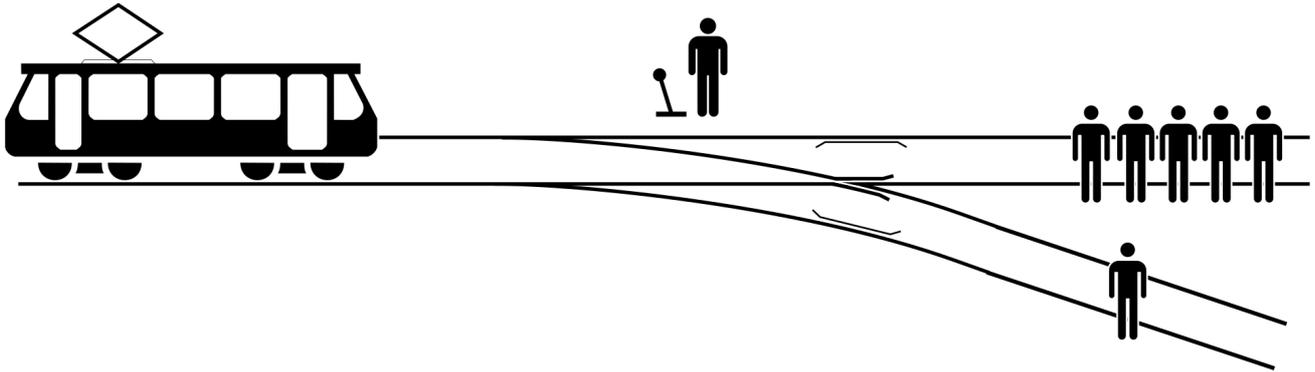


Autonoblog

Autonomes Fahren zwischen Hype und Wirklichkeit

Ethik & autonomes Fahren II: Trolley-Probleme

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



Schematische Darstellung eines Trolley-Problems (copyright: [Zapyon, CC BY-SA 4.0](#))

Wenn man sich das erste Mal mit der Ethik des autonomen Fahrens beschäftigt, ist eines so sicher wie das sprichwörtliche Amen in der Kirche:

Man wird schnell auf das sogenannte Trolley-Problem stoßen.

Das Trolley-Problem prägt die Debatte über die Ethik des autonomen Fahrens mittlerweile so sehr, dass häufig selbst Laien und regelmäßige Zeitungsleser eine Vermutung anstellen, wenn sie danach gefragt werden, worum es gehen könnte: „Ach, ist das nicht das, wo man sich entscheiden muss, ob man die Großmutter oder das Kind tot fährt?“.

Die großzügige Antwort muss lauten: ja, das stimmt.

Allerdings erweitert um die gewichtige Einschränkung: Es ist, wie so häufig bei näherer Betrachtung, komplizierter.

Trolley-Probleme sind kompliziert

Das sogenannte Trolley-Problem wurde erstmals 1967 von der britischen Philosophin [Philippa Foot](#) beschrieben¹ und seitdem von zahlreichen Philosophen aufgegriffen, weiterentwickelt und diskutiert – maßgeblich etwas von

der amerikanischen Philosophin [Judith Jarvis Thomson](#), auf deren Arbeit wir uns hier kurz stützen werden.

Unter dem Begriff Trolley-Problem werden heute zahlreiche philosophische „Gedankenexperimente“ beschrieben, weshalb korrekterweise von *Trolley-Problemen* gesprochen werden sollte, weil es das *eine* dominierende Problem nicht gibt. All diesen Gedankenexperimenten ist gemein, dass sie eine *dilemmatische Entscheidungssituation* beschreiben. Diese zeichnet sich dadurch aus, dass es darin keine offensichtlich und intuitiv als solche erkennbare *objektiv* „gute“ oder „richtige“ Entscheidung gibt, weil jede Option mit hohen Kosten verbunden ist und als unterträglich erscheint. Man hat die Wahl zwischen zwei Übeln, will sich deshalb eigentlich gar nicht entscheiden, ist aber mitunter dazu gezwungen.

Im ursprünglichen Beispiel bei Foot wird der Führer einer außer Kontrolle geratenen Straßenbahn mit folgendem Dilemma konfrontiert: Lässt er die Bahn weiter rollen (die Bremsen haben versagt), so werden fünf Gleisarbeiter mit Sicherheit sterben, die sich im Weg der Straßenbahn befinden. Der Führer hat jedoch die Möglichkeit, auf ein benachbartes Gleis zu wechseln, auf dem sich lediglich ein Gleisarbeiter befindet. Auch er wird mit Sicherheit von der Straßenbahn erfasst und sterben müssen².

Zwar ließen sich beide Optionen semantisch danach unterscheiden, dass es sich einerseits im Falle des „Nichtstun“ um ein passives „Sterben lassen“ („letting die“) und andererseits im Falle der „Weichenstellung“ um ein aktives, wenn auch mittelbares „Töten“ („killing“) handelt, in der Konsequenz läuft es aber auf das gleiche unerwünschte Ergebnis hinaus³, wie Thomson feststellt:

„*Whichever [one] does, turn or not turn, [one] kills.*“⁴

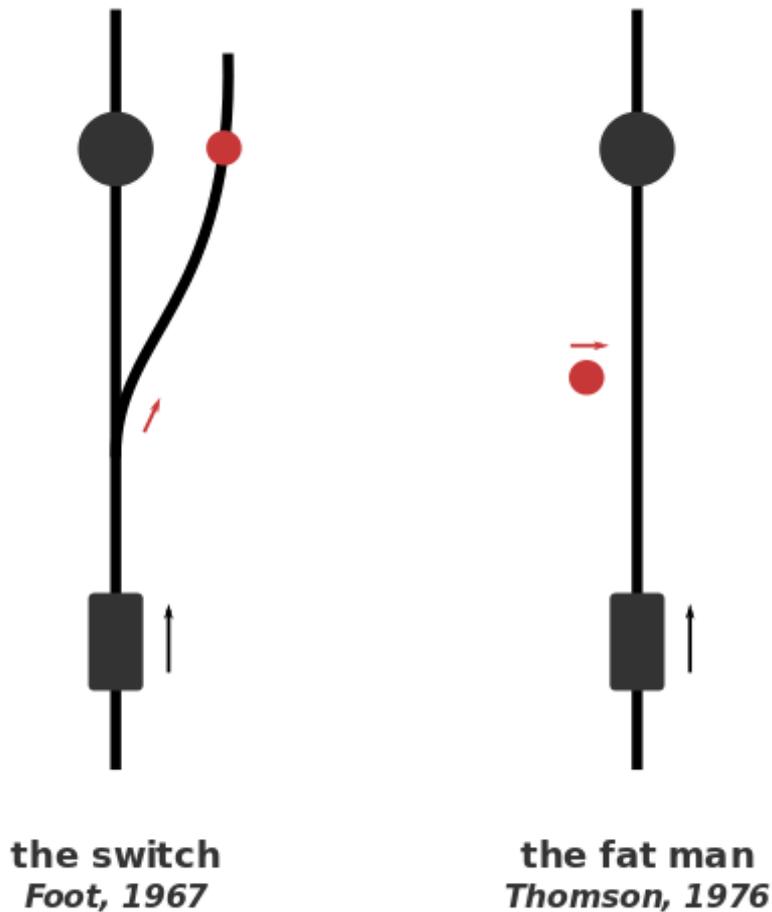
Der Konflikt besteht nun nach Thomson nicht zwischen einer *„positive duty to save five and a negative duty to refrain from killing one“*, sondern zwischen einer *„negative duty to refrain from killing five and a negative duty to refrain from killing one.“* Thomson argumentiert, dass die negative Pflicht, die Weiche zu stellen, um den Tod von Fünf vs. Einem zu vermeiden, stärker („more stringent“) sei als vice versa, so dass der Fahrer die Weiche stellen *dürfe*. Er sei zwar moralisch nicht dazu *verpflichtet*, so zu handeln, die Handlung sei jedoch *erlaubt*, glaubt Thomson ihrer Intuition. Ein Handlungsgebot („positive duty“) verneint Thomson also: Wer sich also nicht entscheiden mag, wird nicht eindeutig „falsch“ handeln.

Bereits mit diesem kurzen Abriss der Diskussion Thomsons wird deutlich: Bei der moralphilosophischen Diskussion wird es mitunter recht kompliziert, es werden Bedeutungen von Begriffen diskutiert und es scheint – jedenfalls aus der Sicht des Laiens – auch um sprachliche Nuancen zu gehen. So ist nicht etwa nur die Rede von Handlungspflichten, sondern diese werden unterschieden nach einer positiven oder einer negativen Pflicht. Letzteres meint nach Thomson, dass man bei Vorliegen einer solchen nicht handeln *muss*, aber die Handlung

doch erlaubt sei. Weiter wird die Diskussion auch dann komplexer, wenn man das Trolley-Problem variiert. Wobei gilt, dass die oberflächliche Präsentation immer durch ein relativ simples Setting besticht, die weitere Diskussion aber immer anspruchsvoller wird.

Zahlreiche Variationen

Eine berühmte Variante des Trolley-Problems stammt von Thomson selbst und führt den „Fat Man“ ein – eine fiktive Person, die es aufgrund ihrer körperlichen Opulenz vermag, eine heranfahrende Straßenbahn zu stoppen.



Trolley-Problem-Variationen (copyright: [public domain](#), [CCo](#))

In diesem Beispiel steht der Entscheider nicht mehr an einer Weiche, die er umlegen kann, um den Fahrtweg der Straßenbahn umzulenken und den Tod von einer Person (roter Punkt) anstatt von fünf Personen (schwarzer Punkt) in Kauf zu nehmen. Stattdessen befindet sich der Entscheider auf einer Brücke über den Gleisen und zufällig befindet sich ein „fat man“ neben ihm. Wirft er den Mann über das Geländer der Brücke auf die Gleise, so stirbt der Mann, die Straßenbahn kommt zum Stehen und die Fünf werden gerettet. Die aktive Handlung (Weiche

umlegen, „Fat Man“ auf die Gleise werfen) hat in beiden Varianten den Tod von einer Person und die Rettung von fünf Personen zur Folge. Und obwohl die Ergebnisse gleich sind, **beurteilen die meisten Menschen beide Varianten intuitiv unterschiedlich**: Wo im ersten Beispiel eher dazu geneigt wird, die Weiche umzustellen, schrecken viele Befragte davor zurück, im zweiten Beispiel den „Fat Man“ zur Rettung Dritter zu instrumentalisieren und zu töten.



Trolley-Problem-Variante „Tunnel“ (copyright: [public domain](#), [Karora](#), [CCo](#))

Auf unterschiedliche Reaktionen wird man auch dann treffen, wenn in den Trolley-Problemen die eigene Betroffenheit in den Blick gerät, also von der dritten in die erste Person gewechselt wird. So **beschreibt Jason Millar das Beispiel** einer alpinen einspurigen Passstraße, auf der der Fahrer fährt. Vor einem Tunnel tritt unvermittelt ein kleines Mädchen auf die Straße – der Fahrer ist mit zwei Optionen konfrontiert: Weicht er dem Mädchen aus, wird er von der Straße in den Abgrund stürzen und sterben; setzt er die Fahrt unverändert fort, stirbt das junge Mädchen. Sollte der Fahrer sich zugunsten des Mädchen opfern? Sollte *ich* das Steuer herumreißen, um das Mädchen zu retten? In der Beurteilung macht es natürlich einen Unterschied, ob man eine Weiche umstellt, um dritte Personen zu retten, indem man eine Person tötet, oder ob man *sich selbst* opfert, um eine Person zu töten.

Denkt man an hochautomatisierte Fahrzeuge, so stellt sich die Frage, wie diese für derartige Fälle programmiert werden sollen: Sollte der Insassenschutz immer an erster Stelle stehen, wenn es darum geht, die Akzeptanz der Technik zu erhöhen? Würde überhaupt jemand ein hochautomatisiertes Fahrzeug kaufen, das die Insassen im

Zweifelsfall zugunsten Dritter opfert? Wäre aber ein unbedingter Insassenschutz ethisch zu rechtfertigen? Was, wenn im oben genannten Beispiel nicht ein Mädchen, sondern eine ganze Schulklasse auf der Passstraße steht und ein mit einer Person besetzter SUV auf Kollisionskurs fährt?



Beispielfoto Trolley-Variante „Radfahrer“ (copyright: [StockSnap](#), Free Pixabay License)

Komplizierter wird es auch, wenn man die Anreize für weitere Verkehrsteilnehmer in den Blick nimmt, die eine Programmierung autonomer Fahrzeuge für bestimmte Dilemma-Situationen hervorrufen kann. Dies lässt sich am Beispiel zweier Fahrradfahrer verdeutlichen, die sich ähnlich wie im ursprünglichen Trolley-Setting auf zwei möglichen Kollisionskursen eines Fahrzeuges befinden. Wiederum hätte die Maschine in diesem Fall nur die Wahl zwischen A und B, wobei beide Optionen mindestens die schwere Verletzung des Radfahrers bedeuten würden. Die Radfahrer unterscheiden sich im Beispiel jedoch danach, dass einer einen Radhelm trägt und der andere nicht. Würde die Maschine im Sinne einer „Schadensoptimierung um jeden Preis“ programmiert, müsste sie die Kollision mit dem helmtragenden Radfahrer präferieren, weil dort der Schaden weniger groß wäre als beim Radfahrer ohne Helm. In einem Regime, das das Tragen von Helmen zur Pflicht für Fahrradfahrer macht, würde dies gleichwohl die Regelverletzung durch den Radfahrer begünstigen, der keinen Helm trägt. In der Folge, so eine Annahme, würden immer weniger Menschen Helm tragen, weil das Risiko, von einem autonomen Auto „ausgewählt“ zu werden, unter der Prämisse der programmierten „Schadensoptimierung“ höher wäre.

Trolley-Probleme als Instrument, um ethische Theorien zu „testen“

Trolley-Probleme sind nun vor allem deshalb so beliebt, weil sie zunächst ein geeignetes Mittel sind, um a) Laien ethische Problemstellungen anschaulich zu machen und b) gleichsam ethische Theorien einem ersten „Test“ zu unterziehen. So lässt sich etwa auf Foots Trolley-Problem mit den Augen eines Deontologen blicken.⁵ Erklärt man das Gebot „Du sollst nicht töten“ zur deontologischen Prämisse, so ist jede Handlung verboten, weil jeweils ein Mensch sterben wird. **Konsequentialisten haben es da einfacher** – sie beurteilen die Sache vom Ende her: Wo nur einer stirbt und nicht fünf, da mag die Handlung erlaubt sein.

Thomsons Diskussion von Foots Ursprungsbeispiel zeigt gleichwohl, was unter Moralphilosophen längst bekannt ist: „Reinformen“ deontologischer wie konsequentialistischer Ethik werden selten vertreten. So spricht Thomson von (negativen oder positiven) Pflichten, was auf ein deontologisches Ethik-Verständnis deutet, blickt gleichwohl auf das Ergebnis der Handlung (fünf Tote vs. ein Toter) und erkennt sodann, dass die negative Pflicht, die Weiche umzustellen und nur einen Toten in Kauf zu nehmen, stärker sei als die negative Pflicht, die Weiche nicht umzustellen, was wiederum auf ein eher konsequentialistisches Ethik-Verständnis deutet.

Und nicht jeder Philosoph, der sich als Utilitarist bezeichnet⁶, würde befürworten, den „Fat Man“ auf die Gleise zu werfen. So ist ein „regelbasierter“ Utilitarismus denkbar, der es grundsätzlich für verwerflich hält, Menschen gegen ihren Willen für einen (vermeintlich) höheren Zweck zu instrumentalisieren. Denn eine Gesellschaft, in der Menschen für einen „guten Zweck“ gegen ihren Willen geopfert werden, ist womöglich nicht wünschenswert. Jedenfalls werden die meisten Utilitaristen unterschreiben, dass es nicht *immer* eine moralisch gute Handlung ist, jemanden gegen seinen Willen zu töten, um einen anderen zu retten.

Ohnehin mag versöhnlich anklingen, dass Deontologen und Konsequentialisten in vielen Dingen womöglich gar übereinstimmen, wie der Philosoph Derek Parfit glaubt: „These people are climbing the same mountain on different sides.“⁷

Autonome Autos für Trolley-Probleme programmieren?

Sogleich sollte deutlich werden, dass man auch autonome Autos nicht nach einer ethischen „Generaltheorie“ programmieren wird können, weil es diese gar nicht gibt. Technisch gesehen ist es aber überhaupt kein Problem, verschiedene „Ethiken“ miteinander zu kombinieren und zu programmieren. So könnte eine auf einer konsequentialistischen Ethik basierende Kostenfunktion um auf deontologischer Ethik basierende „constraints“ erweitert werden, wie Gerdes und Thornton konstatieren:

„[...] for certain objectives, the trade-offs implicit in a cost function may obscure the true importance or priority of specific goals. It may make sense to penalize both large steering changes and collisions with pedestrians but there is a clear hierarchy in these objectives. Instead of simply trying to make a collision a [...] million times more costly

than a change of steering angle, it makes more sense to phrase the desired behavior in more absolute terms: the vehicle should avoid collisions regardless of how abrupt the required steering might be. The objective therefore shifts from a consequential approach of minimizing cost to a deontological approach of enforcing certain rules.”⁸

Schließlich gibt es jedoch ein zentrales Problem bei der Betrachtung von Trolley-Problemen und autonomen Fahren:

Die Ethik des autonomen Fahrens hat womöglich gar nicht soviel mit Trolley-Problemen zu tun.

So gesehen lenkt die Beschäftigung mit Trolley-Problemen von den *tatsächlichen* ethischen Herausforderungen ab, die mit der Automatisierung des Fahrens entstehen. Dazu demnächst mehr.

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. Foot, Philippa (1967): [The Problem of Abortion and the Doctrine of the Double Effect](#). In: The Oxford Review, Number 5. [↩](#)
2. Foot, Philippa (1967): [The Problem of Abortion and the Doctrine of the Double Effect](#). In: The Oxford Review, Number 5, p. 7. [↩](#)
3. Juristisch besteht gleichwohl ein nicht unerheblicher Unterschied zwischen [Tun und Unterlassen](#). [↩](#)
4. Thomson, Judith Jarvis (1976): Killing, Letting Die, and the Trolley Problem. In: Monist 59 (29), pp. 204-217, p. 206. [↩](#)

5. Die Unterschiede zwischen deontologischer und konsequentialistischer Ethik habe ich [hier kurz erläutert](#).



6. Wir erinnern uns: [Der Utilitarismus ist eine Variante konsequentialistischer Ethik](#). 

7. Parfit, Derek (2013): On What Matters Vol. 1. The Berkeley Tanner Lectures. Oxford: Oxford University Press. p. 419. 

8. Gerdes, J. Christian/Thornton, Sarah M. (2016): Implementable Ethics for Autonomous Vehicles. In: Maurer et al.: Autonomous Driving. Technical, Legal and Social Aspects. Heidelberg: Springer. pp. 87-102. p. 94.



[← zurück](#)

[weiter →](#)

Über

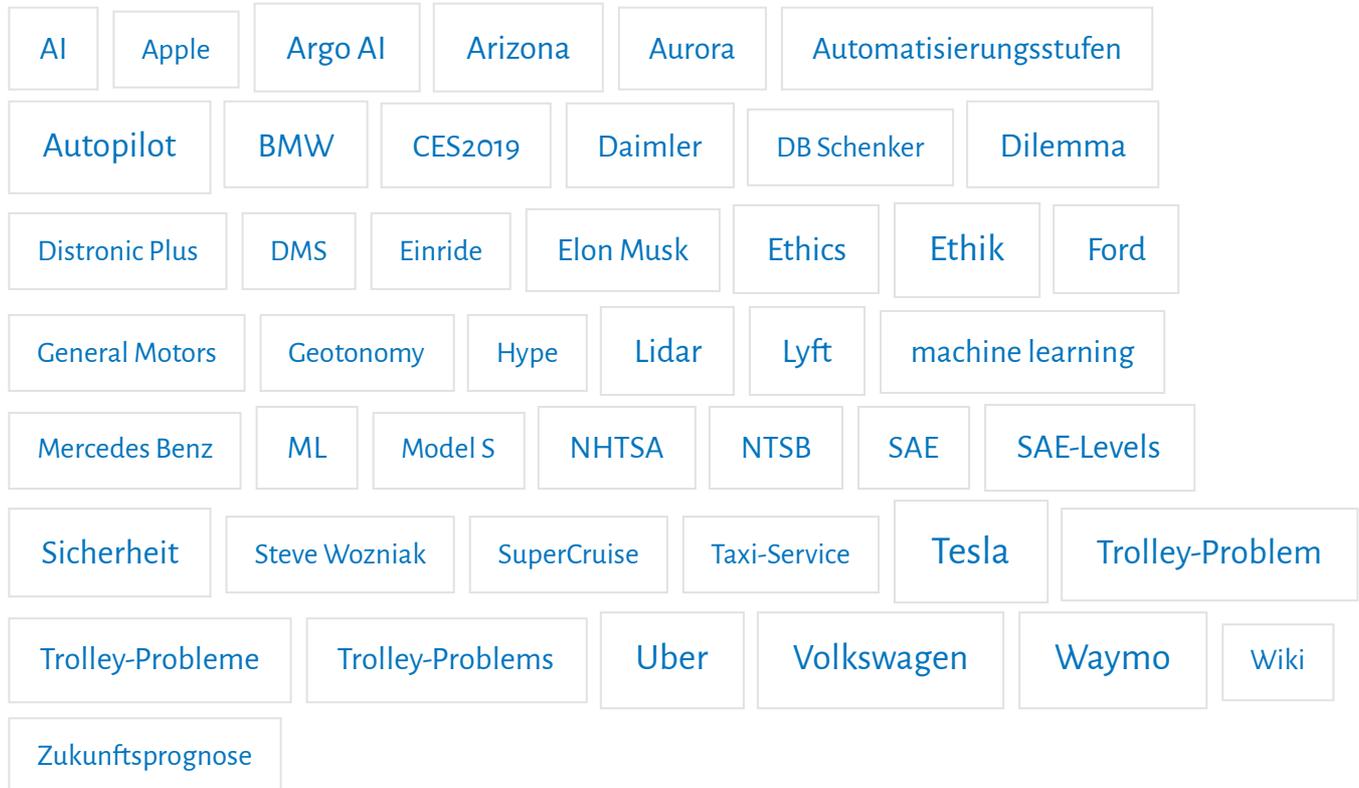
Auf dem Autonoblog schreiben Wissenschaftler unterschiedlicher Disziplinen über sozialwissenschaftliche, ethische wie rechtliche Aspekte des autonomen Fahrens. Unter Leitung von Dr. David Knollmann und Prof. Volker Lüdemann informiert das Autoren-Team regelmäßig über neueste Entwicklungen rund um das 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](#)