



Vanessa Schöffner

Unfallalgorithmen

Eine risikoethische Auseinandersetzung
mit moralischen Dilemma-Strukturen im
Kontext des autonomen Fahrens

VERLAG KARL ALBER



Ethics, Law and AI

Herausgegeben von

Carmine Di Martino (Università degli Studi di Milano)

Federico L.G. Faroldi (Università di Pavia)

Roberto Redaelli (Università degli Studi di Milano)

Band 2

Vanessa Schöffner

Unfallalgorithmen

Eine risikoethische Auseinandersetzung
mit moralischen Dilemma-Strukturen im
Kontext des autonomen Fahrens

VERLAG KARL ALBER



Die Forschungsarbeit wurde durch ein Promotionsstipendium im Rahmen des interdisziplinären, kooperativen Promotionskollegs „Ethik, Kultur und Bildung für das 21. Jahrhundert“ von der Hanns-Seidel-Stiftung gefördert.

Sie entstand weiterhin im Verbundpromotionskolleg „Mobilität & Verkehr“ des Bayerischen Wissenschaftsforums (BayWISS) und wurde vom Bayerischen Staatsministerium für Wissenschaft und Kultur gefördert.

Die Publikation als Open-Access-Werk wurde ermöglicht mit Unterstützung der Barbara-Wengeler-Stiftung.

© Titelbild: Shutterstock, 1162422088

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Zugl.: München, Hochschule für Philosophie, Diss., 2024

u.d.T.: Unfallalgorithmen in risikoethischer Perspektive. Zur Weiterentwicklung des Diskurses moralischer Dilemma-Strukturen im Kontext des autonomen Fahrens

1. Auflage 2024

© Vanessa Schäffner

Publiziert von
Verlag Karl Alber – ein Verlag in der
Nomos Verlagsgesellschaft mbH & Co. KG
Waldseestraße 3–5 | 76530 Baden-Baden
www.verlag-alber.de

Gesamtherstellung:
Nomos Verlagsgesellschaft mbH & Co. KG
Waldseestraße 3–5 | 76530 Baden-Baden

ISBN (Print): 978-3-495-99203-6

ISBN (ePDF): 978-3-495-99204-3

DOI: <https://doi.org/10.5771/9783495992043>



Onlineversion
Nomos eLibrary



Dieses Werk ist lizenziert unter einer Creative Commons Namensnennung 4.0 International Lizenz.

Für meine Familie.

Vorwort

Dieses Buch stellt die gekürzte, redigierte und hinsichtlich ihrer praktischen Kontexteinbettung aktualisierte Fassung meiner Dissertation dar, die im Sommersemester 2024 von der Hochschule für Philosophie München, Philosophische Fakultät S.J. als Promotionschrift angenommen wurde. Als interdisziplinäre Untersuchung aus dem Bereich der praktischen Philosophie ist die Forschungsarbeit thematisch im Grenzbereich von Technologie, Ökonomie, Gesellschaft und Ethik zu verorten – einer Kombination derjenigen Forschungsbereiche, die meinen wissenschaftlichen Werdegang nachhaltig geprägt haben. Die scheinbare Unaufhaltsamkeit technologischer Innovation steht in einem Spannungsverhältnis zur gesellschaftlichen Transformation des Zusammenlebens und dem, was in zukünftigen Gesellschaften als moralisch wünschenswert gelten kann. Dabei bestimmt das autonome Fahren seit einigen Jahren die inter- und transdisziplinären Diskurse in Forschung, Politik und Wirtschaft wie kaum eine andere derzeit in Entwicklung befindliche disruptive Technologie. Die ethische Auseinandersetzung mit den Implikationen und Herausforderungen einer der dynamischsten technologischen Revolutionen der Gegenwart scheint mir ein lohnendes Forschungsziel, das der nachstehenden philosophischen Abhandlung zugrunde liegt.

Der Weg zur Entstehung dieses Buches war durch persönliche und familiäre Ereignisse, vor allem aber durch viele verschiedene Menschen geprägt, die meine Arbeit begleitet haben. An dieser Stelle möchte ich meinen Dank jenen Personen aussprechen, ohne deren Unterstützung diese Forschungsarbeit nicht möglich gewesen wäre. Mein besonderer Dank gilt zuerst meinem Doktorvater Prof. Dr. Alexander Filipović – nicht nur für die fachliche Begleitung meiner Arbeit, zahllose Impulse und kompetente Ratschläge, sondern vor allem auch für seine Geduld, sein Verständnis und die immer unterstützende Form der Förderung und Betreuung. Während der intensiven Arbeit an diesem Buch waren mir die vielen Gespräche

mit ihm stets Ermutigung und Motivation zugleich. Für das mir entgegengebrachte Vertrauen und den Glauben an mich und mein Projekt über die gesamte Entstehungszeit hinweg bin ich sehr dankbar. Ferner danke ich Prof. Dr. Markus Babo in seiner Funktion als Zweitgutachter meiner Dissertation für seine Hilfsbereitschaft und beratende Unterstützung. Ebenfalls bedanken möchte ich mich bei Prof. Dr. Claus Dierksmeier, der nicht nur das Forschungsthema dieses Buches inspirierte, sondern mich auch seit Beginn meines Philosophiestudiums unterstützt und ermutigt hat, meinen Weg in der Philosophie zu finden und weiterzugehen.

Der wesentliche Teil der Abhandlung entstand zwischen Oktober 2018 und Dezember 2022 im Rahmen des interdisziplinären, kooperativen Promotionskollegs »Ethik, Kultur und Bildung für das 21. Jahrhundert«, das von der Kooperationspartnerschaft Katholischer Hochschulen in Bayern getragen wurde. Der fachliche und persönliche Austausch, der während dieses Zeitraums – und darüber hinaus – auf verschiedenen Ebenen mit den Kollegiatinnen und Kollegiaten, der Kollegleitung und dem professoralen Leitungsgremium stattfand, hat den Fortschritt meiner Arbeit sehr bereichert. Mein Dank gilt der Hanns-Seidel-Stiftung für die finanzielle und ideelle Förderung durch ein Promotionsstipendium, ohne das mir die Durchführung dieses Forschungsprojekts nicht möglich gewesen wäre. Ebenfalls bedanken möchte ich mich beim Bayerischen Wissenschaftsforum (BayWISS) für die mehrjährige Förderung meines Projekts im Rahmen des Verbundpromotionskollegs »Mobilität & Verkehr«, insbesondere für die Teilfinanzierung der Printausgabe dieses Buches. Zudem danke ich der Barbara-Wengeler-Stiftung für die zur Verfügung gestellten großzügigen Fördermittel, dank derer die Forschungsarbeit in ihrer elektronischen Form als frei zugängliches Open-Access-Werk erscheinen kann. Dem Verlag Karl Alber und der Nomos Verlagsgesellschaft danke ich für die Aufnahme in die Schriftenreihe und die professionelle, unkomplizierte und wertschätzende Betreuung während des Publikationsprozesses.

Einen besonderen persönlichen Dank widme ich nicht zuletzt meiner Familie, die mir während der intensiven Zeit der Arbeit an diesem Buch auf vielfältige Weise zur Seite gestanden hat. So danke ich zunächst meinen Eltern, Brigitte und Raimund, für ihre Wertschätzung und Unterstützung meines akademischen Weges, ihre praktische Hilfe in den zahlreichen Stunden, die sie dem Korrektorat

meines Manuskripts gewidmet haben – und dafür, dass sie immer an mich geglaubt haben. Tief verbunden und dankbar bin ich meinem Ehemann Benjamin für seine Rücksichtnahme und fortwährende Unterstützung in all den Jahren sowie für seinen unerschütterlichen Optimismus, mit dem er mich an den Tiefpunkten zum Weitermachen ermutigte und mir Hoffnung schenkte. Von ganzem Herzen danke ich schließlich meinen beiden Töchtern Sophia und Lea, die mir vor allem in schwierigen Phasen der intensiven Forschungsarbeit stets Quelle von Kraft, Zuversicht und Freude waren. Ihnen sei dieses Buch gewidmet.

In Bezug auf die im Rahmen der nachfolgenden Untersuchung verwendeten zentralen Begrifflichkeiten und verfolgten Ziele sind vorab einige klärende Anmerkungen hilfreich, um das Verständnis der dargestellten Sachverhalte, Thesen und Argumente zu erleichtern. Im Verlauf dieser Forschungsarbeit werden Unfallszenarien als unlösbare Dilemmata charakterisiert. Diese sind, wie der Name schon sagt, ›unlösbar‹ in dem Sinne, dass keine triviale, eindeutige Lösung für ihre spezifische Problematik existiert; unlösbare Dilemmata werden nicht gelöst, sie werden *entschieden*. Entsprechend ist das erklärte Ziel des Diskurses moralischer Unfalldilemmata nicht die Entwicklung von Lösungs-, sondern von *Entscheidungsstrategien*. Mit ›Entscheidung‹ bzw. ›entscheiden‹ ist dabei das Resultat bzw. der Prozess einer ethischen Reflexion gemeint, die es erlaubt, unter Berücksichtigung aller moralisch relevanten Aspekte die im jeweiligen Einzelfall bestmögliche Antwort zu identifizieren und zu begründen, ohne den für die zurückgewiesene Alternative sprechenden Gründen ihre Geltung abzuerkennen.

Aufgrund der Tatsache, dass autonome Fahrzeuge prinzipiell durch Softwarealgorithmen gesteuert werden, entsteht im Hinblick auf den Entscheidungsbegriff in diesem Kontext ein zusätzlicher Klärungsbedarf, wenn angenommen wird, dass Maschinen nicht in einer dem Menschen ebenbürtigen Weise moralisch handlungsfähig sind. Es sei darauf hingewiesen, dass, sofern in der vorliegenden Arbeit von ›Handlungen‹, ›Entscheidungen‹ oder semantisch ähnlichen Ausdrücken die Rede ist, die im Kontext von autonomen Systemen verwendet werden, diese in einem metaphorischen Sinne zu verstehen sind. Sie beziehen sich nicht auf den philosophischen Handlungsbegriff und schließen explizit die Annahme aus, dass sie das Ergebnis eines kognitiven Prozesses sind, der menschlichem

Handeln und Entscheiden ebenbürtig ist. Die Verwendung des Begriffs ›Entscheidung‹ erfolgt vielmehr analog zu Miller et al. (2017, S. 390): »[...] we will use ›decision-making‹ to describe the following situation: an entity is in a situation, receives information about that situation, and selects and then implements a course of action.« Von ›Handlungen‹ wird im Sinne von maschinell gesteuerten Bewegungsabläufen gesprochen:

[...] the term ›action‹ is elliptical for something more technical: a robot's action is any movement that the robot causes that is not immediately caused by a human programmer or controller. [...] When an autonomous car in ›autopilot mode‹ steers the wheel to stay in its lane or avoid a collision, this is the action of a robot. (Talbot et al., 2017, S. 259–260)

Die Arbeit demonstriert, dass Unfalldilemmata sich als unlösbare Konflikte zwischen legitimen individuellen Interessen darstellen, die Grundrechte der Einzelnen berühren. Die Auseinandersetzung erfolgt hier ausdrücklich auf ethischer Ebene; rechtbasierte Perspektiven finden nur stellenweise Erwähnung, um Argumente zu veranschaulichen oder zu vervollständigen. Der entwickelte alternative Problemzugang versteht sich als dezidiert ethischer Entwurf, der bestrebt ist, Impulse für eine politische Regulierung zu liefern, ohne sich selbst politischer Komponenten zu bedienen.

Darüber hinaus sind an dieser Stelle noch einige formale Hinweise vor auszuschicken, welche die Standards wissenschaftlichen Arbeitens erfordern. So sei erstens erwähnt, dass im Verlauf dieses Buches teilweise spätere Ausgaben zitierter, vor allem philosophischer Literatur verwendet werden. Gemäß des gewählten Zitationsstils geben die Kurzbelege im laufenden Text das Jahr der jeweiligen Erstveröffentlichung unabhängig von der verwendeten Ausgabe an, um eine Einordnung der zitierten Quellen in den jeweiligen historischen Kontext zu ermöglichen. Seitenangaben bei direkten Zitaten beziehen sich hingegen auf die jeweils verwendeten Ausgaben. Im Literaturverzeichnis werden sowohl das Erscheinungsjahr der verwendeten Ausgabe als auch dasjenige der Erstveröffentlichung ergänzend ausgewiesen. Eine Ausnahme bilden die Werke von Immanuel Kant; hier wird das Jahr der ursprünglichen Veröffentlichung jeweils separat per Fußnote vermerkt.

Zweitens wird explizit darauf hingewiesen, dass Teilergebnisse der nachfolgenden philosophischen Untersuchung sowohl in sinn-

gemäß der auch wortgetreuer Form bereits in wissenschaftlichen Sammelbänden und thematisch einschlägigen Fachzeitschriften publiziert worden sind. Die Langbelege der entsprechenden Publikationen sind dem Literaturverzeichnis zu entnehmen.

Drittens ist zu beachten, dass dieses Buch auf die Verwendung von Genderstilen mit Sonderzeichen oder Doppelnennungen zugunsten einer besseren Lesbarkeit verzichtet. Wo immer es möglich ist, werden genderneutrale Formulierungen gewählt. Die verwendeten Personenbezeichnungen sind geschlechtsunspezifisch zu verstehen und beziehen sich – sofern nicht gesondert gekennzeichnet – auf alle Geschlechter (m/w/d/x).

Vanessa Schöffner

Ulm, im Dezember 2024

Inhaltsverzeichnis

1. Einführung in die Untersuchung	21
1.1 Problemaufriss: Autonomes Fahren als sozio-techni- sches Phänomen mit ethischer Dimension	21
1.2 Erkenntnisinteresse und Relevanz	25
1.3 Methodik und Struktur der Untersuchung	28
1.3.1 Methodischer Ansatz	28
1.3.2 Ziele und Hypothesen	30
1.3.3 Gedankengang	31
 I. Autonomes Fahren und Unfalldilemmata: Ethischer Problemhorizont und Relevanz . . .	 35
 2. Das Phänomen ›Autonomes Fahren‹: Agenda, Ziele und Herausforderungen	 37
2.1 Selbstfahrende Fahrzeuge als Treiber der Mobilitäts- wende	37
2.1.1 Der Autonomiebegriff im Kontext technischer Systeme	37
2.1.2 Motivatoren des autonomen Fahrens	39
2.1.3 Von Informanten über Assistenten zu Automa- ten: Evolution und Stufenmodell der Fahrauto- matisierung	45
2.2 Herausforderungen im Kontext der Entwicklungs- agenda	49
2.2.1 Die Wechselbeziehung zwischen technischer Reife und Wirtschaftlichkeit	49
2.2.2 Wo stehen wir heute? Aktueller technischer Stand und regulative Verordnungen	54
2.2.3 Zwischen Utopie und Dystopie: Die Ambiva- lenz des autonomen Fahrens	63
	13

3. Unfallalgorithmen als ›moralischer Kompass‹ in ausweglosen Fahrsituationen	71
3.1 Ethische Problemstellungen und Diskurse im Überblick	71
3.1.1 Ethische Problemfelder im Kontext des autonomen Fahrens	71
3.1.2 Problemfeld Unfallsituationen: Der Verantwortungsdiskurs	75
3.1.3 Praktische Unvermeidbarkeit und dilemmatische Struktur auswegloser Fahrsituationen	81
3.2 Die Relevanz von Dilemma-Szenarien für das autonome Fahren	89
3.2.1 Möglichkeit und Existenz von Unfalldilemmata	89
3.2.2 Sind Unfallalgorithmen normierbar?	98
3.2.3 Gesellschaftliche und technische Relevanz von Dilemma-Szenarien	102
3.3 Zwischenergebnis: Die zentrale Bedeutung von Dilemma-Szenarien	106
 II. Problemzugänge in zwei Diskursen: Darstellung und Kritik	 107
 4. Unfallalgorithmen als moralisches Designproblem: Rekonstruktion und Kritik	 111
4.1 Entscheidungsalgorithmen, Dilemma-Szenarien und vermeintliche Trolley-Analogien	111
4.1.1 Die ethische Dimension von Entscheidungsalgorithmen	111
4.1.2 Maschinelle Moral, kontextsensitive Systeme und maschinelles Lernen	116
4.1.3 Systematisierung repräsentativer Dilemma-Szenarien und ihre moralphilosophische Problematisierung.	126

4.1.4	Dilemma-Szenarien als angewandtes Trolley-Problem? Von Diskrepanzen und Disanalogien	131
4.2	Praktische Kontexteinbettung: Politisch-soziale Dimension und Entscheidungen unter Risiko	136
4.2.1	Die gesellschaftlich-soziale Dimension von Dilemma-Szenarien	136
4.2.2	Politische Regulierung: Unfallalgorithmen im Spannungsfeld zwischen individuellen Präferenzen und pluralistischen Wertvorstellungen	140
4.2.3	Epistemische Diskrepanzen: Sicherheit, Unsicherheit und Risiko im Kontext von Unfallszenarien	151
4.3	Deskriptive Ansätze: Perspektiven aus der Moralphychologie	156
4.3.1	Moralische Präferenzen der Öffentlichkeit im Fokus einer experimentellen Ethik	156
4.3.2	Zur Relevanz deskriptiver Methoden: Eine Kritik	159
4.4	Normative Ansätze: Begründungsversuche der philosophischen Ethik	164
4.4.1	Klassische philosophische Ansätze zur moralischen Relevanz des Intervenierens	166
4.4.2	Utilitaristische Ansätze	170
4.4.3	Deontologische Ansätze	179
4.4.4	Alternative Ansätze und pluralistische Frameworks	186
4.5	Zwischenergebnis: Ungeklärte Fragen des Diskurses	199
5.	Die Komplexität moralischer Dilemma-Strukturen: Rekonstruktion aus metaethischer Sicht	203
5.1	Einführung: Dilemmata als Grenzsituationen moralischen Handelns	203
5.1.1	Beispiele und Narrative aus Philosophie, Literatur und lebenspraktischen Kontexten	203

5.1.2	Kriterien und Definition moralischer Dilemma-Strukturen	207
5.2	Von der (Un-)Möglichkeit und (Nicht-)Existenz moralischer Dilemmata	216
5.2.1	Überblick und Einführung in den Diskurs . . .	216
5.2.2	Phänomenologische und konzeptionelle Per- spektiven	218
5.2.3	(Vermeintliche) Inkonsistenzen in Theoriesys- temen: Argumente der deontischen Logik und Thesen logischer Widersprüchlichkeit	229
5.3	Lösbarkeit, Inkommensurabilität und (Un-)Ver- gleichbarkeit in Wertekonflikten	234
5.3.1	Vorrangbeziehungen und <i>Prima-Facie</i> -Pflich- ten	234
5.3.2	Symmetrie versus Inkommensurabilität: Krite- rien und Konzeptionen	239
5.3.3	Metaethische Konzepte unvermeidbaren Scheiterns: Von unersetzbaren Verlusten und nicht-verhandelbaren moralischen Werten . .	243
5.4	Anwendungsfall Unfalldilemmata: Interpretation aus metaethischer Sicht	247
5.4.1	Dilemmatische Unfallsituationen als Konflikte inkommensurabler Werte	247
5.4.2	Entscheidungsperspektiven für inkommen- surable Wertekonflikte	250
5.4.3	Zwischenergebnis: Argumentative Relevanz der metaethischen Analyse	259

III. Risikoethische Auseinandersetzung: Entwurf eines alternativen Problemzu- gangs	261
6. Theoretische Grundlagen, begriffliche Reflexion und Ziele einer Risikoethik für Unfalldilemmata . . .	265
6.1 Systematische wissenschaftliche Einordnung der Risikoethik	265
6.1.1 Sozialwissenschaftlicher und sozio-technischer Diskurs	265
6.1.2 Von der Technikanalyse zur Technikbewertung: Technikfolgenabschätzung und technischer Diskurs	268
6.2 Risikoethische Grundlagen und Begriffe	274
6.2.1 Risikoethische Grundbegriffe: Unsicherheit, Ungewissheit und Risiko	274
6.2.2 Risiken im Handlungskontext: Risikosituationen und Risikokonstellationen	278
6.2.3 Grundfragen der Risikoethik: Zulässigkeit, Fairness und Verantwortung im Kontext von Risikoübertragungen	284
6.3 Grundzüge der (rationalen) Risikopraxis: Paradigmen und entscheidungstheoretische Ansätze	287
6.3.1 Risikopraktische Paradigmen	287
6.3.2 Entscheidungstheoretische Kriterien rationaler Risikopraxis	291
6.3.3 Zur Kritik traditioneller Risikopraxis	296
7. Unfallalgorithmen als risikoethisches Verteilungsproblem	299
7.1 Die (risiko-)ethische Problematisierung von Mobilitätsrisiken im Kontext autonomer Fahrsysteme	301
7.1.1 Autonome Fahrzeuge im Spannungsfeld zwischen soziologischer Risikoakzeptanz und ethischer Risikoakzeptabilität	301
7.1.2 Unfallalgorithmen und Risikoethik: Ansätze bisheriger (risikoethischer) Forschung	306

7.1.3	Gegenstand und Ziele eines alternativen risikoethischen Entwurfs	312
7.2	Analyse der Risikokonstellationen in Dilemma-Szenarien entlang von Kriterien der Risikoakzeptabilität	315
7.2.1	Akteure, Beziehungsnetzwerke und private Risiken	315
7.2.2	Szenarien der Risikoübertragung	319
7.2.3	Diskussion aus Sicht konsequentialistischer und kontraktualistischer Kriterien	323
7.2.4	Deontologische Risikoethik: Begründung, Ansätze und Konzeptionen	333
7.3	Grundzüge einer deontologischen Risikoethik für Unfallalgorithmen	336
7.3.1	Kohärente Risikopraxis nach Julian Nida-Rümelin: Grundlinien, Ziele und Anwendung.	336
7.3.2	Die (absolute) Frage der Zumutbarkeit: Eine moralische Gratwanderung entlang von Risikoschwellen	341
7.3.3	Die (relative) Frage der Gerechtigkeit: Zwischen Reziprozität und Vorteilsausgleich	351
8.	Fazit und Ausblick	375
8.1	Ergebnisse der philosophischen Untersuchung: Zusammenfassung	375
8.2	Kritische Reflexion und Ausblick: Wissenschaftliche Relevanz, Forschungsdesiderate und Limitationen	384
	Literaturverzeichnis	387

Facere docet philosophia, non dicere, et hoc exigit, ut ad legem suam quisque vivat, ne orationi vita dissentiat vel ipsa inter se vita; ut unus sit omnium actio[dissentio] num color [sit].

– **Seneca**, *Epistulae morales ad Lucilium*, Liber II, Epistula 20, 2

Was kann als Kompaß dienen? Die vorausgedachte Gefahr selber! In ihrem Wetterleuchten aus der Zukunft, im Vorschein ihres planetarischen Umfangs und ihres humanen Tiefganges, werden allererst die ethischen Prinzipien entdeckbar, aus denen sich die neuen Pflichten neuer Macht herleiten lassen.

– **Hans Jonas**, Vorwort zu *Das Prinzip Verantwortung* (1979)

