

**Freie Universität Berlin, Fachbereich Mathematik und Informatik
Institut für Informatik, Studiengang Monobachelor Informatik**

31.03.2020

Autonomes Fahren: Das Trolley Problem und seine Konsequenzen

Von

Nina Papenfuß

5007688
papenfun@zedat.fu-berlin.de

Wissenschaftliche Ausarbeitung zum Vortrag im Seminar:
Machine Ethics and Normative Reasoning

Bei Christoph Benzmüller
Im Wintersemester 2019/2020

Inhaltsverzeichnis

Einleitung	3
Philosophischer Hintergrund	3
Praktische Lösungsansätze	6
<i>Das Moral Machine Experiment und seine Ergebnisse</i>	<i>6</i>
<i>Bericht der Ethik-Kommission für automatisiertes und vernetztes Fahren.....</i>	<i>8</i>
Konsequenzen	10
Literaturverzeichnis	12

Einleitung

Angenommen Sie fahren ein Auto um eine Kurve. Plötzlich befinden sich fünf Menschen vor Ihnen, die die Straße überqueren. Sie könnten nach links ausweichen – dort befindet sich lediglich eine Person. Angenommen Sie hätten ausreichend Zeit, diese Entscheidung zu treffen, was würden sie tun? Möglichst viele Menschen verschonen und die eine Person auf der linken Seite verletzen, oder gar töten? Oder die Spur halten und in Kauf nehmen, dass fünf Menschen sterben? Vielleicht lehnen Sie sich jetzt zurück in dem Glauben, dass ein solches hypothetisches Dilemma niemals eintreffen werde. In dem Bewusstsein, dass ein Mensch in einer solchen Extremsituation nicht innerhalb von Sekunden nach moralisch-rationalen Prinzipien entscheiden könne.

Ihre Einschätzung mag stimmen, doch im Kontext des autonomen Fahrens stellt sich die Frage unausweichlich. Egal wie vorausschauend gefahren wird, plötzliche, unvermeidlich tödliche Unfälle können und werden auftreten. Das autonome Fahrzeug¹ muss entscheiden, wie deren Folgen zu optimieren sind – das heißt wie Schaden minimiert und *verteilt* werden soll (Vgl. Lin 2015). Und die ethischen Prinzipien, nach denen über diese Verteilung entschieden wird, müssen programmiert werden. Anders formuliert: verfolgen wir das Ziel, autonome Fahrzeuge in das öffentliche Leben zu integrieren, müssen wir notwendigerweise einen Umgang mit dem geschilderten moralischen Dilemma finden. Die Natur dieser Entscheidung verändert sich, denn wir müssen ex-ante reflektieren, was die Bedingungen und Umstände, unter denen es als moralisch akzeptabel angesehen wird zu töten, sind. In einem ersten Schritt wird in den philosophischen Hintergrund des Trolley Dilemmas eingeführt. In einem zweiten Schritt werden die praktischen Lösungsansätze des *Moral Machine Experimentes* und des Berichtes der *Ethik-Kommission für automatisiertes und vernetztes Fahren* skizziert und diskutiert. Schließlich werden Konsequenzen für ein weiteres Vorgehen benannt.

Philosophischer Hintergrund

Um einer praktischen Lösung näher zu kommen, lohnt sich ein Blick in die Moralphilosophie und die Ethik, in die Debatten rund um moralische Dilemmata. Unter einem moralischen Dilemma versteht man eine Situation, in der ein Akteur moralisch dazu verpflichtet ist, mehrere Handlungen auszuführen. Er kann jeweils eine der möglichen Handlung ausführen, aber nicht alle. Der Akteur kann nur moralisch versagen, da er immer etwas Falsches tun wird,

¹ Im Folgenden auch mit AV abgekürzt. Steht für autonomous vehicle.

bzw. eine Handlung unterlässt, die er ausführen sollte (Vgl. SEP MD). Im Unterschied zu einer *Konfliktsituation*, lässt sich ein solches dilemmatisches moralisches Entscheidungsproblem in keine Richtung auflösen, indem z.B. widersprüchliche moralische Gebote hierarchisiert werden. In der Philosophie wird in diesem Kontext von *echten* moralischen Dilemmata gesprochen (Vgl. SEP MD).

In der oben geschilderten Situation – einer Abwandlung des so genannten Trolley Problems oder auch „Weichenstellerproblem“² – ist das autonome Fahrzeug mit der Entscheidung konfrontiert, ob es in der Spur bleiben und entweder *fünf* Personen zu töten, oder nach links zu schwenken und *eine* Person zu töten. Die moralische Pflicht nicht zu töten, verbietet beide Handlungen, und doch muss es sich entscheiden. Einer utilitaristisch-konsequentialistischen³ Theorie folgend, könnte *für* ein Umschwenken nach links argumentiert werden, da hier die Anzahl der überlebenden Menschen in der Konsequenz höher ist. Neben der reinen Beurteilung der Konsequenzen kann außerdem – insbesondere wenn es darum geht, unter der Fülle an Trolley Problemen zu differenzieren⁴ – mit der Länge der Kausalkette oder mit der Umleitung, im Unterschied zur Neuschaffung, einer Bedrohung argumentiert werden. Im genannten Beispiel würde die Kausalkette durch das Ausweichen nicht verlängert, und die real existierende Bedrohung des sich einer Menschenansammlung nähernden AVs umgelenkt, aber nicht neu geschaffen. Mit einer streng deontologischen Ethik könnte *gegen* das Ausweichen nach links argumentiert werden. Das Verbot zu töten steht über dem Gebot Menschen zu retten. Ein Mensch darf nie als Mittel zum Zweck behandelt werden und seine Rechte und Würde müssen stets akzeptiert werden. Eine rein konsequentialistische Argumentation stößt genau hier an ihre Grenzen, denn sie missachtet das Instrumentalisierungsverbot. Das Problem einer rein deontologischen Argumentation hingegen besteht darin, dass sie nicht in allen Situationen widerspruchsfreie Handlungsanweisungen gibt – was im Kontext von autonomen Systemen jedoch notwendig ist (Vgl. Goodall 2014).

In der Diskussion des originalen Trolley Problems und seiner Variationen wird zumeist zwischen „sterben lassen“, also nicht intervenieren, und „töten“, also *aktivem* Umschwenken, unterschieden.⁵ Das aktive Töten wäre demnach das stärkere Verbot im Gegensatz dazu, dem

² Siehe SEP PF. Für einen systematischen Überblick der philosophischen Diskussion des Trolley Problems siehe Bruers und Braeckman 2014.

³ Für den philosophischen Hintergrund siehe: SEP CON.

⁴ Für einen systematischen Überblick siehe Bruers und Braeckman 2014.

⁵ Thomson 1985. Für eine Kritik an Thomson siehe Bruers und Braeckman 2014.

Kausalzusammenhang seinen Lauf zu lassen. In der Beurteilung spielt die Intention des moralischen Akteurs eine zentrale Rolle, wie beispielsweise bei der Doktrin des Doppeleffektes (Vgl. SEP DE). Im Kontext von autonomen Fahrzeugen greifen diese psychologischen Erwägungen nicht. Anthropozentrische Konzepte wie *agency*, also bewusst-eingreifendem Handeln eines Akteurs mit freiem Willen, können nicht auf ein AV übertragen werden. Dennoch muss an die im Vorhinein getroffene Programmierentscheidung ein ebenso moralischer Maßstab angelegt werden. Mit Blick auf das Trolley Problem des AV gibt es also legitime Gründe und Fallstricke *für* und *gegen* eine Entscheidung zum Ausweichen und der Streit um die richtige Entscheidung ist auch in der Philosophie nicht entschieden.

Bisher haben wir das Trolley Problem als ein Gedankenexperiment betrachtet und versucht, es auf dem Terrain der Ethik zu lösen. Praktisch lassen sich die Lösungsansätze erweitern: Da die Moralphilosophie kein eindeutiges, also unumstrittenes Regelsystem bereitstellt, könnte sich *erstens* in einer solchen Situation zufällig entschieden werden, also eine Münze geworfen werden. Diese Entscheidung basiert somit auf keiner moralischen Grundlage. Da es sich hier aber um Entscheidungen über Leben und Tod handelt, sollten diese im Zweifel auch begründet und verteidigt werden können. Somit ist der Münzwurf kein sinnvoller Umgang mit dem Dilemma. Hinzu kommt, dass es für andere Verkehrsteilnehmer bei einer Zufallsentscheidung nicht vorauszusagen ist, wie sich ein Fahrzeug verhält. Im Gegensatz zu einem nicht-autonomen Fahrzeug, kann beispielsweise kein Blickkontakt mit dem Fahrer aufgenommen werden, was zu Misstrauen gegenüber autonomen Fahrzeugen führen könnte. Eine *zweite* Lösung könnte sein, die Fahrzeugkontrolle dem Fahrer zu übergeben. Diese *man-in-the-loop*-Lösung wird gerne herangezogen, da hier im Zweifel der Fahrer juristisch zur Rechenschaft gezogen würde. Das Problem ist allerdings, dass der Fahrer in einer plötzlichen Unfallsituation keine Zeit hat sich einen Überblick zu verschaffen, und dementsprechend nicht moralisch entscheiden kann. Er würde quasi-zufällig handeln und diese Lösung weist ergo dieselben Probleme wie der Münzwurf auf.

Eine konsistente und moralische Lösung des Trolley Problems und seiner Abwandlungen muss die Vielfalt der moralischen Intuitionen der Menschen zur Grundlage haben, denn moralische Normen und Werte sind historisch kontingent. Zugleich muss die Lösung durch einen Algorithmus beschrieben werden können.

Praktische Lösungsansätze

Das Moral Machine Experiment und seine Ergebnisse

Um eine genauere Kenntnis der alltäglichen moralischen Vorstellungen in Trolley-Dilemma-Situation zu erlangen, wurde das *Moral Machine Experiment* ins Leben gerufen. Die *Moral Machine* ist eine Online-Plattform, die Variationen des Trolley-Problems erzeugt, und den Nutzer zwischen zwei Übeln entscheiden lässt (Vgl. Awad et al. 2018). Das Szenario entspricht der klassischen Switch-Situation – das heißt das Fahrzeug kann die Spur halten oder in eine Richtung ausweichen. Das Fahrzeug fährt in beiden Optionen entweder in ein Hindernis, was den Tod der Fahrzeuginsassen zur Folge hat, oder in Personen, was deren Tod zur Folge hat. Insgesamt wurden so über einen Zeitraum von anderthalb Jahren 40 Millionen Entscheidungen von 2,3 Millionen Nutzern aus über 233 Ländern zusammengetragen und analysiert. Trotz seiner enormen Größe ist der Datensatz nicht repräsentativ und die Ergebnisse sollten kritisch hinterfragt werden. Dies scheint den Autoren des Experimentes bewusst, wenn sie den Anspruch formulieren, dass die Analyse von politischen Entscheidungsträgern bei der Formulierung von Richtlinien *beachtet* nicht aber blind befolgt werden sollte.

Die Auswertung der Daten hat ergeben, dass es drei *universale moralische Präferenzen* gibt: *Erstens* verschonen Menschen eher Menschen, als Tiere. *Zweitens* gibt es eine starke Tendenz dazu, so viele Leben wie möglich zu retten. Und *drittens* verschonen Menschen eher junge, als alte Leute. Die Personen, die am ehesten verschont wurden waren Babys, Kinder, Schwangere und Ärzte. Zudem konnte über ein Clustering eine Einteilung der Länder in drei Gruppen erlangt werden. Ergeben haben sich ein Süd-, ein Ost-, und ein Westcluster, zwischen denen die Gewichtung einzelner Präferenzen unterschiedlich ausfiel. So ist beispielsweise die Präferenz jüngere Menschen eher zu verschonen als ältere, im östlichen Cluster deutlich schwächer ausgeprägt als in den anderen Clustern. Zu beobachten ist außerdem, dass die von der Moral-Maschine aufgedeckten Vorlieben in hohem Maße mit den kulturellen und wirtschaftlichen Unterschieden zwischen den Ländern korrelieren. So korreliert der Grad an Individualisierung in einem Land mit der Präferenz, möglichst viele – individuelle – Leben zu retten. Aufgrund der Unterschiede zwischen den Clustern muss die Frage diskutiert werden, ob, und vor allem wie sich die Fahrweise des AVs sich an lokal herrschende Moralvorstellungen anpassen sollte. Dafür spricht, dass dies eine höhere Akzeptanz fördern könnte, da so die Fahrzeuge in vermeintlich geltende Moralsysteme integriert würden.

Um ein differenzierteres Bild der global variierenden Moralvorstellungen zu erhalten, wurde eine weitere Studie derselben Autoren durchgeführt (Vgl. Awad et al. 2020). In dieser Studie wurden die Szenarien durch das *Loop*- und das *Footbridge*-Szenario erweitert, allerdings in der Aufmachung der klassischen Straßenbahn auf Schienen. Im *Loop*-Szenario kann der Weichensteller entscheiden, ob die Straßenbahn geradeaus fährt und mehrere Menschen überrollt, oder ob die Straßenbahn auf ein Ausweichgleis umgeleitet wird. Dort würde sie eine Person umfahren, aber durch die Masse der Person zum Halt gebracht werden, und somit nicht auf das Hauptgleis mit den anderen Menschen zurückrollen. Im *Footbridge*-Szenario⁶ gibt es nur ein Gleis, auf dem sich mehrere Personen befinden. Zwischen der Straßenbahn und den Personen befindet sich jedoch eine Brücke, auf der ein fatter Mann und der moralische Agent stehen. Der moralische Agent könnte den fetten Mann die Brücke hinunterstoßen und somit die Straßenbahn zum Stoppen bringen, bevor sie in die Menschenmenge hineinrollt. Der Großteil der Teilnehmer der Studie fand es für das *Switch*-Szenario moralisch akzeptabel, die Weiche umzustellen, was dem Ergebnis des *Moral Machine* Experimentes entspricht. Der Grad der Akzeptanz, die Weiche umzustellen, nahm beim *Loop*-Szenario ab, und war beim *Footbridge*-Szenario am geringsten. Zudem wurde festgestellt, dass der Grad an Zustimmung zum Töten weniger Menschen für das Gemeinwohl mit der Ausprägung des Faktors der *relationalen Mobilität*⁷ in einem Land zusammenfällt.

Möchte man die Erkenntnisse der fortgesetzten Studie für autonomes Fahren fruchtbar machen, muss zum einen die Frage der Übertragbarkeit der abgefragten Trolley-Situationen beantwortet werden. Ein *Loop*-Szenario kann leicht modelliert werden. Der Unterschied zum *Switch* Szenario besteht darin, dass das AV kalkuliert einen Menschen umbringt, ihn also *instrumentell* benutzt, um das Fahrzeug zu stoppen und so vermeidet, eine größere Zahl an Menschen zu töten. Der Unterschied zum *Footbridge*-Szenario ist weniger einfach zu fassen, da eine, der Situation externe, Person aktiv zum Hindernis gemacht werden müsste. Mit dieser Beschreibung bleibt der Unterschied zum *Loop*-Szenario allerdings schwammig, da in beiden Fällen der Körper als Hindernis genutzt wird. Der Unterschied besteht im Grad der Involviertheit – einmal steht die Person auf einer Brücke, einmal auf einem Ausweichgleis. Und zumindest im Originalproblem gibt es einen Unterschied in der personalen Nähe des moralischen Akteurs. Es müsste erst einmal definiert werden, wann eine Person Teil des

⁶ In der Literatur auch „Fetter-Mann-Problem“ genannt. Ursprünglich formuliert von Thomson 1985.

⁷ Für den theoretisch-konzeptionellen Hintergrund siehe RM.

Geschehens ist, und wann nicht. Muss sie sich auf der Straße befinden? Auf dem Gehweg? Oder generell in Reichweite? Die Grenze ist nicht klar zu ziehen. Eine andere Modellierung des Footbridge-Szenarios setzt kommunizierende autonome Fahrzeuge voraus. In der modellierten Situation könnte dann ein AV A vor der Entscheidung stehen eine Person als Hindernis in die Unfallsituation des AV B zu involvieren, z.B. weil B an das A ein Notfall-Signal gesendet hat. Für eine genaue Programmierung eines moralischen autonomen Fahrzeuges sollten allerdings entsprechende Dilemma-Situationen modelliert werden, und weitere Studien durchgeführt werden. Unabhängig davon wie eine Modellierung aussieht kann festgehalten werden, dass moralische Entscheidungen, in denen Menschen instrumentalisiert werden, wenig Anklang finden.

Zugleich stellt sich die Frage der Vereinbarkeit der Ergebnisse mit national und international geltendem Recht. Im Folgenden sollen die Ergebnisse des Moral Machine Experimentes mit den Richtlinien der von der Bundesregierung einberufenen Ethikkommission zum autonomen Fahren verglichen werden.

Bericht der Ethik-Kommission für automatisiertes und vernetztes Fahren

Die Ethik-Kommission formuliert das Prinzip, im Zweifel Tier-, und Sachschäden in Kauf zu nehmen, sollten so Menschenleben ausgespart werden. Dieses Prinzip stimmt mit der im *Moral Machine Experiment* festgestellten Präferenz überein. Zugleich muss das AV eine präzise Folgeabschätzung leisten können, da auch Sachschäden einen Dominoeffekt in Gang setzen können, wodurch hohe Personenschäden entstehen können. Ein von der Ethikkommission genanntes Beispielszenario ist der Zusammenbruch eines Stromnetzes. Die Ethikkommission untersagt des Weiteren „jede Qualifizierung nach persönlichen Merkmalen (Alter, Geschlecht, körperliche oder geistige Konstitution)“, und folgt somit dem gesetzlich festgeschriebenen Diskriminierungs- und Instrumentalisierungsverbot, das aus Artikel 1, Absatz 1 des Grundgesetzes folgt. Es lautet: „Die Würde des Menschen ist unantastbar. Sie zu achten und zu schützen ist Verpflichtung aller staatlichen Gewalt.“ (Vgl. GG) Aus diesem Grund darf sich die dritte universale Präferenz – die Bevorzugung von Jungen im Gegensatz zu Alten – nicht in der Programmierung widerspiegeln.

In dem Papier der Ethikkommission heißt es weiter: „Eine *Aufrechnung* von Opfern ist untersagt. Eine allgemeine Programmierung auf eine *Minderung* der Zahl von Personenschäden kann vertretbar sein.“ [Markierungen von mir] Diese Passage bedarf einer

genaueren Analyse: Die Gegenüberstellung von der Aufrechnung von Personen und der Minimierung an Personenschäden scheint auf den ersten Blick widersprüchlich. Die zu beantwortende Frage ist, wie eine Minimierung von Personenschäden möglich ist, ohne die betroffenen Personen zu instrumentalisieren. Der zentrale Unterschied ist, dass es sich bei der Minimierung der Personenschäden um einer *allgemeine ex-ante* Entscheidung handelt. Für diese Art von Entscheidungen gelten andere Maßstäbe als im konkreten Trolley Dilemma⁸. In einer solchen, allgemeinen Situation, ist „eine Wahrscheinlichkeitsprognose aus der Situation zu treffen, bei der die Identität der Verletzten oder Getöteten noch nicht feststeht“ (Vgl. EK). Das Prinzip, im Zweifel mehr Menschen zu retten ist auch im Interesse der Geopferten, bevor sie „situativ als solche identifizierbar waren.“ (Vgl. EK) Grundsätzlich könnte jeder Beteiligte die geopferte Person sein. Das moralische Entscheidungsdilemma wird hier als eine *risikoethische* Problemstellung begriffen (Vgl. Schäffner 2020). Laut Vanessa Schäffner ist „Grundlage jeder risikoethischen Betrachtung [...] die Annahme, dass auch wenn es moralisch unzulässig ist[,] einem Individuum eine bestimmte Konsequenz *direkt* zuzufügen, dennoch Umstände vorliegen können, die es erlauben, das Individuum einem entsprechenden Risiko auszusetzen, infolgedessen die Konsequenz möglicherweise eintreten kann.“ (Vgl. Schäffner 2020) Die Frage, die sich nun bei der Implementierung stellt, lautet also, welches die einem Individuum zumutbaren Risiken sind, und wie eine gerechte Verteilung der Risiken zwischen allen Beteiligten aussieht. Dies betrifft nicht nur Entscheidungen in Extremsituationen, wie Bonnefon, Shariff und Rahwan mit ihrer Notation des *statistischen* Trolley Problems deutlich machen. Fußgänger und Radfahrer sind ungeschützter als Autos. Diese haben wiederum eine deutlich geringere Masse als Lastkraftwagen. Im gemischten Straßenverkehr bewegen sich all diese Akteure und beeinflussen durch kleine Fahrentscheidungen jeweils die Risikowahrscheinlichkeit der anderen in einen Unfall verwickelt zu werden. Aggregiert man diese Entscheidungen über Millionen von Fahrzeugen, so hat dies eine statistische Auswirkung. „Anstatt zu fragen, ob ein einziger Passagier sterben sollte, um fünf Fußgänger bei einem bestimmten Unfall zu retten, würden wir fragen, ob ein Passagier für fünf Fußgänger sterben sollte, und zwar bei allen in einem einzigen Jahr registrierten Unfällen.“ (Bonnefon et al. 2019) Dies ist eine Diskussion, die alle Beteiligten des öffentlichen Lebens mit einbeziehen sollte, und die deshalb nicht nur im Rahmen einer Ethik-Kommission diskutiert

⁸ Ein Bereich, in dem so verfahren wird ist die Medizin. Und zwar beim Zuteilen von Spenderorganen. Siehe Birnbacher und Birnbacher 2016.

werden sollte.

Konsequenzen

Empirische Studien, wie das *Moral Machine Experiment* haben gezeigt, dass utilitaristisch handelnde autonome Fahrzeuge eine hohe moralische Akzeptanz genießen. Auch wenn es durch das Grundgesetz untersagt ist, Menschenleben gegeneinander aufzurechnen, so kann dennoch dieser Präferenz mit einer risikoethischen Betrachtungsweise Rechnung getragen werden. Zur Weiterentwicklung eines solchen Ansatzes sollten entsprechende Studien – mit modellierten Fragestellungen, wie z.B. Risikozuweisungen⁹ oder auch Abfragen über die empfundene Gerechtigkeit eines beispielsweise rawl'schen Algorithmus¹⁰ – ins Leben gerufen werden. Entsprechend komplexe, bzw. differenzierte ethische Prinzipien sollten reguliert werden, da sich sonst soziale Dilemmata ergeben. Denn überlässt man die Regulierung den Kräften des Marktes, können Wettbewerbsvorteile auf Kosten der Allgemeinheit entstehen (Vgl. Birnbacher und Birnbacher 2016, Bonnefon et al. 2016). Die den autonomen Fahrzeugen inhärenten ethische Prinzipien müssen verständlich sein, und auf einen möglichst breiten gesellschaftlichen Konsens gründen.

Ein Ansatz zur Herstellung dieses Konsenses, der rein auf die empirisch zu ermittelnden Moralvorstellungen setzt, verkennt die Gefahr, vorhandene diskriminierende Neigungen innerhalb der Gesellschaft zu reproduzieren. Die Schaffung einer Ethik-Kommission für automatisiertes und vernetztes Fahren war daher ein Schritt in die richtige Richtung. Sie besteht nach Angaben des Bundesministeriums für Verkehr und digitale Infrastruktur aus Akteuren aus Wirtschaft, Wissenschaft, Verbänden und Verwaltung (Vgl. BMVI). An der Verfassung des Berichtes 2017 waren allerdings keine Vertreter der Zivilgesellschaft, insbesondere zivilgesellschaftlicher Initiativen im Bereich Verkehr beteiligt. Diese sollten ebenso Teil einer solchen Kommission sein, um eine hohe gesellschaftliche Akzeptanz zu ermöglichen. Hierfür sollten auch Stimmen aus anderen wissenschaftlichen Disziplinen mit einbezogen werden.¹¹ Denn Ziel autonomer Fahrzeuge sollte nicht nur sein, die Sicherheit auf den Straßen und Schienen zu erhöhen, sondern auch sozial fortschrittliche Systeme zu

⁹ Wie z.B. in Meder et al. 2019 referenziert.

¹⁰ siehe Leben 2017. Freilich muss bei der Abfrage auf die Form geachtet werden, damit die Abfrage nicht zu kompliziert wird.

¹¹ Bissell et al. 2018 argumentieren z.B. mit einem Mobilitätsansatz.

entwerfen, die der Mehrheit der Gesellschaft zu Gute kommen und Ungleichheiten entgegenwirken.

Auch wenn Unfälle, in denen ein autonomes Fahrzeug einen Menschen umbringt, selten sein werden, ist es wahrscheinlich, dass sie Aufsehen in der Öffentlichkeit erregen. Der Zwang, moralische Entscheidungen im Voraus zu programmieren, eröffnet die Möglichkeit, bewusste und reflektierte Entscheidungen zu treffen. Eine kollektive und breit geführte Diskussion über die ethischen Richtlinien und ihre Begründung ist in diesem Sinne im Interesse aller Beteiligten – den politischen Entscheidungsträgern, der Industrie, und nicht zuletzt auch der Zivilgesellschaft.

Literaturverzeichnis

- Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon J.-F., Rahwan, I.:** *The Moral Machine Experiment*. In: Nature 563, S. 59-64 (2018).
- Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon J.-F., Rahwan, I.:** *Universals and variations in moral decisions made 42 countries by 70,000 participants*. In: PNAS 117, S. 2332-2337 (2020).
- Birnbacher, D. und W. Birnbacher:** *Automatisiertes Fahren. Ethische Fragen an der Schnittstelle von Technik und Gesellschaft*. In: Information Philosophie 4, S. 8-15 (2016).
- Bissell, D., Birtchnell, T., Elliott, A., und Hsu, E.L.:** *Autonomous Automobilities: The social impacts of driverless vehicles*. In: Current Sociology, S.1-19 (2018).
- Bonnefon, J.-F. Shariff, A., Rahwan, I.:** *The social dilemma of autonomous vehicles*. In: Science 352, S. 1573-1576 (2016).
- Bonnefon, J.-F., Shariff, A., Rahwan, I.:** *The trolley, the bull bar, and why engineers should care about the ethics of autonomous cars*. In: Proceedings of the IEEE 107, S. 502-504 (2019).
- Bruers, S.und J.Braeckman:** *A Review and Systematization of the Trolley Problem*. In: Philosophia 42 (2014)
- Goodall, N. J.:** *Ethical decision making during automated vehicle crashes*. In: Transportation Research Record 2424, S. 58-65 (2014).
- Leben, D.:** *A Rawlsian algorithm for autonomous vehicles*. In: Ethics and Information Technology 19, S. 5-24 (2017).
- Lin, Patrick:** *Why Ethics matters for Autonomous Cars*. In: Autonomes Fahren, Maurer, Gerdes et al, S. 69-85 (2015).
- Meder, B., Fleischhut, N., Krumnau, N.-C., Waldmann, M.R.:** *How should Autonomous Cars Drive? A Preference for Defaults in Moral Judgements Under Risk and Unvertainty*. In: Risk Analysis 39, S. 295-314 (2019).
- Thomson, J. J.:** *The trolley problem*. In: Yale Law Journal 94, S. 1395-1415 (1985).
- Schäffner, V.:** *Wenn Ethik zum Programm wird: Eine risikoethische Analyse moralischer Dilemmata des autonomen Fahrens*. In: Zeitschrift für Ethik und Moralphilosophie (2020).
- <https://plato.stanford.edu/entries/moral-dilemmas/> (SEP MD)
- <https://plato.stanford.edu/entries/philippa-foot/> (SEP PF)
- <https://plato.stanford.edu/entries/consequentialism/> (SEP CON)
- <https://plato.stanford.edu/entries/double-effect/> (SEP DE)
- <https://www.bmvi.de/SharedDocs/DE/Publikationen/DG/bericht-der-ethik-kommission.html> (EK)
- <https://www.bmvi.de/DE/Themen/Digitales/Automatisiertes-und-vernetztes-Fahren/automatisiertes-und-vernetztes-fahren.html> (BMVI)
- https://www.bundestag.de/parlament/aufgaben/rechtsgrundlagen/grundgesetz/gg_01-245122 (GG)
- <http://relationalmobility.org/> (RM)