

Autonoblog

Autonomes Fahren zwischen Hype und Wirklichkeit

Ethik & autonomes Fahren III: Das Problem mit dem Trolley-Problem

[Ethik & autonomes Fahren, Wiki / Von David Knollmann](#)

Wenn von der Ethik des autonomen Fahrens die Rede ist, spricht man unvermeidlich [über das Trolley-Problem](#). Jedenfalls in der populärwissenschaftlich-journalistischen Betrachtungsweise erfreut sich das Phänomen großer Beliebtheit. Und das muss nicht überraschen, ist das Trolley-Problem zunächst doch ein vorteilhaftes Vehikel, um ethische Fragen anschaulich zu diskutieren und unterschiedliche ethische Theorien zu „testen“. Erst in einem zweiten Schritt [zeigt sich dann](#): Bei genauerer Betrachtung erwachsen aus der anfänglichen Anschaulichkeit schnell ethische Komplikationen, deren Auflösung mitunter Kopfzerbrechen bereitet.

Aber nicht nur die Kompliziertheit im zweiten Schritt mag ein Nachteil der Diskussion ethischer Probleme des autonomen Fahrens anhand des Trolley-Problems sein. Es gibt [zahlreiche Experten](#), die viel weiter gehen und kritisieren: Die Dilemma-Betrachtung führt uns am Kern des Problems vorbei. Oder anders:

Womöglich hat das Trolley-Problem mit der *tatsächlichen* Ethik des autonomen Fahrens und den entsprechenden Herausforderungen *gar nicht soviel zu tun*.

Wo sich die „Ethik des autonomen Fahrens“ vom Trolley-Problem unterscheidet

Denn fraglich ist vielmehr, ob sich die Szenarien, die Trolley-Probleme entwerfen, von den Szenarien unterscheiden, denen sich ein autonomes Fahrzeug *in der Realität* ausgesetzt sieht.

Wenn wir nach der Ethik des autonomen Fahrens fragen, darf die Suche nach Erkenntnis und ethischen Gewissheiten streng genommen nicht bei philosophischen Gedankenexperimenten enden, sondern man muss sich ebenso damit auseinandersetzen, wie autonome Fahrzeug nicht nur für einen Fall, sondern für *alle denkbaren Fälle* programmiert werden sollten.

Die Unterschiede fangen schon dort an, wo auf die Ergebnisse – neudeutsch: Outcomes – möglicher Kollisionsszenarien geblickt wird. Beim Trolley-Problem stehen diese Outcomes *von vornherein fest* und sind

bekannt. Man *weiß*, dass Person X sterben wird, wenn sie vom autonomen Fahrzeug erfasst wird. Es besteht dahingehend *keine* Unsicherheit, die Trolley-Probleme sind deshalb auch *deterministisch*.

Damit unterscheiden sich Trolley-Probleme aber fundamental von der Programmierung autonomer Fahrzeuge. Diese muss nämlich mit *Unsicherheit* umgehen, die eine zentrale Eigenschaft aller realen Unfallszenarien ist: Behält die Radfahrerin ihren Kollisionskurs bei oder bremst sie im letzten Moment doch noch ab? Tritt vielleicht ein Fußgänger unvermittelt auf die Straße? Wie tief ist die Pfütze auf dem Asphalt und wie verändert sich der Bremsweg bei nassen Reifen?

Die Modelle, die die Rechner im autonomen Fahrzeug verwenden, sind deshalb *probabilistisch*: Sie rechnen mit Wahrscheinlichkeiten. Die Maschine *schätzt* Bewegungsbahnen, sie *schätzt* die Auswirkungen von Straßenkonditionen auf diese Bewegungsbahnen, sie *schätzt* mögliche Ergebnisse einer Kollision – kurzum: sie ist kein omnipotenter Gott, sondern lediglich ein guter Rechner, der mit maximaler Komplexität und begrenzten kognitiven Ressourcen (Sensoren, Rechenpower) umgehen muss und deshalb nicht alles *wissen kann*.

Im Gegensatz zur ethischen Beurteilung innerhalb von Trolley-Problemen, muss sich im Zusammenhang mit autonomen Fahrzeugen deshalb einer *Risikoethik* zugewandt werden, einer moralischen Betrachtung von Risiko-Management und Entscheidungen unter Unsicherheit. Nyholm und Smids verdeutlichen dies [in Anlehnung an das ursprüngliche Trolley Problem von fünf vs. einem Opfer](#) wie folgt:

„A significant number of people may find hitting the pedestrian morally unacceptable if this was certain to kill him. [...] But what if the estimated chance of a fatal collision were 10%? Or just 1%? To many people, imposing a 1% chance of death on an innocent pedestrian in order to save five car-passengers might appear to be the morally right choice. The trolley cases don't require any such judgments. In the scenarios involved in the trolley cases, all outcomes are assumed to be a 100% certain, and hence there is no need to reflect on how to weigh different uncertain and/or risky outcomes against each other.“¹

Nyholm und Smids nennen in ihrer Kritik an der „Trolley-Sichtweise“ noch fünf weitere Kategorien, die die Unterschiede verdeutlichen, in der folgenden Tabelle veranschaulicht werden und danach kurz zusammengefasst werden sollen:

**Programmierung von
Entscheidungsalgorithmen für
autonome Fahrzeuge** **Trolley-Problem**

<i>Entscheidungssubjekt</i>	Gruppe; zahlreiche stakeholder	Individuum
<i>Zeitdimension</i>	Zukünftige Entscheidungssituation; Kontingenzplanung	Unmittelbarer Handlungsdruck, „hier und jetzt“
<i>Zu berücksichtigende Entscheidungsparameter, Umweltfaktoren</i>	Potentiell unbegrenzt; frei	Wenige Parameter, Kontext ausgeklammert
<i>Juristische und moralische Verantwortung</i>	Zu berücksichtigen	Nicht berücksichtigt
<i>Wissen</i>	Risikoabschätzung; Entscheidung unter Unsicherheit	Perfektes Wissen; Outcomes sicher und bekannt
<i>Soziale Interaktion</i>	Zu berücksichtigen	Nur einseitig, keine Wiederholung

1. *Entscheidungssubjekt*: Beim Trolley-Problem lastet die Entscheidungsverantwortung auf dem Individuum. Beim autonomen Fahrzeug ist es völlig anders: Es wird von mehreren Personen programmiert, wobei nicht nur die jeweiligen Experten, Informatiker, sondern auch weitere Personen im Unternehmen sowie mittelbar auch Politiker, zivilgesellschaftliche Akteure und weitere Stakeholder und Interessenvertreter die Möglichkeit haben können, auf die Programmierung nach ethischen Gesichtspunkten Einfluss zu nehmen und eine gemeinsame Lösung zu erarbeiten.
2. *Zeitdimension*: Das Trolley-Problem sieht den Entscheider meist unter Zeitdruck; es muss im „hier und jetzt“ entschieden werden, bevor ein Unglück geschieht. Es geht um „split-second decisions“, die über Leben und Tod entscheiden. Beim autonomen Fahrzeug als Zukunftstechnologie und dessen Programmierung ist es wiederum völlig anders: Es gibt keine Zeitnot. Im Gegenteil: Es bestehen zeitliche Freiräume für eine deliberative Abwägung des Ethik-Designs, an dem sich eine Vielzahl von Personen und Institutionen beteiligen können, und nimmt somit eher die Gestalt einer „Kontingenzplanung“ an, also des Managements von potentiellen Unwägbarkeiten.
3. *Umweltparameter*: Trolley-Probleme konzentrieren sich auf wenige Faktoren, die das Szenarioergebnis verändern können. So werden die Gedankenexperimente einfach gehalten, ihre ethische Diskussion ist schließlich kompliziert genug. Bei der Programmierung autonomer Fahrzeuge kann man sich diesen

Luxus nicht erlauben: Die Sensoren der Fahrzeuge müssen potentiell *alle* Umweltfaktoren beachten, die auf das Fahrzeug einwirken können und vice versa. Das ist auch für die ethische Beurteilung relevant, weil damit besonders deutlich wird, dass es keine *sicheren* Ergebnisse geben wird, sondern eine potentiell unbeherrschbare Umweltkomplexität.

4. *Verantwortung*: Trolley-Probleme stellen die moralische und insbesondere die juristische Verantwortung in den Hintergrund. Bei der Programmierung autonomer Fahrzeuge müssen diese Fragen jedoch mit beachtet werden. Wer haftet? Wer trägt die strafrechtliche Verantwortung? Wo es im Trolley-Problem mitunter keinen Unterschied zwischen Tun oder Unterlassen geben mag, wird dieser etwa in der juristischen Betrachtung besonderer Bedeutung zukommen.
5. *Soziale Interaktion*: Trolley-Probleme blenden regelmäßig aus, welche Anreize eine bestimmte Handlung hervorruft. Insbesondere dann, wenn sie wiederholt wird. Bei der Programmierung autonomer Fahrzeuge müssen jedoch Anreize, die sich aus ihr ergeben, besonders berücksichtigt werden. Eine schadensoptimierende Dilemma-Programmierung würde es ggf. für Radfahrer unattraktiv machen, [einen Helm zu tragen](#). Eine Programmierung, die den Insassenschutz über alles stellt, würde ggf. den Verkauf von schweren SUV befeuern.

Anhand dieser kurzen Zusammenfassung der Kritik von Nyholm und Smids wird sogleich deutlich, dass sich die jeweiligen Entscheidungssituationen im Trolley-Problem und in der „real world“ fundamental voneinander unterscheiden. Nun beurteilt Ethik Handlungen. Um zu einem Urteil über die Ethik autonomer Fahrzeuge zu gelangen, muss man sich also in einem nächsten Schritt genau anschauen, wie die vermeintlich „intelligente“ Maschine im Fahrzeug *handelt* – wie sie *entscheidet*.

Sogleich wird man feststellen, dass

- a) die Maschine völlig anders entscheidet als ein Mensch und mithin auch völlig anders als ein Mensch, der sich einer Dilemma-Situation gegenüber sieht und
- b) dass sich aufgrund dieses Entscheidungsdesigns ganz andere Arten von ethischen wie sozialen Dilemmata ergeben.

Dazu mehr im nächsten Beitrag dieser Reihe.

Reihe „Ethik und autonomes Fahren“

1. [Was autonomes Fahren mit Ethik zu tun hat](#)
2. [Ethik & autonomes Fahren II: Trolley-Problem](#)
3. [Ethik & autonomes Fahren III: Das Problem mit dem Trolley-Problem](#)

4. [Ethik & autonomes Fahren IV: Wie autonome Fahrzeuge wirklich entscheiden](#)
5. Ethik & autonomes Fahren V: Warum das Trolley-Problem doch wichtig ist
6. Ethik & autonomes Fahren VI: Ein selbstbestimmtes Ethik-Setting für mehr Akzeptanz?
7. Ethik & autonomes Fahren VII: Die deontologische Kritik an der Schadensoptimierung
8. Ethik & autonomes Fahren VIII: Recht und Dilemma
9. Ethik & autonomes Fahren IX: Die Ergebnisse der Ethik-Kommission „Automatisiertes und vernetztes Fahren“
10. Ethik & autonomes Fahren X: Ethik und Datenschutz

1. Nyholm, Sven; Smids, Jilles (2016): The Ethics of Accident-Algorithms for Self-Driving Cars: an Applied Trolley Problem? In: Ethical Theory and Moral Practice 19 (5), pp. 1275–1289. p. 1286. [↩](#)

[← zurück](#)

[weiter →](#)

Über

Auf dem Autonoblog schreiben Wissenschaftler unterschiedlicher Disziplinen über sozialwissenschaftliche, ethische wie rechtliche Aspekte des autonomen Fahrens. Unter Leitung von Dr. David Knollmann und Prof. Volker Lüdemann informiert das Autoren-Team regelmäßig über neueste Entwicklungen rund um das autonome Fahren. Der Autonoblog ist eine Publikation des [Niedersächsischen Datenschutzzentrums \(NDZ\)](#), einer wissenschaftlichen Einrichtung der [Hochschule Osnabrück](#), und des vom [Europäischen Fonds für regionale Entwicklung](#) geförderten Forschungsprojekts „[Demokratie des Fahrens – Sollen Autos moralische Entscheidungen treffen?](#)“ (DeFrAmE).

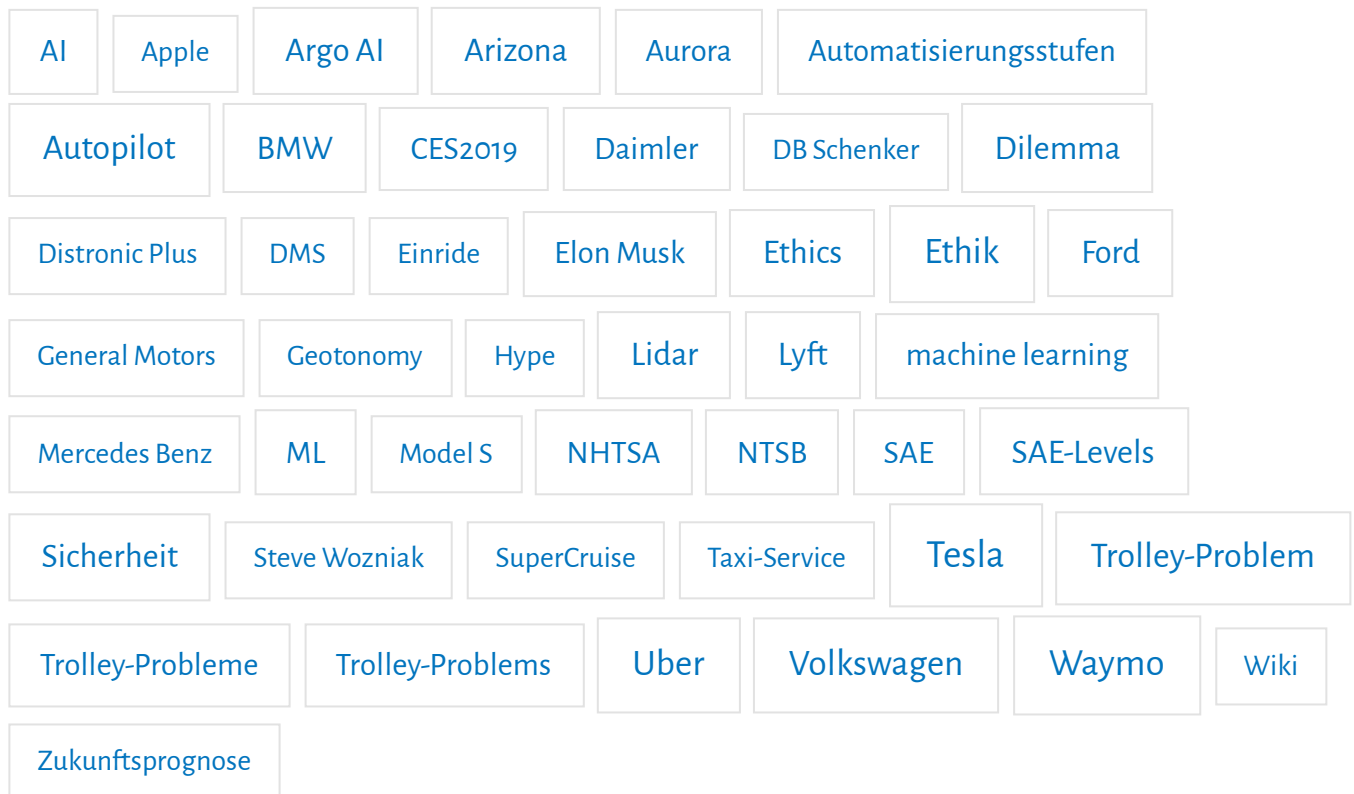




HOCHSCHULE OSNABRÜCK
UNIVERSITY OF APPLIED SCIENCES



Tags



Kategorien

[Analyse](#)

[Ethik & autonomes Fahren](#)

[Kommentar](#)

[Longform](#)

[News](#)

[Wiki](#)

Neueste Beiträge

[News zum autonomen Fahren {KW16/2019}](#)

[Ethik & autonomes Fahren IV:](#)

[Wie autonome Fahrzeuge wirklich entscheiden](#)

[Ethik & autonomes Fahren III:](#)

[Das Problem mit dem Trolley-Problem](#)

[News zum autonomen Fahren {KW13/2019}](#)

[Ethik & autonomes Fahren II: Trolley-Probleme](#)

Archive

[April 2019](#)

[März 2019](#)

[Februar 2019](#)

[Januar 2019](#)

[Dezember 2018](#)

[November 2018](#)

Copyright © 2020 Autonoblog

[Über](#) [Datenschutzerklärung](#) [Impressum](#)