would be to allow personal settings, that is, that individuals can adjust the driving behavior of their autonomous vehicles, perhaps for a monetary payment (Millar 2017). When you are late to a meeting, you can pay your way through traffic or override the safety features. The extent to which this should be allowed, if it should be allowed at all, raises an ethical issue.

Furthermore, mundane situations might give rise to trade-offs with values external to the traffic system such as environmental protection. Given the challenge of scale, a vehicle's performance settings in mundane situations as simple as cornering will have significant environmental impact, concerning greenhouse gas emissions or traffic noise. Depending on how fast a vehicle accelerates and breaks, the environmental impact due to emissions and material wear will differ (Millar 2017). Because a great number of vehicles will follow the same algorithm of how to handle mundane situations, incremental changes will have significant effects on reducing or increasing environmental impact in the aggregate.

Finally, situations as mundane as pedestrian crossings illustrate that the introduction of autonomous vehicles opens the opportunity for transformative and novel approaches to urban design. Given that autonomous vehicles might be much more reliable in yielding to pedestrians safely and efficiently, the question arises of

[23] Other examples of relevant values are values of social justice, such as sustainability, privacy, and equality of access (Mladenovic and McPherson 2016).

whether pedestrians should be granted greater priority when crossing streets anywhere.[24] Perhaps the idea of dedicated crosswalk areas should be abandoned. This is a crucial question affecting urban design and it is a normative question. This crucial question raises subsequent questions. Suppose pedestrians were given priority over vehicles in crossing streets anywhere in a city – should pedestrians then be required to indicate that they want to cross? This question involves ethical issues concerning the roles and responsibilities of pedestrians and users of autonomous vehicles. Reflections on how autonomous vehicles will change quotidian and mundane traffic situations can help motivate these questions.

In short, mundane situations embody trade-offs between different values such as safety, mobility, efficiency, environmental impact, and how pedestrians' responsibilities should be taken into account in urban design. Which of these values are important, how important they are vis-à-vis one another, and how the trade-offs between them should be made is – given the challenge of scale – a significant ethical issue.

4.3 Adjusting Legal Incentives

Compared to the *status quo*, manufacturers of autonomous vehicles are likely to face an increase in liability exposure and lawsuits based on novel failure modes (Marchant and Lindor 2012; Crane, Logue, and Pilz 2017). Given that a large proportion of traffic accidents occur in mundane driving situations, we should examine mundane situations with respect to the legal frameworks that govern liability in these situations. Given restrictions of space, we cover the complex legal landscape in very general terms and only mention one ethical issue to illustrate our case.

It is a basic principle in US tort law that liability damages are a function of the income lost to dependents (Posner and Sunstein 2005). The more you earn, the greater your liability claim. Assuming that manufacturers will want to keep the exposure to liability claims constant, they are incentivized to adjust driving behavior depending on average income in an area. In an affluent area, an autonomous vehicle would drive more carefully than in an economically a deprived area. In short, the existing legal framework incentivizes discriminatory driving behavior (Casey 2017).

The challenge emerges from the fact that a basic principle of tort law is not easily changed. How exactly legal frameworks should be adjusted – should the principle of tort law be suspended only for accidents involving autonomous vehicles? – is a relevant ethical question at the intersection of applied ethics and law. This case hence illustrates that even in mundane situations, ethical questions arise, for example, when legal frameworks pose objectionable incentives.

[24] In most legislations, pedestrians' responsibilities are higher when crossing the street outside of dedicated crossings. Unlike in crosswalks, drivers might not have to yield to pedestrians.

5 Conclusion

We have identified several ethical challenges of autonomous vehicles that arise from mundane situations. In comparison with trolley cases, these challenges might seem less obvious and pressing. To emphasize the relative importance of mundane situations, we have discussed four worries about taking trolley cases to be central to the ethics of autonomous vehicles. First, trolley cases might rest on assumptions that are in tension with one another, given technical limitations. Second, trolley cases cohere with a top-down design approach. Reflection on mundane traffic situations, by contrast, invites us to consider the relevance of such engineering and design decisions. Third, trolley cases address the ethics of autonomous vehicles on the wrong level. They seek to elicit an individual choice (a moral solution) when, in fact, a social choice (a political solution) is called for. Mundane traffic situations illustrate how the driving behavior of autonomous vehicles meshes with the rights and responsibilities of other traffic participants and moral values held in society at large. Fourth, solutions to trolley cases, to the extent they are widely acceptable, are likely to be only of limited help in informing decisions in novel and uncertain situations. Reflection on mundane situations, in contrast, can inform the development of ethical vehicle behavior immediately.

Mundane situations, in sum, give rise to important ethical issues for autonomous vehicles and they do so because of the two fundamental challenges of specificity and scale. Whereas human drivers decide intuitively and on a small-scale, with autonomous vehicles, behavior in mundane situations becomes a matter of policy. Small differences in an algorithm pertinent to a mundane situation might have a significant effect in the aggregate. This gives rise to three kinds of ethical issues. First, the optimization problem to make autonomous vehicles as safe as possible puts at stake issues of economic freedom and intellectual property rights. This is an internal value conflict that arises in the process of achieving global safety optima. Second, further values – such as mobility, environmental impact, or values in urban design and traffic planning – might conflict with safety. How these concurring values are balanced against each other is an important ethical question. Third, existing legal frameworks give rise to perverse incentives. Adjusting the framework and mitigating against these incentives is a delicate issue because the effects of legal changes are potentially widespread.

In conclusion, while we concede that trolley cases may be useful as thought experiments in trolley problems, as a method to gather evidence, as a didactical device, and to illustrate a social dilemma, we have argued that, when it comes to the ethics of autonomous vehicles, their usefulness is limited. Mundane situations deserve more attention in reflections on the moral and political issues involved in the development of autonomous vehicles.

References

Achenbach, J (2015) Driverless Cars Are Colliding with the Creepy Trolley Problem. Washington Post. December 29, 2015. https://www.washingtonpost.com/news/innovations/wp/2015/12/29/will-self-driving-cars-ever-solve-the-famous-and-creepy-trolley-problem/. Accessed 31 October 2017

Bjorndahl A, London AJ, Zollman KJS (2017) Kantian Decision Making Under Uncertainty: Dignity, Price, and Consistency. Phil Impr 17

Bonnefon JF, Shariff A, Rahwan I (2016) The Social Dilemma of Autonomous Vehicles. Science 352:1573-76

Borenstein J, Herkert JR, Miller KW (2017) Self-Driving Cars and Engineering Ethics: The Need for a System Level Analysis. Sci Eng Ethics. https://doi.org/10.1007/s11948-017-0006-0

Casey BJ (2017) Amoral Machines, or: How Roboticists Can Learn to Stop Worrying and Love the Law. Northwest U Law Rev 11:231-50

Coughenour C, Clark S, Singh A, Claw E, Abelar J, Huebner J (2017) Examining Racial Bias as a Potential Factor in Pedestrian Crashes. Accid Anal & Prev 98:96–100. https://doi.org/10.1016/j.aap.2016.09.031

Crane D, Logue K, Pilz B (2017) A Survey of Legal Issues Arising from the Deployment of Autonomous and Connected Vehicles. Mich Telecommun Technol Law Rev 23:191-320

Elster J (2011) How Outlandish Can Imaginary Cases Be? J Appl Phil 28:241-258. https://doi.org/10.1111/j.1468-5930.2011.00531.x

Etzioni A, Etzioni O (2017) Incorporating Ethics into Artificial Intelligence. J Ethics 21:403-18. https://doi.org/10.1007/s10892-017-9252-2

Fleetwood J (2017) Public Health, Ethics, and Autonomous Vehicles. Am J Public Health 107:532-37. https://doi.org/10.2105/AJPH.2016.303628

Foot P (1967) The Problem of Abortion and the Doctrine of Double Effect. Oxford Rev 5:5-15

Fraichard T, Asama H (2004) Inevitable Collision States – a Step towards Safer Robots?. Adv Robotics 18:1001-24. https://doi.org/10.1163/1568553042674662

Fried BH (2012) What Does Matter? The Case for Killing the Trolley Problem (Or Letting It Die). Phil Q 62:505-29. https://doi.org/10.1111/j.1467-9213.2012.00061.x

Frison AK, Wintersberger P, Riener A (2016) First Person Trolley Problem: Evaluation of Drivers' Ethical Decisions in a Driving Simulator. Adjun Proc 8th Intern Conf Automoti User Interfaces Interact Vehicular Appl, 117–122. https://doi.org/10.1145/3004323.3004336

Goddard T, Kahn KB, Adkins A (2015) Racial Bias in Driver Yielding Behavior at Crosswalks. Transp Res Part F: Traffic Psychol Behav 33:1-6. https://doi.org/10.1016/j.trf.2015.06.002

Gogoll J, Müller JF (2017) Autonomous Cars: In Favor of a Mandatory Ethics Setting. Sci Eng Ethics 23:681-700. https://doi.org/10.1007/s11948-016-9806-x
Goodall NJ (2016) Away from Trolley Problems and Toward Risk Management. Appl Artif Intell 30(8):810-21. https://doi.org/10.1080/08839514.2016.1229922
Greene JD, Cushman FA, Stewart LE, Lowenberg K, Nystrom LE, Cohen JD (2009) Pushing Moral Buttons: The Interaction between Personal Force and Intention in Moral Judgment. Cogn 111:364-71. https://doi.org/10.1016/j.cognition.2009.02.001
Greene JD, Sommerville RB, Nystrom LE, Darley JM, Cohen JD (2001) An FMRI Investigation of Emotional Engagement in Moral Judgment. Science 293:2105-8. https://doi.org/10.1126/science.1062872
Hansson SO (2013) The Ethics of Risk: Ethical Analysis in an Uncertain World. Palgrave Macmillan, New York.
Jackson F, Smith M (2006) Absolutist Moral Theories and Uncertainty. J Phil 103:267-83. https://doi.org/10.2307/20619943
Jackson F, Smith M (2016) The Implementation Problem for Deontology. In Lord E, Maguire B (eds) Weighing Reasons. Oxford: Oxford University Press, pp, 279–92. https://doi.org/10.1093/acprof:oso/9780199315192.003.0014
Kagan S (2015) Solving the Trolley Problem. In Rakowski E (ed) The Trolley Problem Mysteries. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190247157.001.0001.
Kamm FM (2008) Intricate Ethics: Rights, Responsibilities, and Permissible Harm. Oxford: Oxford University Press.
Kamm FM (2016) The Trolley Problem Mysteries. In Rakowski E (ed) The Trolley Problem Mysteries. Oxford: Oxford University Press.
Lazar S (2018) In Dubious Battle: Uncertainty and the Ethics of Killing. Phil Stud 175:859-883. https://doi.org/10.1007/s11098-017-0896-3
Lazar S, Lee-Stronach C (2017) Axiological Absolutism and Risk. Noûs. https://doi.org/10.1111/nous.12210
Lin P (2014) The Robot Car of Tomorrow May Just Be Programmed to Hit You. WIRED. https://www.wired.com/2014/05/the-robot-car-of-tomorrow-might-just-be-programmed-to-hit-you/. Accessed 31 October 2017
Luetge C (2017) The German Ethics Code for Automated and Connected Driving. Philos & Technol. https://doi.org/10.1007/s13347-017-0284-0
Marchant GE, Lindor RA (2012) The Coming Collision between Autonomous Vehicles and the Liability System Symposium Article. Santa Clara Law Rev 52:1321-40
Marcus G (2012) Moral Machines. The New Yorker. https://www.newyorker.com/news/news-desk/moral-machines. Accessed 26 October 2017
Millar J (2014) An Ethical Dilemma: When Robot Cars Must Kill, Who Should Pick the Victim? Robohub. http://robohub.org/an-ethical-dilemma-when-robot-cars-must-kill-who-should-pick-the-victim/. Accessed 31 October 2017

Millar J (2017) Ethics Settings for Autonomous Vehicles. In Lin P, Jenkins R, Abney K (eds) Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence. New York: Oxford University Press, pp. 20–34
Millard-Ball A (2016) Pedestrians, Autonomous Vehicles, and Cities. J Plan Educ Res 38:6-12. https://doi.org/10.1177/0739456X16675674
Mladenovic MN, McPherson T (2016) Engineering Social Justice into Traffic Control for Self-Driving Vehicles?. Sci Eng Ethics 22:1131-49. https://doi.org/10.1007/s11948-015-9690-9
Nyholm S, Smids J (2016) The Ethics of Accident-Algorithms for Self-Driving Cars: An Applied Trolley Problem?. Ethical Theory and Moral Pract 19:1275-89. https://doi.org/10.1007/s10677-016-9745-2
Posner EA, Sunstein CR (2005) Dollars and Death. Univ of Chic Law Rev 72:537-98
Rakowski E (2016) Introduction. In Rakowski E (ed) The Trolley Problem Mysteries. Oxford: Oxford University Press
Rawls J (1993) Political Liberalism. New York: Columbia University Press.
Rosenbloom T, Nemrodov D, Eliyahu AB (2006) Yielding Behavior of Israeli Drivers: Interaction of Age and Sex. Percept Mot Skills 103:387-90. https://doi.org/10.2466/pms.103.2.387-390
SAE International (2016) Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. http://standards.sae.org/j3016_201609/. Accessed 29 November 2017
Santoni de Sio F (2017) Killing by Autonomous Vehicles and the Legal Doctrine of Necessity. Ethical Theory and Moral Pract 20:411-29. https://doi.org/10.1007/s10677-017-9780-7
Schneider RJ, Sanders RL (2015) Pedestrian Safety Practitioners' Perspectives of Driver Yielding Behavior Across North America. Transp Res Re: J Transp Res Board 2519:39-50. https://doi.org/10.3141/2519-05
Shariff A, Bonnefon JF, Rahwan I (2016) Whose Life Should Your Car Save? The New York Times. https://www.nytimes.com/2016/11/06/opinion/sunday/whose-life-should-your-car-save.html. Accessed 27 October 2017
Tenenbaum S (2017) Action, Deontology, and Risk: Against the Multiplicative Model. Ethics 127:674-707. https://doi.org/10.1086/690072
Thomson JJ (1976) Killing, Letting Die, and the Trolley Problem. The Monist 59: 204-17. https://doi.org/10.5840/monist197659224
Thomson JJ (1985) The Trolley Problem. The Yale Law J 94:1395. https://doi.org/10.2307/796133
Thomson JJ (2008) Turning the Trolley. Phil Public Aff 36:359-74. https://doi.org/10.1111/j.1088-4963.2008.00144.x
Wallach W, Allen C (2008) Moral Machines: Teaching Robots Right from Wrong. Oxford: Oxford University Press
Wood A (2013) Humanity as End in Itself. In Scheffler S (ed) On What Matters vol 2. Oxford: Oxford University Press, pp. 58-82