

Ethische Anforderungen an autonom fahrende Personenkraftwagen im Straßenverkehr

Anne Borchard

28. Januar 2017

Humboldt-Universität zu Berlin

Inhaltsverzeichnis

1	Einleitung	4
2	Begriffe und Annahmen	4
3	Grundlegende Problematik	5
4	Deontologische Sichtweise	5
5	Utilitaristische Sichtweise	6
6	Weitere Sichtweisen	7
7	Fazit	8

Abkürzungsverzeichnis

AA Autonomer Agent

IRA Intelligent Road Agent

StVO Straßenverkehrs-Ordnung

1 Einleitung

IRA (Intelligent Road Agent)¹ fährt in einer regnerischen Nacht durch die Stadt. Die Straßen sind leer und sie kann die Ampelschaltung optimal nutzen. Direkt nach einer Kurve sieht IRA auf der Straße zwei Frauen und bremst. Nasses Laub verlängert den Bremsweg drastisch, IRA hat nur folgende Optionen: bei Beibehaltung des Kurses beide Frauen treffen oder sich für eine Frau als Ziel entscheiden. Die Entscheidung fällt auf die Übergewichtige (Bezeichnung: A), da sie voraussichtlich eine geringere Lebenserwartung hat als die andere Frau. A stirbt an Ort und Stelle. IRA hat ihre Insassen geschützt und ihren Algorithmen entsprechend gehandelt. Leider war A nicht übergewichtig, sondern hochschwanger. Die Angehörigen von A werden versuchen, den Automobilhersteller für ihren Verlust haftbar zu machen.

Dieses Dilemma kann sich, wenn auch mit geringer Wahrscheinlichkeit, in einer Zukunft mit autonom fahrenden Fahrzeugen zutragen. Es zeigt, dass eine Auseinandersetzung mit den moralischen Dimensionen des autonomen Fahrens nötig ist. Welche ethischen Richtlinien sollen solche Fahrzeuge befolgen? Wer entscheidet darüber? Wie können Dilemma dieser Art gelöst werden? Besteht überhaupt Konsens über die „richtige“ Ethik?

2 Begriffe und Annahmen

Da es in der Literatur verschiedenen Graden der Automatisierung gibt, wird implizit vom höchsten Grad² ausgegangen. Zudem wird angenommen, dass der AA (autonomer Agent) fehlerfrei funktioniert³.

Um in Deutschland ein Fahrzeug im öffentlichen Straßenverkehr bewegen zu dürfen, ist, sofern nicht anders vermerkt, eine Fahrerlaubnis nötig. Das Regelwerk hierzu bildet die FeV (Fahrerlaubnis-Verordnung). Ziel ist es, sowohl das Fahrzeug kontrollieren als auch im Rahmen der StVO (Straßenverkehrs-Ordnung) am Straßenverkehr teilnehmen zu können.

Da sich das Essay an Personen aus dem Bereich der Informatik richtet, werden im Folgenden zwei hierfür relevante Formen der Ethik kurz dargestellt. Zunächst wird die *Deontologische Ethik* erläutert, welche auch als *Deontologie* bezeichnet wird. Dabei handelt es sich um eine moralische Handlungsanweisung, welche eine strikte Befolgung von Regeln oder auch Pflichten verlangt (vgl. [2], S. 47). Die Konsequenzen einer Handlung, egal ob im Allgemeinen als wünschenswert

¹Fiktives, voll-autonomes Assistenzsystem

²Fahrerloses Fahren, Insassen sind reine Passagiere (vgl. [1], S. 6).

³Implementierungsfehler, fehlerhafte Sensoren oder unbefugte Eingriffe von Außen werden ausgeschlossen, so dass der AA in einem idealen Umfeld agieren kann. Ausgenommen ist die korrekte Interpretation der genommenen Messdaten.

angesehen oder nicht, werden nicht betrachtet. Des weiteren ist die *Utilitaristischen Ethik* als Ausprägung der *Konsequenzialistische Ethik* wichtiger Bestandteil folgender Kapitel. Hierbei werden moralische Entscheidungen ausschließlich anhand ihrer direkten und indirekten Folgen getroffen. Baase [2] formuliert wie folgt: „we consider the impact on utility and judge the action by its net impact“ ([2], S. 48).

3 Grundlegende Problematik

Zu den angestrebten Zielen durch die Verwendung von AA gehören die drastische Reduzierung von Verkehrstoten, die effizientere Nutzung der Infrastruktur und das Integrieren von Personenkreisen, welche zuvor nicht am Straßenverkehr ungehindert teilnehmen konnten. Die Einführung von AA ist unter Betrachtung dieser Aspekte für weite Teile der Bevölkerung wünschenswert. Im Gegensatz zu einem menschlichen Fahrer werden an einen AA offensichtlich Anforderungen gestellt, welche über technische Aspekte hinaus gehen. Nürnberger [3] formuliert diesen Aspekt treffend: „Interessanter Weise haben wir an Algorithmen höhere ethische Ansprüche als an uns selbst“ ([3], S. 2).

Den erhofften Vorzüge stehen jedoch einer neuen Problematik gegenüber. Ein AA kann in einer wie in der Einleitung geschilderten Situation *Entscheidungen treffen*, während ein menschlicher Fahrer in Bruchteilen von Sekunden fast ausschließlich *reagieren* kann. Der Programmierer⁴ des Entscheidungsalgorithmus hätte genug Zeit im Voraus, einen Ablaufplan für diverse Situationen vorzugeben. Es besteht Unklarheit darüber, welche ethischen Richtlinien bei der Entscheidungsfindung zu beachten sind und ob diese auch gesellschaftliche akzeptiert werden. Im Folgenden werden verschiedene ethische Sichtweisen und teilweise auch Lösungsvorschläge erläutert.

4 Deontologische Sichtweise

Eine deontologische Ethik scheint keine Lösung für Dilemma-Situationen zu bieten. Regeln wie die StVO können als Rahmen für *alltägliche* Situationen im Straßenverkehr dienen, müssen aber ggf. gebrochen werden. Aus diesem Grund würden für AA „soft constraint“ ([4], S. 94) gebraucht, die den Umständen entsprechend gebrochen werden dürfen. Diese Forderung widerspricht der reinen Deontologie.

⁴Aus Gründen der Lesbarkeit wird in vielen Fällen die maskuline Form eines Wortes gewählt, jedoch sind beide Geschlechter angesprochen.

5 Utilitaristische Sichtweise

Computer sind für den Umgang mit Zahlen prädestiniert. Die utilitaristische Ethik kann vergleichsweise einfach als Algorithmus implementiert werden. Elemente des Straßenverkehrs könnten als Variablen mit festgelegten Werten belegt, im Entscheidungsfall miteinander abgewogen werden und müssen gewissen Konstanten entsprechen (z.B. StVO). Es gestaltet sich schwierig, sowohl den richtigen Wert jener Elemente zu ermitteln, als auch nicht-quantifizierbare Attribute – z.B. der „Wert“ eines Menschen – für einen Computer messbar zu machen.

Gogoll und Müller [5] stellten sich zunächst die Frage: Sollte die Ethik eines AA durch eine festgelegte Instanz oder durch den Einzelnen selbst festgelegt werden (vgl. [5], S. 3). Letzteres lässt zu befürchten, dass der AA als „moral proxy“ ([5], S. 8) seines Besitzers dessen mitunter diskriminierende Haltung gegenüber bestimmten Bevölkerungsgruppen ausübt. Ihren Berechnungen zufolge bestünde das Optimum – hier das Wohl der Allgemeinheit – darin, ein utilitaristisches, staatlich vorgegebenes Modell zu implementieren, was im Extremfall die Opferung der eigenen Insassen des AA einschloße (vgl. [5], S. 12 ff.). Jedoch räumen sie ein, dass viele Menschen sich gegen einen selbst opfernden AA entscheiden würden, obwohl dieser im Sinne aller wäre (vgl. [5], S. 18).

In der Annahme, dass sich Unfälle mit AA nicht vermeiden lassen werden, spricht Lin [6] gar von einer „crash-optimization“ ([6], S. 72), welche für AA in Betrachtung gezogen werden könnte. Lin [6] führt an, der Blickwinkel entscheide hierbei, *wessen* Schaden minimiert werden soll. Wenn als „Ziele“ eine Grundschülerin, ihre 80-jährige Großmutter oder ein massiver SUV⁵ zur Auswahl stünden, gäbe es unterschiedliche Arten, den minimalen Schaden zu berechnen. Für die Insassen des AA wäre das Kind das leichteste Objekt und somit bei einer Kollision eine geringere Gefahr, verletzt oder getötet zu werden. Ist beispielsweise nur ein Insasse im AA, könnte es zielführend sein, in den SUV zu steuern, da dessen Insasse(n) mit großer Wahrscheinlichkeit unverletzt blieben. Somit wäre nur eine Person, der Insasse des AA, zu Schaden gekommen. Die ältere Dame käme unter dem Aspekt infrage, als dass sie bereits ein langes Leben geführt hat und es ein Unrecht wäre, das Leben von deutlich jüngeren Personen zu beenden (vgl. [6], S. 70 ff.). Problematisch wird es, falls bekannt ist, *wie* AA Unfälle zu vermeiden versuchen und „other drivers may be tempted to ‘game’ it“ ([6], S. 81). Eine konkrete Lösung stellt Lin [6] nicht vor, jedoch kommt er zu dem Schluss, dass eine gesellschaftliche Diskussion über Ethik in Bezug auf AA helfen kann, *realistische* Erwartungen an diese Technologie zu stellen (vgl. [6], S. 82).

Die zuvor genannten Modelle sollten kritisch hinterfragt werden: Auf der einen Seite würden

⁵Sport Utility Vehicle

bestimmte gesellschaftliche Gruppen (z.B. Senioren) öfter als „Ziel“ von AA gewählt werden als andere. Auf der anderen Seite sollte es die Aufgabe jedes demokratischen Staates sein, das individuelle Leben zu schützen und nicht für das Wohl einer Mehrheit zu opfern (vgl. [7], S. 622). Aus dieser Perspektive betrachtet wäre im Szenario aus der Einleitung die beste Entscheidung gewesen, emphkeine Entscheidung zu treffen und mit beiden Frauen zu kollidieren. Dies erscheint jedoch grausamer, als nur ein Todesopfer verantworten zu müssen (vgl. [6], S. 71).

Denkbar wäre zudem, dass ein utilitaristischer, potentiell sich selbst opfernder AA seine Einführung in den allgemeinen Straßenverkehr verzögern oder gar verhindern könnte, womit die erhofften Ziele des autonomen Fahrens nicht mehr erreicht würden. Der Utilitarismus stünde sich paradoxerweise selbst im Wege (vgl. [8], S. 1575 f.).

6 Weitere Sichtweisen

Aus dem vorhergehenden Kapitel wird ersichtlich, dass eine rein deontologische bzw. utilitaristische Ethik voraussichtlich von vielen Menschen abgelehnt würde. Unter keinen Umständen Regeln wie die StVO zu brechen entspricht nicht realistischen Situation im heutigen Straßenverkehr und Berechnungen anzustellen, welche Person zum Wohle einer Mehrheit sterben möge, scheint unvereinbar zu sein mit dem Recht auf Leben und körperliche Unversehrtheit.

In Anbetracht der erwarteten Vorzüge des autonomen Fahrens mag es ein pragmatischer Ansatz sein, Dilemma-Situationen als Ausnahmefälle zu akzeptieren. Laut Gasser [9] sind solche zugespitzten Szenarien kritisch zu bewerten, da der allgemeinen Straßenverkehr zu dynamisch sei, als dass es zwangsweise zu einem tödlichen Personenschaden kommen muss. Vielmehr könnten AA durch eine defensive Fahrweise solche Situationen rechtzeitig erkennen und ggf. vermeiden. Sollten Dilemma, entgegen seiner Meinung, von größerer Relevanz sein, müsse man ethische Entscheidungskriterien finden, die allgemein anerkannt seien. (vgl. [9], S. 555 ff.). Folgendes sollte jedoch nicht vergessen werden: „Auch kommt es in ‚Dilemma-Situationen‘ unterschiedslos im Fall eines menschlichen Fahrers wie im Fall maschinellen Wirkens letztlich zu einer Schädigung.“ ([9], S. 558).

Bezogen auf das letzte Zitat bestünde die Option, den Insassen eines AA schlicht als „Notfall-Fahrer“ in Dilemma-Situationen die Verantwortung zu übergeben (vgl. [6]). Dilemma lassen sich in diesem Modell offensichtlich nicht lösen, jedoch entfällt die ethische Verantwortung der Entscheidungsfindung. Der menschliche Fahrer wird den drohenden Unfall nicht verhindern können, da die Übergabe des Fahrauftrags Zeit beansprucht und der Unfall sich in dieser Zeit bereits ereignet haben wird. Fordert man vom Fahrer wiederum, stets konzentriert den Verkehr zu

beobachten, führt dies den Zweck eines AA ad absurdum. Wolf [10] formuliert wie folgt: „einerseits werden Systemfunktionen aufgrund der Fehlerhaftigkeit des Menschen automatisiert, und andererseits soll genau dieser Mensch das System überwachen und im Notfall als Rückfalloption zur Verfügung stehen“ ([10], S. 105). Hevelke und Nida-Rümelin [7] prognostizieren zudem, dass sich mit AA im Durchschnitt nur ca. alle 2 Millionen Kilometer ein Unfall ereignen wird (vgl. [7], S. 624 f.).

Abschließend sei noch ein Lösungsvorschlag von Scholz und Kempf [11] vorgestellt. Dilemma sollen durch vier Lösungswege vermieden werden: eine graduelle Einführung von AA in bestimmten Szenarien, Koordination (zwischen AA, anderen Verkehrsteilnehmern und der Infrastruktur), eine defensive Fahrweise und ein nicht-deterministisches Entscheidungssystem für Dilemma-Situationen. Die Ansätze zielen darauf ab, die gesellschaftliche Akzeptanz für AA nach ihrer Einführung zu festigen, da sie zunächst nur in bestimmten, tendenziell vorhersehbaren Szenarien (z.B. Autobahn) eingesetzt würden und sich etablieren können. Eine vernetzte Infrastruktur in Kombination mit einer zurückhaltenden Fahrweise kann im Idealfall auch in urbanen Gebieten Konflikte vermeiden (vgl. [11], S. 223 ff.). Aus ethischer Sicht ist der Vorschlag, ein nicht-deterministisches Entscheidungssystem zu implementieren, ambivalent. Verhielte sich der AA „wie ein perfekter Autofahrer“ ([11], S. 226), wäre die Diskussion über ethische AA unbegründet, da an menschliche Fahrer keine besondere ethischen Forderungen gestellt worden sind. Andererseits könnten Todesfälle im Straßenverkehr auftreten, die mit einem utilitaristischen AA hätten verhindert werden können.

7 Fazit

Es konnte keine Übereinstimmung gefunden werden, welche Ethik die „richtige“ sei. Welche Instanz darüber zu entscheiden hat ergab zum Teil sogar widersprüchliche Ansichten [5], [7]. Man könne zudem von „den Programmierern dieser Fahrzeuge... nicht das Auflösen moralischer Dilemmata [erwarten]... wenn selbst in der Moralphilosophie keine Einigkeit herrscht“ ([11], S. 222). Generell scheint der utilitaristische Ansatz eher Zustimmung zu finden als andere, wobei jedoch die fehlende Achtung vor dem individuellen Leben scharf kritisiert wird. Das Dilemma kann mit technischen Mitteln umgangen, aber im Eintrittsfall nicht mehr gelöst werden. Es liegt in der Natur eines Dilemmas, dass es keine zufriedenstellende Lösung geben kann.

Trotzdem besteht Einigkeit darüber, dass die Einführung von AA nicht wegen seltener Extremfälle verhindert werden sollte. Transparenz über die Mechanismen von AA und eine öffentliche Diskussion über deren Ethik können helfen, den Entscheidungsvorgang eines AA zu verstehen und nicht als „Black-Box“ aufzufassen.

Literatur

- [1] Bundesministerium für Verkehr und digitale Infrastruktur, *Strategie automatisiertes und vernetztes Fahren: Leitanbieter bleiben, Leitmarkt werden, Regelbetrieb einleiten*, Broschüre, Berlin, 2015. Adresse: http://www.bmvi.de/SharedDocs/DE/Publikationen/StB/broschuere-strategie-automatisiertes-vernetztes-fahren.pdf?__blob=publicationFile.
- [2] S. Baase, *A gift of fire : Social, legal, and ethical issues for computing technology*, 4. ed., internat. ed. Boston, Mass. [u.a.]: Boston, Mass. [u.a.] : Pearson, 2013.
- [3] S. Nürnberger und S. Bugiel, »Autonome Systeme«, *Datenschutz und Datensicherheit - DuD*, Bd. 40, Nr. 8, S. 503–506, 2016, ISSN: 1862-2607. DOI: 10.1007/s11623-016-0646-2. Adresse: <http://dx.doi.org/10.1007/s11623-016-0646-2>.
- [4] J. C. Gerdes und S. M. Thornton, »Implementable ethics for autonomous vehicles«, in *Autonomes Fahren: Technische, rechtliche und gesellschaftliche Aspekte*, M. Maurer, J. C. Gerdes, B. Lenz und H. Winner, Hrsg. Berlin, Heidelberg: Springer Berlin Heidelberg, 2015, S. 87–102, ISBN: 978-3-662-45854-9. DOI: 10.1007/978-3-662-45854-9_5. Adresse: http://dx.doi.org/10.1007/978-3-662-45854-9_5.
- [5] J. Gogoll und J. F. Müller, »Autonomous cars: In favor of a mandatory ethics setting«, *Science and Engineering Ethics*, S. 1–20, 2016, ISSN: 1471-5546. DOI: 10.1007/s11948-016-9806-x. Adresse: <http://dx.doi.org/10.1007/s11948-016-9806-x>.
- [6] P. Lin, »Why ethics matters for autonomous cars«, in *Autonomes Fahren: Technische, rechtliche und gesellschaftliche Aspekte*, M. Maurer, J. C. Gerdes, B. Lenz und H. Winner, Hrsg. Berlin, Heidelberg: Springer Berlin Heidelberg, 2015, S. 69–85, ISBN: 978-3-662-45854-9. DOI: 10.1007/978-3-662-45854-9_4. Adresse: http://dx.doi.org/10.1007/978-3-662-45854-9_4.
- [7] A. Hevelke und J. Nida-Rümelin, »Responsibility for crashes of autonomous vehicles: An ethical analysis«, *Science and Engineering Ethics*, Bd. 21, Nr. 3, S. 619–630, 2015, ISSN: 1471-5546. DOI: 10.1007/s11948-014-9565-5. Adresse: <http://dx.doi.org/10.1007/s11948-014-9565-5>.
- [8] J.-F. Bonnefon, A. Shariff und I. Rahwan, »The social dilemma of autonomous vehicles«, *Science*, Bd. 352, Nr. 6293, S. 1573–1576, 2016, ISSN: 0036-8075. DOI: 10.1126/science.aaf2654. eprint: <http://science.sciencemag.org/content/352/6293/1573.full.pdf>. Adresse: <http://science.sciencemag.org/content/352/6293/1573>.

- [9] T. M. Gasser, »Grundlegende und spezielle Rechtsfragen für autonome Fahrzeuge«, in *Autonomes Fahren: Technische, rechtliche und gesellschaftliche Aspekte*, M. Maurer, J. C. Gerdes, B. Lenz und H. Winner, Hrsg. Berlin, Heidelberg: Springer Berlin Heidelberg, 2015, S. 543–574, ISBN: 978-3-662-45854-9. DOI: 10.1007/978-3-662-45854-9_25. Adresse: http://dx.doi.org/10.1007/978-3-662-45854-9_25.
- [10] I. Wolf, »Wechselwirkung Mensch und autonomer Agent«, in *Autonomes Fahren: Technische, rechtliche und gesellschaftliche Aspekte*, M. Maurer, J. C. Gerdes, B. Lenz und H. Winner, Hrsg. Berlin, Heidelberg: Springer Berlin Heidelberg, 2015, S. 103–125, ISBN: 978-3-662-45854-9. DOI: 10.1007/978-3-662-45854-9_6. Adresse: http://dx.doi.org/10.1007/978-3-662-45854-9_6.
- [11] V. Scholz und M. Kempf, »Autonomes Fahren: Autos im moralischen Dilemma?«, in *Nationale und internationale Trends in der Mobilität: Technische und betriebswirtschaftliche Aspekte*, H. Proff und T. M. Fojcik, Hrsg. Wiesbaden: Springer Fachmedien Wiesbaden, 2016, S. 217–230, ISBN: 978-3-658-14563-7. DOI: 10.1007/978-3-658-14563-7_14. Adresse: http://dx.doi.org/10.1007/978-3-658-14563-7_14.