

## 7. Unfallalgorithmen als risikoethisches Verteilungsproblem

Bevor in diesem siebten Kapitel schließlich zur risikoethischen Analyse übergegangen wird, ist an dieser Stelle zunächst eine klärende Vorbemerkung zu einer zentralen Begrifflichkeit notwendig. In den bisher einschlägigen risikoethischen Forschungsbeiträgen ist im Hinblick auf das zugrundeliegende Entscheidungsproblem beinahe ausschließlich von einer »Risikoverteilung« (*risk distribution*) die Rede (vgl. z. B. Berkey, 2022; Dietrich, 2021; Dietrich & Weisswange, 2019; Geisslinger et al., 2021; Smith, 2022). »The ethics of automated vehicles may be better framed as the fair distribution of risks, both during and before forced-choice situations«, stellt auch Goodall (2017, S. 496) fest. Aus strikt definitorischer Sicht ist die Verwendung dieses Begriffs jedoch nicht korrekt. Der Terminus der ›Distribution‹ bzw. ›Verteilung‹ ist ökonomischen Ursprungs und wird insbesondere im Kontext von Konzepten der distributiven Gerechtigkeit, z. B. bei Rawls, mit einer Verteilung von Gütern bzw. Ressourcen assoziiert. Mit ihm eng verwandt und häufig verwechselt ist der inhaltlich weiter gefasste Begriff der ›Allokation‹, der sich ebenfalls mit distributiven Fragen auseinandersetzt. In den Wirtschaftswissenschaften meint ›Allokation‹ die Verteilung bzw. Zuordnung knapper Ressourcen bzw. Produktionsfaktoren mit dem Ziel, diese denjenigen Produktionsmöglichkeiten einer Volkswirtschaft im Sinne von Gütern und Dienstleistungen zuzuführen, welche den bestmöglichen Einsatz der Produktionsfaktoren verwirklichen (vgl. Bundeszentrale für politische Bildung, 2016). ›Allokation‹ bezieht sich also auf *Inputfaktoren*, mittels derer *Outputs* erzeugt werden; die Verteilung Letzterer auf Wirtschaftssubjekte wird als ›Distribution‹ bezeichnet.

Diese begriffliche Differenzierung hat bedeutende Implikationen für ein korrektes Verständnis der Entscheidungsproblematik, die der Programmierung von Unfallalgorithmen zugrunde liegt. Um was geht es eigentlich genau? Der Versuch, entstehende Schäden anhand

ethischer Kriterien zu verteilen, ist schon deshalb unplausibel, weil die Entscheidungssituation sich gerade dadurch auszeichnet, dass die sich tatsächlich manifestierenden Folgen kontingent sind. Es geht vielmehr um die Inputfaktoren, die Schäden hervorbringen – um etwas, aus dem potenziell eine negative Konsequenz entstehen kann, aber nicht zwangsläufig muss: um Risiken. Streng genommen wäre es also angemessener, von einem Allokationsproblem von Risiken zu sprechen:

The distribution of the risks of operating autonomous vehicles [...] may fall short of the targeting paradigm specific individuals, yet the centralised and coordinated nature of algorithmic risk transforms this process to one of risk allocation rather than mere risk distribution. This is because allocation infers direction, edging towards intentionality, and emphasises the perspective of the patient-victim in the accident relationship by connoting the imposition of risk. (Liu, 2017, S. 201)

Der Begriff der ›Verteilung‹ ist hingegen dann adäquat, wenn es um die Outputs geht, die durch Risiken generiert werden. Diese lassen sich allgemein als die Vor- und Nachteile beschreiben, die aus einer Risikokonstellation resultieren und anhand von Fairnesskriterien verteilt werden sollten (siehe Kap. 7.3.3): »Distributive justice, [...] is the fairness between those who directly and indirectly benefit and those who directly and indirectly carry the burdens«. (Goér Herve et al., 2023, S. 4) Tatsächlich werden Fragen einer fairen Risikoverteilung im Kontext von Unfallalgorithmen zumeist anhand einer Diskussion von Nutzen und Lasten im Zuge von Risikoübertragungen erörtert. Das akzentuierte Problem ist damit im Kern durchaus ein Verteilungsproblem, wenn auch nur indirekt eines von Risiken. Aus diesem Grund wird in der in diesem Buch vorgelegten risikoethischen Abhandlung von einem *risikoethischen Verteilungsproblem* bzw. einer *Verteilung von Vor- und Nachteilen*, nicht aber von einer Risikoverteilung gesprochen. Wann immer in direkter oder indirekter Weise auf Literatur verwiesen wird, die den Begriff der ›Risikoverteilung‹ verwendet, sei angemerkt, dass zitierte Inhalte sinngemäß vor dem Hintergrund der erläuterten Begriffsreflexion zu verstehen sind.

## 7.1 Die (risiko-)ethische Problematisierung von Mobilitätsrisiken im Kontext autonomer Fahrsysteme

### 7.1.1 Autonome Fahrzeuge im Spannungsfeld zwischen soziologischer Risikoakzeptanz und ethischer Risikoakzeptabilität

Obwohl Mobilitätsrisiken gegenwärtig vor allem im Zusammenhang mit neuen Technologien diskutiert werden, sind sie kein Spezifikum der Verkehrsautomatisierung, sondern vielmehr seit jeher Bestandteil des Systems Mobilität. Jedoch nimmt ihre Komplexität seit den Anfangstagen des motorisierten Verkehrs und damit verbundenen höheren Fahrgeschwindigkeiten stetig zu. Am Verkehrsgeschehen beteiligte Personen sind zum einen jenen Risiken ausgesetzt, die durch andere Verkehrsteilnehmer herbeigeführt werden, und zum anderen verursachen sie selbst Risiken, die wiederum andere betreffen. Dabei sind Risiken und Gefahren im Kontext von Mobilitätssystemen, wie auch in vielen anderen Bereichen unseres alltäglichen Lebens, weitgehend gesellschaftlich akzeptiert. Sie werden aufgrund ihres bedeutenden gesellschaftlichen Nutzens oftmals bedenkenlos in Kauf genommen. Das lässt sich einerseits durch psychologische Effekte erklären, denn aus psychologischer Sicht klaffen Risikorealität und Risikowahrnehmung oft auseinander. Sichtbar wird dies zum einen hinsichtlich der subjektiven Wahrnehmung von Eintrittswahrscheinlichkeiten. Bei sehr seltenen Ereignissen bzw. sinkenden Wahrscheinlichkeiten rückt das Wissen um die kausale Verknüpfung zwischen einer Handlung und deren Folgen tendenziell in den Hintergrund (vgl. Hansson, 2003, S. 292), während unmittelbare Vorteile der Mobilität in den Vordergrund treten. Zum anderen werden neuartige Risiken tendenziell eher überschätzt, etablierte hingegen unterschätzt. Risiken, die man glaubt steuern zu können, werden als weniger bedrohlich empfunden; stehen Risiken mehr in der öffentlichen Diskussion, verändert dies das subjektive Risikobewusstsein ebenfalls (vgl. Nida-Rümelin, 2005, S. 868–871; Rath, 2011, S. 120).

Andererseits wird auf gesamtgesellschaftlicher Ebene zwischen Kosten bzw. Risiken und Nutzen eines Mobilitätssystems abgewogen, wobei Sicherheit (*safety*) und Komfort (*convenience*) die zentralen Aspekte darstellen. Ist der gesellschaftliche Nutzen höher als die Kosten, werden Risikoübertragungen innerhalb des betreffenden Systems als zulässig bewertet und ein gewisses Risiko ist grund-

sätzlich zu tolerieren (vgl. Gasser, 2015, S. 548). Mobilität als gesellschaftliches System kann damit auf quasi-kontraktualistische Weise als im Kern gerechtfertigt erachtet werden (vgl. Müller & Gogoll, 2020, S. 1554). Dies muss umso mehr für automatisierte Fahrsysteme gelten, von deren Einführung man sich gerade eine signifikante Erhöhung der Verkehrssicherheit erhofft.

Derartige Argumentationsstrukturen sind zwar logisch valide und damit gesellschaftlich akzeptiert, aus (risiko-)ethischer Sicht jedoch prinzipiell problematisch. Philosophische Kritik rückt vor allem die Diskrepanz zwischen gesellschaftlichem und individuellem Nutzen in den Vordergrund. Sie richtet sich dabei gegen den konsequentialistischen Kern des Gedankengangs, unter dessen Primat individuelle Interessen im Rahmen einer gesamtgesellschaftlichen Nutzenabwägung dem großen Ganzen untergeordnet werden, wie es auch im Grundkonzept rationaler Risikopraxis verankert ist. Wie bereits die kritische Auseinandersetzung in Kap. 6.3.3 gezeigt hat, ist ein solches Vorgehen nicht haltbar. Verkehr ist ein soziales System, das sich an den Bedürfnissen des Einzelnen ausrichten muss und letztlich nur durch diese legitimiert werden kann.

Dennoch ist es plausibel anzunehmen, dass gesamtgesellschaftlich getragene Risiken des automatisierten Verkehrs nicht notwendigerweise für alle Individuen vorteilhaft sind; vielmehr ist davon auszugehen, dass gewisse Gruppen – i. e. die Halter bzw. Insassen eines selbstfahrenden Fahrzeugs – in den Genuss der Vorteile kommen, während andere vor allem die resultierenden Nachteile tragen. Auch das auf den ersten Blick einschlägige Argument, dass aufgrund des Sicherheitspotenzials autonomer Fahrzeuge alle Mitglieder einer Gesellschaft von einer erhöhten Verkehrssicherheit profitieren würden, kann die Skepsis gegenüber einer ethisch fragwürdigen Verteilung von Vor- und Nachteilen nicht aufheben. Erstens ist zweifelhaft, ob sich das visionäre Sicherheitsversprechen autonomer Fahrzeuge überhaupt erfüllen wird (siehe Kap. 2.2.3). Zweitens schließt eine allgemeine Verbesserung des Sicherheitsniveaus nicht aus, dass Individuen im Einzelfall schlechter gestellt werden können, z. B. in Situationen, die durch die rezeptiven und kognitiven Limitationen technischer Systeme erst verursacht werden (vgl. Brändle & Grunwald, 2019, S. 289).

An dieser Problematik setzt die Risikoethik an, die die Legitimität eines risikobehafteten Systems stets unter Bezugnahme auf Einzelin-

teressen begründet: Ein System, das Risiken produziert, kann nur unter Berücksichtigung distributiver Effekte ethisch gerechtfertigt werden, d. h. wenn die entstandenen Vor- und Nachteile für alle Betroffenen fair verteilt sind. So argumentiert Ferretti (2010, S. 506) von einem institutionalistischen Standpunkt aus, dass eine bestimmte Risikohandlung, die andere betrifft, unter diesen Voraussetzungen für eine gerechte Gesellschaft akzeptabel sein kann. Dies ist beispielsweise dann der Fall, wenn ein »equitable social system of risk-taking« (Hansson, 2003, S. 305) vorliegt, das auf der folgenden zentralen These basiert: Das *Prima-Facie*-Recht, keinen Risiken ausgesetzt zu werden, kann dann überschrieben werden, wenn die Risikoübertragung Teil eines gerechten Systems von Risiken ist, welches dem Einzelnen Vorteile bringt.<sup>243</sup>

Auch wenn die soziologisch begründete Akzeptanz aggregierter Nutzenabwägungen unabhängig von ihrem konkreten Anwendungskontext prinzipiell ethisch fragwürdig ist, wird sie zumeist nicht gesondert ethisch thematisiert, sondern als etabliertes gesellschaftsstrukturelles Phänomen betrachtet.<sup>244</sup> Ein möglicher Grund dafür mag darin liegen, dass die Risiken des herkömmlichen Straßenverkehrs durch die Annahme impliziter Zustimmung der Einzelnen als legitimiert angesehen werden, um das *Problem of Paralysis* – und damit einen drohenden Verlust der durch das Mobilitätssystem generierten Vorteile – zu umgehen. Ungeachtet dessen, dass eine solche Argumentationsbasis generell ethisch fragwürdig ist, muss sie angesichts des autonomen Fahrens erst recht kritisch adressiert werden. Mögliche Zustimmungsvorbehalte sind angesichts neuartiger technikinduzierter Risiken neu zu verhandeln (siehe Kap. 7.2.3.3); insbesondere dann, wenn die Nutzung autonomer Fahrzeuge tatsächlich eines Tages verpflichtend werden sollte.

243 Diese These Hanssons wird im weiteren Verlauf des siebten Kapitels noch intensiv erörtert.

244 Als einflussreich gilt in diesem Kontext das Modell des Ingenieurs Chauncey Starr (1969), der sich als einer der ersten Wissenschaftler mit der gesellschaftlichen Akzeptanz neuer technischer Risiken auseinandergesetzt hat. Ausgehend von einer Analyse der in Gesellschaften historisch nachweisbaren und traditionell akzeptierten Risiken hat er eine Theorie gesellschaftlicher Risikoakzeptanz entwickelt, in deren Rahmen Letztere nicht auf eine moralische Bewertung, sondern auf soziologisch feststellbare Gewohnheiten hinsichtlich der Risikoakzeptanz zurückgeführt wird.

Autonome Fahrsysteme verändern die Risikolandschaft des zukünftigen Straßenverkehrs auf disruptive Weise. Ein zentraler Grund, weshalb Mobilitätsrisiken im Kontext des autonomen Fahrens im Vergleich zu herkömmlichen Fahrzeugen ethisch stärker problematisiert werden, liegt in der Natur der betreffenden Risiken. Die ›neuen‹ Risiken sind technikinduziert, d. h. sie werden durch spezifische Funktionsweisen technischer Systeme in einer Weise transformiert, die ethisch problematisch ist: Diese Risiken sind systemisch und verändern das Wesen risikobehafteter Entscheidungen – aus einer intuitiv-situativen Reaktion in einer spontan auftretenden Verkehrssituation wird eine überlegte, bewusste Entscheidung, die sich in Algorithmen niederschlägt (vgl. Lin, 2015, S. 74; Nyholm & Smids, 2016, S. 1278–1279). Die im Rahmen der Fahrautomatisierung herbeigeführten Risiken sind in hohem Maße abhängig von einem fehlerfreien Betrieb der entsprechenden Systeme, der jedoch angesichts der generellen Fehleranfälligkeit von Maschinen nie zu vollständiger Sicherheit führen kann.

Wie Sahlin und Persson (1994, S. 38–53) betonen, müssen neben Risiken, die sich direkt auf zu erwartende Konsequenzen beziehen, auch epistemische Risiken Bestandteil einer validen Risikobewertung sein, welche die Zuverlässigkeit des vorausgesetzten Wissens hinterfragt. Hinzu kommt schließlich, dass Technik an sich grundsätzlich nicht wertneutral ist, sondern vielmehr geeignet, wesentliche Aspekte des menschlichen Lebens zu konditionieren (vgl. Mladenovic & McPherson, 2016, S. 1132–1133). Die Erkenntnis, dass eine explizite Risikobetrachtung bei KI-Systemen unerlässlich ist, reiste spätestens seit dem Inkrafttreten der Datenschutz-Grundverordnung (DSGVO) im Jahr 2018 und der damit verbundenen Forderung nach der Etablierung von *Privacy-by-Design*-Ansätzen. Der im Herbst 2021 vom Ingenieurverband IEEE herausgegebene Ethikstandard für intelligente und autonome Systeme (*IEEE 7000*) mit dem Titel »IEEE Standard Model Process for Addressing Ethical Concerns during System Design« trägt dem zunehmenden Bewusstsein Rechnung, dass Ethik bereits von Beginn eines Entwicklungsprojekts an mitgedacht werden muss, um spätere negative Auswirkungen zu verhindern. Grundlage dieses Standards ist das von Sarah Spiekermann erarbeitete Prinzip des *Value Based Engineering*, welches das Ziel verfolgt, ethische Werte und Normen in allen Phasen der Systementwicklung zu berücksichtigen (*Ethics by Design*). Der Standard

enthält zehn Handlungsempfehlungen, zu denen auch eine risiko-bewusste und verantwortungsvolle Entwicklung von KI-Systemen zählt (vgl. Institute of Electrical and Electronics Engineers, 2021; Spiekermann & Winkler, 2022).

Vor diesem Hintergrund ist festzuhalten, dass die Aufgabe der ethischen Gestaltung von Unfallalgorithmen verschiedene Problemstellungen mit sich bringt. Ziel des dritten und letzten Teils der Forschungsarbeit ist es, ausgewählte Aspekte aus dem Blickwinkel einer risikoethischen Perspektive zu beleuchten und damit über den Horizont einer rein theoretischen Ebene hinauszugehen, auf welcher der bisherige Forschungsdiskurs in weiten Teilen stattfindet (vgl. Poszler et al., 2023, S. 19). Als Pointierung des inkompatiblen Aufeinandertreffens individueller Interessen werden Dilemma-Szenarien als besonders geeignet erachtet, dieses Ziel zu erreichen. Wenn sich Situationen, in denen es unweigerlich zu Schäden kommt – mindestens in dem Sinne, dass das Recht, keinem Risiko ausgesetzt zu werden, verletzt wird –, nicht völlig ausschließen lassen, dann verbleiben im Hinblick auf die automatisierte Mobilität nur zwei Möglichkeiten: Entweder sind Risikoübertragungen in jeglicher Form und rigoros unzulässig, was ein Verbot der Einführung des autonomen Fahrens an sich zur Folge hätte. Das Postulat eines Nullrisikos würde eine kategorische Zurückweisung jeglicher Risikoübertragungen implizieren und somit eine zu starke und absurde Einschränkung des Raums erlaubter Handlungen darstellen. Dies wiederum stünde dem Zweck der Moral als Ermöglichung guten Handelns zuwider (vgl. Hansson, 2003, S. 297–299; Hayenjelm & Wolff, 2012, e37; Rippe, 2013, S. 534) und erscheint aus pragmatischer Sicht unplausibel. Oder es gibt spezifische Umstände, unter denen dieses Recht überschrieben werden darf und die es zu identifizieren gilt; Hansson beschreibt dies als »exemption problem« (2003, S. 302). Letztgenanntem Auftrag widmen sich die Ausführungen in diesem dritten Teil des Buches im Hinblick auf das Anwendungsproblem. Zunächst wird näher erläutert, inwiefern diese einen originären Beitrag zum Forschungsdiskurs leisten.

### 7.1.2 Unfallalgorithmen und Risikoethik: Ansätze bisheriger (risikoethischer) Forschung

Im Zuge dominanter Forschungszugänge, wie sie in Kap. 4 expliziert wurden, wird die Frage nach der Programmierung von Unfallalgorithmen als moralisches Designproblem konzipiert, das die praktische Gegebenheit unsicherer Handlungsfolgen zum entscheidungsrelevanten Zeitpunkt vernachlässigt. Zwar wird stellenweise ange deutet, dass Unsicherheit in Dilemma-Szenarien eine Rolle spielt;<sup>245</sup> konkrete Forschungsdesiderate werden daraus zunächst jedoch nicht abgeleitet. Erst im Zuge einer kritischen Auseinandersetzung mit den bis dato vorgeschlagenen Ansätzen einerseits sowie einer stärkeren, auch technischen Anwendungsorientierung andererseits rücken neuere Beiträge zunehmend Risikoaspekte in den Vordergrund. Diverse von Bund und Ländern finanzierte Forschungsprojekte haben die Bildung interdisziplinärer Forschungsteams forciert, die anwendungs näher forschen als noch zu Beginn des Diskurses. Auf technischer Seite sind Risikoanalyse und Risikomanagement als zentrale Bestandteile von Forschung und Entwicklung fest verankert. Impulse aus dem ingenieurwissenschaftlichen Diskurs, der Risikoaspekte wie technikgenerierte Unsicherheiten bzw. Unreife der Technik bereits intensiv erforscht (vgl. Dietmayer, 2015; Hubmann et al., 2017), bereichern zunehmend verwandte Debatten in anderen Disziplinen.

So ist auch im ethischen Diskurs der letzten drei bis vier Jahre eine stetig steigende Anzahl an akademischen Beiträgen zu verzeichnen, die die Problemstellung unter Bezugnahme auf Komponenten aus Risikomanagement, Risikoanalyse und Risikoethik diskutieren. Grundlage dieser Entwicklung im deutschen Forschungsdiskurs sind nicht zuletzt auch die Richtlinien der Ethik-Kommission (Di Fabio et al., 2017), die eine Entscheidungsstrategie fordern, welche Risikoaspekte in systematischer Weise berücksichtigt. Der kürzlich in Kraft getretene und daher verstärkt in die öffentliche Wahrnehmung gerückte *AI Act* trägt ebenfalls erheblich dazu bei, dass risikobasierte Konzepte in Zukunft alternativlos sein werden, und daher verstärkt an Relevanz gewinnen. Das neue europäische KI-Gesetz

---

245 Benannt wird dieser Aspekt beispielsweise bei Bhargava und Kim (2017), Birnbacher und Birnbacher (2016), Davnall (2020), Goodall (2016a, 2016b, 2017), Grunwald (2015), Hübner und White (2018), Keeling (2020) sowie Nyholm und Smids (2016).

legt Unternehmen zukünftig die Verpflichtung auf, die mit den von ihnen entwickelten KI-Anwendungen in Zusammenhang stehenden Risiken mithilfe leistungsfähiger Test- und Validierungsverfahren umfassend zu identifizieren, zu analysieren, zu bewerten und entsprechend zu handeln. Die Durchführung von Folgenabschätzungen in Bezug auf die Grundrechte wird fortan von Betreibern von Hochrisiko-KI-Systemen, wie sie auch selbstfahrende Fahrzeuge darstellen, eingefordert (vgl. Future of Life Institute, 2024).

Ein Schwerpunkt der bisherigen Literatur zu Unfallalgorithmen, die sich risikoethischer Konzepte im weiteren Sinne bedient, liegt auf der Betrachtung von Fragen sozialer Gerechtigkeit.<sup>246</sup> Lebens (2017) prominente Anwendung von Rawls' *Differenzprinzip*, welches das Wohlergehen der am schlechtesten gestellten Personen in einer Gesellschaft fördert, bildet einen zentralen Ausgangspunkt kontroverser Diskussionen hinsichtlich der Analyse distributiver Effekte in Dilemma-Szenarien. In ähnlicher Weise verteidigt auch Smith (2022) eine konstruktivistische Rawls'sche Methode bei der Herleitung gerechtigkeitsethischer Prinzipien für Unfallalgorithmen. Im Zuge seiner Kritik an der Verwendung des Trolley-Problems fordert er eine institutionalistische Orientierung hin zu einem »distributiven Paradigma« in der normativen Analyse autonomer Fahrsysteme, in dessen Rahmen Fragen einer gerechten Verteilung ausgehend von gesellschaftlichen Kernstrukturen erörtert werden.

[...] we have considerable reason to adopt institutionalism about our moral obligations in the context of AVs. Rather than try to understand our obligations in the domain of AV by using trolley problems to interrogate how an ideal agent would behave and then trying to replicate that behavior for AV, we should determine the set of principles or rules that govern the fair distribution of benefits and burdens for the deployment of AVs. Our personal obligations, then, would be to attempt to instantiate those principles, but the primary normative principles would apply to institutions and practices. (Ebd., S. 286)

---

246 Dieser Fokus entspricht der gerechtigkeitsethischen Agenda, die unter dem Terminus »Verkehrsgerechtigkeit« wissenschaftlich erforscht wird: »Transport justice describes a normative condition in which no person or group is disadvantaged by a lack of access to the opportunities they need to lead a meaningful and dignified life [...] but also emphasises a more equal distribution of transport benefits and burdens in society [...].« (Martínez-Buelvas et al., 2022, S. 3) Für einführende Beiträge in dieses Forschungsfeld siehe z. B. Karner et al. (2020), Martens (2016) sowie Pereira et al. (2017).

Smith unterstreicht die Bedeutung einer ganzheitlichen Sichtweise, die über die situativ beschränkten Konsequenzen einzelner Dilemma-Szenarien hinausgeht und die Wirkungen auf Strukturen sozialer Gerechtigkeit mitveranschlagt. Mladenovic und McPherson (2016) hingegen gehen im Rahmen ihrer Untersuchung gerechtigkeitsethischer Fragen über den Anwendungsfokus unvermeidbarer Unfallsituationen hinaus, indem sie die Problematik aus der Makroperspektive eines Systems vernetzter Straßenverkehrsüberwachung beurteilen.

Auch wenn diese Forschungsbeiträge nicht explizit risikoethisch ausgerichtet sind, können sie doch in gewissem Sinne als Grundlage für weiterführende risikoethische Forschung dienen. So kombinieren Dietrich und Weisswange (2019) Aspekte beider Ansätze, indem sie sowohl eine theoretische Begründung als auch eine praktische Implementierung eines Planungsmoduls für einen Prototypen vorlegen, welcher der explizit implementierten Regel des *Differenzprinzips* (»Minimiere die maximalen Kosten aller Beteiligten«)<sup>247</sup> folgt. Dabei legen sie den Schwerpunkt nicht auf Dilemma-Szenarien, sondern auf Routinesituationen mit geringen oder moderaten Risiken. In einem weiteren Beitrag entwirft Dietrich (2021) eine Architektur eines automatisierten Fahrzeugbetriebssystems, das ein spezielles Modul für die Risikoabschätzung und das Risikomanagement umfasst, sodass anfängliche Ungleichheiten bei der Risikoexposition ausgeglichen werden. Auch hierbei rekurriert er auf das Rawls'sche *Differenzprinzip* in seiner ursprünglichen Interpretation als institutionelle Zuweisungspolitik. In diesem Sinne plädiert er dafür, das automatisierte Fahren als institutionelle Aktivität zu gestalten (vgl. Dietrich, 2020).<sup>248</sup>

Weitere einschlägige Artikel fokussieren sich auf die Identifizierung und Analyse ethischer Entscheidungsmetriken und die Konzeption kontinuierlicher Entscheidungsprozesse für die Auswahl geeigneter Trajektorien, bei denen risikoethische Aspekte sowohl implizit als auch explizit berücksichtigt werden. Poszler et al. (2023, S. 15–16) geben eine Übersicht über entsprechende Ansätze, zu de-

---

247 Auf Rawls' *Differenzprinzip* wird in Kap. 7.3.3 vertiefend eingegangen.

248 Für eine Perspektive, die Risiko im Kontext (bio-)technologischer Innovationen und vor dem Hintergrund eines nicht-welfaristischen Verständnisses als eine Last der Kooperation auf den Schultern von Regulierungsinstitutionen interpretiert, siehe Ferretti (2010).

nen die *Data Theories Method* (vgl. Robinson et al., 2021), die *Ethical Valence Theory* (vgl. Evans et al., 2020), der *Expected Moral Value Approach* (vgl. Bhargava & Kim, 2017) sowie diverse Konzepte expliziter ethischer Algorithmen zur Trajektorienplanung zählen. Zu Letzteren gehört auch die von Németh (2022) vorgeschlagene Methode zur Routenauswahl, welche quantitative und qualitative Entscheidungsebenen integriert: Zunächst wird auf der Basis einer quantitativen Wahrscheinlichkeitsauswertung die unkritischste Route ermittelt, bevor bestehende Konflikte mittels zusätzlicher qualitativer ethischer Prinzipien entschieden werden. Auf ähnliche Weise operiert auch die Optimierungsfunktion im Rahmen des Steuerungskonzepts von Thornton et al. (2017).

Geisslinger et al. (2021) nehmen eine praxisorientierte Perspektive ein, indem sie (rationale) Entscheidungsprinzipien der Risikoethik auf die Trajektorienplanung autonomer Fahrsysteme anwenden. Dabei benennen sie zunächst interdisziplinäre Anforderungen, die Unfallalgorithmen im Hinblick auf eine mögliche Implementierung erfüllen müssen.<sup>249</sup> Von diesen Kriterien ausgehend, erarbeiten sie ein Framework, das die Spezifikationen der zentralen rationalen Entscheidungsprinzipien (Bayes-Regel, *Maximin*-Prinzip und Vorsorgeprinzip) in mathematische Gleichungen übersetzt und mittels einer komplexen, situativ gewichteten Kostenfunktion integriert. Dieser Ansatz ist aufgrund seiner pluralistischen Struktur flexibel auf eine Vielzahl möglicher Szenarien anwendbar, kann jedoch interne Konflikte und Widersprüche, die zwischen den einzelnen Prinzipien bestehen, nicht aufheben (vgl. Fossa, 2023, S. 76). In einem neueren Ansatz beschäftigen sich Geisslinger et al. (2023a) mit der Auswertung der Risiken möglicher Trajektorien für alle Verkehrsteilnehmer. Anhand einer Kategorisierung in vier Gültigkeitsstufen definieren sie gültige Trajektorien und identifizieren mittels einer ethischen Kostenfunktion diejenigen, die im Hinblick auf verschiedene konse-

---

249 Diese lauten: 1. Repräsentation der Realität, d. h. Berücksichtigung von Kontextfaktoren bei zweckmäßiger Vereinfachung komplexer Sachverhalte, 2. Technische Realisierbar- bzw. Formalisierbarkeit, 3. Allgemeingültigkeit bzw. Universalität für eine Vielzahl denkbarer Szenarien, 4. Soziale Akzeptanz als Voraussetzung für erfolgreiche Nutzeradoption, 5. Transparenz als Absicherung bezüglich rechtlicher Anforderungen (vgl. Geisslinger et al., 2021, S. 1037–1038).

quentialistische und deontologische Kriterien die niedrigsten Kosten verursachen.

Einer der zentralen Kritikpunkte an den bisher vorgelegten risikoethischen Ansätzen betrifft die primäre Folgenorientierung bei der Bewertung risikobehafteter Aktionen des Fahrsystems, infolge derer ethisch relevante Aspekte wie Handlungsmotivationen oder spezifische Pflichten vernachlässigt werden, ebenso wie potenzielle Vorteile aus Risikoübertragungen. Ein möglicher Grund für die Priorisierung konsequentialistischer Kriterien ist, dass bisherige Forschung sich der Problematik aus der Perspektive möglicher Implementierungen annähert. Wie bereits in Kap. 4.4.2 und Kap. 4.4.3 dargelegt, lassen sich insbesondere utilitaristische Kriterien einfacher formalisieren als z. B. deontologische. Es stehen Aspekte einer fairen Verteilung von Risiken einerseits (vgl. Dietrich, 2021; Dietrich & Weisswange, 2019) sowie der Risikobewertung und Risikoabschätzung mittels mathematischer Kostenfunktionen andererseits (vgl. Geisslinger et al., 2021, 2023a; Németh, 2022; Thornton et al., 2017) im Fokus. Indem sie sich ausschließlich auf Kriterien rationaler Risikopraxis stützen, gelingt es den bisherigen Ansätzen jedoch nur begrenzt, zentrale ethische Probleme zu klären.

Unter Bezugnahme auf die fünf Problemfelder der Risikoethik, die in Kap. 6.2.1 beschrieben wurden, bleiben im Hinblick auf die bisherige risikoethische Forschung wesentliche Fragen offen. Beispielsweise sind Summenbildungen, wie sie in Kostenfunktionen meist erfolgen, immer in der einen oder anderen Form aggregativ, wohingegen aus risikoethischer Sicht stets die Perspektive der Einzelnen einzunehmen ist. Die Beurteilung komplexer Beziehungsgüsse zwischen Akteuren steht im Zentrum der Risikoethik, wird in gegenwärtigen Implementierungsversuchen jedoch kaum berücksichtigt. So versäumen es die Ansätze bislang, verantwortungsethische Aspekte umfassend in Betracht zu ziehen.<sup>250</sup> Geisslinger et al. (2021) unterstreichen die Notwendigkeit, ihren Entwurf durch die Integration spezieller Gewichtungsfaktoren weiterzuentwickeln, um Verantwortlichkeiten formal zu repräsentieren. Im Rahmen eines neueren Ansatzes versuchen Geisslinger et al. (2023a), konkrete

---

250 Auch in dieser Forschungsarbeit spielt der Aspekt der Verantwortung keine explizite Rolle in der inhaltlichen Gestaltung des risikoethischen Entwurfs, ist jedoch in Form der Prämissen der deontologischen Grenzkriterien, die in Kap. 7.3.2 und Kap. 7.3.3 begründet werden, implizit enthalten.

Ausprägungen von Verantwortung über die Reduzierung der Risikokosten innerhalb des bayesianischen Funktionsterms auszudrücken. Ihre Auffassung ist dabei einerseits sehr spezifisch; andererseits ist generell fraglich, inwiefern sich (moralische) Verantwortung in quantifizierbaren Faktoren überhaupt adäquat abbilden lässt. Formale Modellierungen, die für Implementierungen unverzichtbar sind, stellen letztlich immer Vereinfachungen dar; komplexe Konzepte können bei einer Darstellung in für Maschinen verständlichen Formeln meist nur verkürzt wiedergegeben werden, wobei u. U. essenzielle Aspekte verloren gehen oder in ihrem ethischen Kern nicht adäquat abgebildet werden. So versuchen Geisslinger et al. (ebd.), eine auf den Anwendungskontext reduzierte Auffassung von Verantwortung in ihre Kostenfunktion zu integrieren, die der Vielschichtigkeit des ethischen Konzepts an solches nicht gerecht werden kann.

Dietrich (2021) hingegen argumentiert zwar explizit verantwortungsethisch, um zu begründen, weshalb mögliche Ungerechtigkeiten bei der Programmierung von Systemalgorithmen aktiv durch die Integration von Gerechtigkeitszielen thematisiert werden müssen. Für deren konkrete Implementierung greift er jedoch wiederum auf die nach Rawls'schem Verständnis zwar kontraktualistisch begründete, aber konsequentialistisch operierende *Maximin*-Regel zurück, welche ihrerseits ethische Schwierigkeiten aufweist (siehe Kap. 6.3.2.2).<sup>251</sup>

Sowohl Geisslinger et al. (2023a) als auch Dietrich (2021) verzichten darauf, die Rechtfertigbarkeit der verwendeten rationalen Prinzipien kritisch zu prüfen. Einen entsprechenden argumentativen Versuch legt Berkey (2022, S. 217–222) vor, indem er eine faire Zuweisung von Risiken im Kontext automatisierter Mobilität fordert:

[...] firms ought to program autonomous vehicles in ways that aim to ensure that the safety-related benefits of those vehicles are distributed as fairly as possible among all of those who could potentially be harmed as a result of the use of motor vehicles. (Ebd., S. 216)

---

251 Als Grundlage einer verantwortungsethischen Diskussion technologiebezogener Risiken siehe Van de Poel und Nihlén Fahlquist (2013), die das Verhältnis von Risiko und Verantwortung sowie Verantwortungskonzeptionen, insbesondere das *Problem of Many Hands*, thematisieren.

Er diskutiert mögliche Gründe, mit denen ein Abweichen von anfänglich gleichverteilten Risiken gerechtfertigt werden könnte. Ein solcher besteht zum einen in der Vorrangstellung, die er der Reduzierung des absoluten Risikolevels bzw. des Gesamtschadens gegenüber dem Postulat gleichverteilter Risiken einräumt: »The aim of reducing injuries and deaths as much as possible should, on my view, take priority over distributing risks equally when there is a conflict.« (Ebd., S. 225) So würde die Annahme, dass autonome Fahrzeuge das Sicherheitsniveau erhöhen, implizieren, dass im Mischverkehr diejenigen weniger Risiko tragen sollten, die anstelle von konventionellen selbstfahrende Fahrzeuge nutzen.<sup>252</sup> Zum anderen ist eine Differenzierung zwischen Insassen motorisierter Verkehrsmittel einerseits und Fußgängern bzw. Radfahrern andererseits ethisch relevant, da Letztere signifikant niedrigere Risiken für andere heraufbeschwören. In diesem Kontext sieht Berkey vor allem eine prinzipielle Priorisierung von Fahrzeuginsassen kritisch.

Auch Abney (2022) deutet mögliche Grundpfeiler einer dezidiert risikoethischen Perspektive an. Im Kontext von Missbrauch und Manipulation von Systemalgorithmen beschreibt er verschiedene Faktoren, die die Akzeptabilität eines Risikos konstituieren. Zu diesen zählen eine (volumfängliche) Zustimmung bzw. Freiwilligkeit hinsichtlich eines eingegangenen Risikos, der Umfang der betroffenen Gruppen (Beteiligte vs. Unbeteiligte), die Natur relevanter Risiken (»state risk« vs. »step risk«), Schwellenwerte bei Ausmaß und Wahrscheinlichkeiten sowie die Legitimation von Entscheidungsinstanzen. Er erörtert diese Aspekte jedoch nicht näher, sondern formuliert sie lediglich als Forschungsimpulse, die in Richtung einer fundierten risikoethischen Analyse weisen. Eine solche zu skizzieren ist das Anliegen der vorliegenden Forschungsarbeit. Nachfolgend werden zunächst deren Gegenstand und Ziele näher spezifiziert.

### *7.1.3 Gegenstand und Ziele eines alternativen risikoethischen Entwurfs*

Zentrales Ziel der vorliegenden Arbeit ist es, den Forschungsdiskurs moralischer Unfalldilemmata weiterzuentwickeln, indem untersucht

---

<sup>252</sup> Plausibel ist dies jedoch nur unter der Voraussetzung, dass alle Individuen potenziell gleiche Zugangschancen zu automatisierter Mobilität haben.

wird, welchen Beitrag eine risikoethische Perspektive zur Klärung offener Fragen leisten kann. In Anbetracht der im vorhergehenden Unterkapitel aufgezeigten Forschungslücke ist festzuhalten, dass dem Forschungsdiskurs eine systematisch entwickelte, dezidiert risikoethische Grundlage fehlt, auf der sich sowohl Dilemma-Szenarien als auch Routinefahrsituationen ganzheitlich ethisch wie auch gesellschaftlich-sozial interpretieren lassen. Der hier vorgelegte Entwurf knüpft sowohl an diejenigen Beiträge an, welche die Problematik aus einem implementierungsnahen Blickwinkel betrachten, als auch an diejenigen, die sich auf distributive Aspekte fokussieren. Im Gegensatz zu Ersteren wird in dieser Arbeit unabhängig von (technischen) Erfordernissen möglicher Implementierungsansätze eine ethische Perspektive entworfen, ohne jedoch technisch relevante Spezifika des Anwendungsproblems gänzlich aus den Augen zu verlieren. Es werden keine konkreten Implementierungsprinzipien erarbeitet, sondern grundlegende ethische Fragen diskutiert, aus denen sich normative Implikationen für Implementierungsversuche ableiten lassen. Diese Herangehensweise trägt dem interdisziplinären Charakter der Fragestellung insofern Rechnung, als sie betont, dass auch anwendungsnahe Forschung nicht ohne fundierte, theoretische (ethische) Grundlagen auskommt, die zunächst unabhängig von (formalen) Realisierungsfragen zu klären sind. Auf diese Weise soll vermieden werden, dass es durch eine zu starke Fokussierung auf die technische Realisierung bzw. Realisierbarkeit zu einer Vernachlässigung der Plausibilität und Überzeugungskraft – und damit der Qualität – der zugrundegelegten ethischen Argumente kommt.

Wie in Kap. 4.2 gezeigt, zeichnen sich Dilemma-Szenarien im Kontext automatisierter Mobilität durch eine gesellschaftlich-soziale Dimension aus, der im Zuge des moralphilosophisch dominierten Zugangs kaum Beachtung geschenkt wurde. Der in diesem dritten Teil entwickelte Entwurf kann als Versuch verstanden werden, der Forderung nach einer institutionalistischen Problemperspektive zu entsprechen, wie sie beispielsweise von Smith (2022) formuliert wird. Unter Bezugnahme auf das »distributive Paradigma« wird an die bisherigen risikoethischen Beiträge angeknüpft und zugleich über diese hinausgegangen. Dabei werden u. a. Rawls'sche Konzepte vor dem Hintergrund einer risikoethischen Gesamtinterpretation von Unfallszenarien und als Teil eines deontologischen Entwurfs in-

tegriert, während bisherige Forschungsarbeiten in einer ausschließlich distributiven Perspektive verbleiben.

In der nachfolgend dargestellten risikoethischen Auseinandersetzung werden ausgehend von den in Kap. 6 vorgestellten theoretischen Grundlagen zentrale risikoethische Grundfragen in Bezug auf das Anwendungsproblem auf systematische Weise thematisiert. Einzelne risikoethische Argumente werden verknüpft und schließlich im Rahmen einer Skizze der Grundzüge deontologischer Risikoethik zusammengeführt. Der auf diese Weise erarbeitete risikoethische Entwurf kann als Alternative zu den in Teil II rekonstruierten Diskursen begriffen werden und bietet neue Impulse für zukünftige Auseinandersetzungen mit Unfallalgorithmen. So werden mögliche Ansätze der Rechtfertigung der moralischen Zulässigkeit von Risikoübertragungen anhand von Dilemma-Szenarien diskutiert, welche im Kontext der automatisierten Mobilität auftreten können. Diese dürfen jedoch nicht so strikt sein, dass jegliche praktische Relevanz verloren geht; dies wird vor allem hinsichtlich des in Kap. 5 erarbeiteten metaethischen Postulats der Nicht-Verrechenbarkeit inkomensurabler Werte problematisiert. Zudem werden Entscheidungsstrategien bezüglich der Problematik einer gerechtfertigten Verteilung individueller Vor- und Nachteile im spezifischen Problemkontext kritisch erörtert, wobei deontologische Prinzipien – anders als in den bisherigen implementierungsnahen Ansätzen – als *hard constraints* aufgefasst werden.

Den formalen Anforderungen hinsichtlich des begrenzten Umfangs dieser Arbeit ist es geschuldet, dass dabei nicht alle Themenkomplexe<sup>253</sup> der Risikoethik gleichermaßen intensiv bearbeitet werden können. Einige werden nur angedeutet und bilden weiterführende Forschungsdesiderate (siehe Kap. 8.2). Die risikoethisch relevanten Aspekte der (numerischen) Risikobewertung und Risikoabschätzung sowie verantwortungsethische Fragen werden im Folgenden daher nur gestreift, jedoch nicht umfassend behandelt. Im nächsten Unterkapitel wird die risikoethische Analyse anhand von zentralen Konzepten und Kriterien systematisch entfaltet.

---

253 Siehe die fünf Problemfelder der Risikoethik aus Kap. 6.2.3.

## 7.2 Analyse der Risikokonstellationen in Dilemma-Szenarien entlang von Kriterien der Risikoakzeptabilität

### 7.2.1 Akteure, Beziehungsnetzwerke und private Risiken

Wie in Kap. 4 ausgeführt wurde, lässt der bisher dominante Forschungszugang zentrale ethische Fragen im Kontext der Gestaltung von Unfallalgorithmen ungeklärt; der Fokus auf die (moralphilosophische) Frage nach der ethisch ›richtigen‹ Entscheidung ergibt keine final rechtfertigbaren Strategien. In diesem Kapitel wird eine risikoethische Auseinandersetzung vorgelegt, welche es erlaubt, einen alternativen Betrachtungswinkel auf das Anwendungsproblem einzunehmen, der Impulse für neue Entscheidungsperspektiven bereitstellt. In einem ersten Schritt wird zunächst eine risikoethische Interpretation des zugrundeliegenden Entscheidungsproblems skizziert, das durch seine dilemmatische Struktur eine besondere Risikokonstellation begründet.<sup>254</sup>

Risikoethische Überlegungen kommen immer dann zum Zuge, wenn ethisch relevante Unsicherheiten im Hinblick auf Handlungsfolgen bestehen. Grundlage jeder risikoethischen Betrachtung ist 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. Hinsichtlich der Gestaltung von Unfallalgorithmen kann zum Zeitpunkt der Entscheidungsfindung, und damit auch der ethischen Bewertung entsprechender Algorithmen, nur von Risiken, nicht aber von tatsächlichen Schädigungen ausgegangen werden:

The harm that interests us in the case of traffic is not one of intentionally harming third parties. Our main focus lies on actions that might cause harm to others but usually do not, meaning actions that put other members of society at risk, even if the risk of getting harmed for each individual is relatively low. (Müller & Gogoll, 2020, S. 1552)

---

254 Die Grundstruktur einer risikoethischen Perspektive auf Unfalldilemmata hat die Autorin bereits an anderer Stelle in groben Zügen entwickelt (vgl. Schäffner, 2020b).

Gegenstand des ethischen Entscheidungsproblems sind also nicht tatsächliche Schäden, sondern vielmehr *Schadensrisiken*. Weiterhin folgt bereits aus der Definition von Dilemmata, dass Schaden in diesen spezifischen Situationen unausweichlich ist, d. h. für mindestens ein beteiligtes Subjekt muss plausiblerweise das Eintreten einer negativen Konsequenz angenommen werden.

Diese Betrachtungsweise wirft die Frage auf, wie mit drohenden Schäden (= Risiken) in Dilemma-Szenarien umgegangen werden soll. Die Programmierung von Unfallalgorithmen lässt sich als risikoethisches Verteilungsproblem interpretieren; die Frage nach der moralphilosophischen Begründung von Entscheidungsprinzipien für Unfalldilemmata, die im Zentrum des dominanten Forschungszugangs steht, wird dabei transformiert in die risikoethische Frage der Rechtfertigbarkeit von Dilemma-Risiken sowie der fairen Verteilung daraus resultierender Vor- und Nachteile. Im Rahmen der nachfolgenden Auseinandersetzung wird das Anwendungsproblem vor dem Hintergrund zweier miteinander zusammenhängender risikoethischer Grundfragen neu entworfen: Einerseits stellt sich die Frage, wie sich entstehende Risikoübertragungen moralisch rechtfertigen lassen; andererseits geht es darum, unweigerlich bestehende Risiken in einer Weise zuzuweisen, die ethisch plausiblen und validen Kriterien genügt. Beide Problemstellungen sind eng miteinander verwoben und lassen sich nicht vollständig unabhängig voneinander untersuchen. Aus diesem Grund werden beide Aspekte im Folgenden stets im Zusammenhang betrachtet.

Als sozio-technische Systeme bewegen sich autonome Fahrzeuge in einem komplexen Verkehrsgeschehen, das durch das Zusammenwirken miteinander interagierender und aufeinander reagierender Subjekte charakterisiert ist. Dabei sind am Straßenverkehr teilnehmende Personen zum einen jenen Risiken ausgesetzt, die durch andere Verkehrsbeteiligte entstehen, und zum anderen verursachen sie selbst Risiken für andere; diese Praxis wechselseitiger Risikoübertragungen gilt sowohl für die Insassen autonomer Fahrzeuge als auch für andere Verkehrsteilnehmer (vgl. Grunwald, 2015, S. 667). Das zentrale Element, das die Risikokonstellationen im Kontext von Unfalldilemmata kennzeichnet, besteht in der Reziprozität von Risikoübertragungen. Jedes aktive Eingreifen in das Verkehrsgeschehen – ob Ausweichmanöver oder Neupositionierung innerhalb einer Fahrspur, Überqueren einer Straße oder Überholmanöver – stellt eine

dynamische Umverteilung von Risiken dar. Sicherheit und Risiko der Einzelnen hängen daher entscheidend von den Handlungen anderer ab:

The choice situation—in terms of levels of autonomy and risk to driver—is almost entirely determined by others. Drivers may select certain ›categories‹ of autonomy or risk, but they have very little control over the delineation of the categories themselves. The expected consequences of these decisions are deeply entangled with the choices of others on the road, from pedestrians to animals to cyclists to other vehicles (some of whom may not have drivers at all)—and attempts to manage these risks are dependent on the other actors. (Smith, 2022, S. 284)

Auf diese Weise übertragene Risiken existieren nicht einfach; sie sind vielmehr Bestandteil eines Beziehungsnetzwerks, welches die Akteure durch ihre wechselseitige Risikourheberschaft generieren: »Risks do not just ‚exist‘ as free-floating entities; they are taken, run or imposed. Risk-taking and risk imposition involve problems of agency and interpersonal relationships [...].« (Hansson, 2007b, S. 27) Risiken sind stets an Handlungen gebunden und etablieren ein spezifisches Beziehungsgefüge zwischen Entscheidenden (*decision-makers*), Betroffenen (*risk-exposed*) und potenziellen Begünstigten (*beneficiaries*). Für die risikoethische Beurteilung konkreter Risiken ist es essenziell, wie sich diese drei Rollen zueinander verhalten: Wer trifft die Entscheidung über eine Risikoübertragung und mit welcher Absicht? Profitieren Entscheidende oder andere Personen davon, Dritte einem Risiko auszusetzen? Findet eine Kompensation von Betroffenen durch Begünstigte statt? (Vgl. ebd., S. 27–28)

Im Zuge der Konkretisierung derartiger Fragestellungen auf das Anwendungsproblem von Unfallalgorithmen sind zunächst die risikoethisch relevanten Akteure zu spezifizieren. Bezogen auf Szenarien der Risikoübertragung können unter ›Entscheidenden‹ diejenigen verstanden werden, welche eine risikobehaftete Handlung initiieren, in deren Folge Risiken für andere entstehen. In der Regel sind in diesem spezifischen Kontext Entscheider und Begünstigte gleichzusetzen, denn diese sind durch ihre (risikoverursachenden) Aktionen in der Lage, ihre eigenen Projekte zu verfolgen, und profitieren daher. Dies könnte beispielsweise auf die Insassen eines autonomen Fahrzeugs zutreffen, die durch die automatisierte Beförderung Zeit für andere Tätigkeiten wie Lesen oder Arbeiten gewinnen.

Da sie sich jedoch durch ihre Teilnahme am Straßenverkehr immer zugleich auch selbst einem Risiko aussetzen, können sie – u. U. und in Abhängigkeit von der spezifischen Konstellation – auch zu den Betroffenen zählen, ebenso wie jede andere beteiligte Person. Weiterhin ergeben sich aus der Analyse der Beziehungsverhältnisse zwischen Risikoakteuren auch Implikationen für die Identifikation relevanter Risikokonstellationen. Diese werden zunächst nur neutral dargestellt und erst in einem zweiten Schritt ethisch bewertet bzw. hinsichtlich ihrer Akzeptabilität diskutiert.

Was die Beziehung zwischen einem Risiko und den von diesem Betroffenen angeht, erscheint es strittig, inwiefern individuelle Risiken, in denen sich Personen selbst einem (privaten) Risiko aussetzen, ohne dass Auswirkungen für andere bestehen, im Kontext automatisierter Mobilität als relevant anzusehen sind. Wie Nida-Rümelin et al. (2012, S. 154) anmerken, stellt das Konzept der allgemeinen Gefährdungshaftung klar, dass allein durch die »Beteiligung am Straßenverkehr auch bei korrektem Verhalten ein gewisses Risiko für andere Personen« besteht. Eine Teilnahme am Straßenverkehr ist daher nicht widerspruchsfrei ohne die Verursachung von Externalitäten denkbar. Dennoch kann eine individuelle Risikosituation vorliegen, denn eine solche kann unter bestimmten Voraussetzungen auch durch ein Kollektiv verursacht werden:

Alle Mitglieder müssen willentlich diesem Kollektiv beigetreten und eine mögliche ungleiche Verteilung von Konsequenzen, resultierend aus einer Risikosituation, muss von allen akzeptiert sein. Sind diese Bedingungen erfüllt, dann kann von einer individuellen Risikosituation gesprochen werden, dessen Urheber ein Kollektiv ist. (Ebd., S. 29)

Es ist grundsätzlich denkbar, ein solches Kollektiv dynamisch als die Gesamtheit aller Verkehrsbeteiligten aufzufassen, die sich zu einem gewissen Zeitpunkt quasi »im System« befinden und deren jeweilige Risiken innerhalb der Systemgrenzen bleiben; die Teilnahme am Verkehr kann dann als willentliche Akzeptanz der »Spielregeln« ausgelegt werden. Fraglich ist allerdings, inwiefern diese Zustimmung mangels Alternativen freiwillig erfolgt. Ebdies greifen Nida-Rümelin et al. in ihrer Darstellung eines weiteren Typus von Risikosituationen auf, »in denen zwar ein Risiko übertragen wird, dies aber durch die Zustimmung eines ausreichend informierten Risiko-Betroffenen in Abwesenheit jeglichen Zwangs legitimiert wird.« (Ebd., S. 29) Ist die Zustimmung freiwillig und auf Basis ausreichender

Informationen erfolgt, so gilt das Risiko als kollektiv eingegangen, wobei die potenziellen Konsequenzen nicht über die beteiligten Individuen hinausgehen dürfen.

Dennoch ist es im Hinblick auf das Anwendungsproblem nicht sinnvoll, Dilemma-Szenarien primär als individuelle Risikosituationen zu interpretieren. Erstens sind individuelle Risikosituationen mit Konsequenzen katastrophalen Ausmaßes definitionsgemäß unvereinbar, was den praktischen Gegebenheiten jedoch widerspricht, wie in den nachfolgenden Unterkapiteln gezeigt wird. Zweitens ist anzumerken, dass – vorausgesetzt, die von Nida-Rümelin et al. beschriebenen Sonderfälle individueller Risikosituationen würden hier greifen – sich die risikoethischen Fragen nach der Zulässigkeit von Risikoübertragungen bzw. den Kriterien einer fairen Risikoverteilung gar nicht oder nur in sehr begrenztem Maße stellen. Aufgrund der unterschiedlichen Voraussetzungen hinsichtlich individueller Risikoverursachung und Risikotoleranz sind erwarteter Nutzen und erwartete Risiken zwischen Anspruchsgruppen in der Regel derart asymmetrisch verteilt bzw. umstritten, dass eine freiwillige Akzeptanz derselben nicht zweifelsfrei angenommen werden kann. Daher erscheint es sowohl praktisch als auch ethisch notwendig, diese Fragen zu klären. Aus diesem Grund richtet sich der Fokus der vorgelegten risikoethischen Auseinandersetzung auf übertragene Risiken, d. h. Situationen, in denen ein Individuum oder eine Gruppe von einem Risiko betroffen ist, welches nicht selbstverursacht ist; das Risiko geht von den verursachenden Personen auf den bzw. die Betroffenen über (vgl. Rath, 2011, S. 30). Die real-lebensweltliche Manifestation dieser Konstellation ist im Kontext des autonomen Fahrens auf vielfältige Weise denkbar. Im Folgenden werden zur Veranschaulichung der Problematik zunächst verschiedene relevante Szenarien kurz umrissen.

### 7.2.2 Szenarien der Risikoübertragung

Aufgrund der Vielfalt möglicher Szenarien, in denen moralische Dilemmata mit Beteiligung autonomer Fahrzeuge auftreten können, sind Risikoübertragungen in verschiedenen Beziehungskonstellationen denkbar. Diese lassen sich hinsichtlich ihrer risikoethischen Problematik kategorisieren. Im einfachen Unfallszenario gehen von einem selbstfahrenden Fahrzeug Risiken für andere Verkehrsbetei-

ligte aus, wie in den Beispielszenarien 1 ›Großmutter versus Kind‹ und 2 ›Einzelperson versus Gruppe‹. Umgekehrt verursachen all diese Akteure ebenfalls Risiken für im konkreten Szenario involvierte Parteien, i. e. die Insassen des autonomen Fahrzeugs sowie die jeweils anderen Fußgänger(gruppen). Diese klassische Dilemma-Konstellation kann aus risikoethischer Sicht als Standardfall gelten, der sich durch komplexere ethisch relevante Beziehungsverhältnisse erweitern und ausdifferenzieren lässt.

Bevor diese Varianten im Folgenden erläutert werden, ist zunächst eine Anmerkung zur Situationsstruktur nötig. Da Schaden in moralischen Dilemmata definitionsgemäß unvermeidbar ist, stellen diese Nullsummenspiele dar. Dieses ursprünglich aus der Spieltheorie stammende Konzept impliziert im ökonomischen Sinne, dass die Summe aller Gewinne auf dem Markt stets null ist, weshalb der Erfolg eines Marktteilnehmers immer zulasten des Verlustes eines anderen geht. Übertragen auf den Kontext von Dilemma-Szenarien bedeutet dies, dass die Summe aller Risikowahrscheinlichkeiten stets ›1‹ ergibt, d. h. wenn sich das Risiko einer Partei reduziert, erhöht sich dasjenige der anderen. Daraus folgt, dass risikorelevante Handlungen in geschlossenen Systemen stets mit der Übertragung von Risiken verbunden sind. Voraussetzung dafür ist die Annahme, dass Risikokonstellationen dynamisch sind; jede noch so geringe Aktion eines jeden Beteiligten hat das Potenzial, die Risikokonstellation grundlegend zu verändern.

Hinsichtlich der Schadenskomponente des Risikobegriffs besteht ein solcher interpersoneller Zusammenhang zwischen den Risiken der Beteiligten nur indirekt bzw. in Fällen, in denen risikorelevante Aktionen direkte Externalitäten aufweisen. So senkt die Wahl schwerer Fahrzeugtypen auf der einen Seite tendenziell das zu erwartende Schadensausmaß für die Insassen, erhöht dieses aber zugleich für Radfahrer oder Fußgänger. Auf der anderen Seite wirken sich erhöhte individuelle Sicherheitsmaßnahmen wie das Tragen spezieller Schutzkleidung durch Motorradfahrer nicht direkt auf die zu erwartenden Schadenshöhen anderer aus – wohl aber indirekt auf deren Risiken über die Wahrscheinlichkeitskomponente, wie es in Beispielszenario 4 ›Motorradfahrer mit/ohne Helm‹ unter der gleichzeitigen Annahme eines utilitaristischen Schadenskalküls der Fall ist.

Vor diesem Hintergrund konstituieren Handlungen, die sowohl in Bezug auf Eintrittswahrscheinlichkeit als auch auf Schadenshöhe

direkt oder indirekt Einfluss auf Risikoübertragungen nehmen, verschiedene vom Standardfall abweichende Dilemma-Varianten. Eine solche ist beispielsweise auf verkehrswidriges Verhalten zurückzuführen. So kommt es in Beispielszenario 3 ›Rote Ampel‹ aufgrund des regelwidrigen Überquerens einer Fußgängerampel zu einer Risikoverschiebung zu Ungunsten des nachfolgenden Verkehrs, in Beispielszenario 6 ›Tunnel‹ ist das abrupte Abbremsen die Ursache für eine erhöhte Risikoexposition desselben. Einen ethisch komplexen Spezialfall innerhalb dieser Kategorie stellen Szenarien dar, in denen Risikoübertragungen durch bewusstes Fehlverhalten bzw. manipulatives Verhalten provoziert werden (vgl. Abney, 2022, S. 259–260; Hevelke & Nida-Rümelin, 2015c, S. 19–20). Schützen Algorithmen z. B. im Zweifelsfall stets Fußgänger, so könnten sich Letztere durch rücksichtsloses Handeln einen Sicherheitsvorteil verschaffen. Auch dem ohne Helm fahrenden Motorradfahrer in Beispielszenario 4 ›Motorradfahrer mit/ohne Helm‹ bieten sich im Falle eines auf die Realisierung der geringsten Schadenssumme ausgerichteten Algorithmus Fehlanreize, aufgrund derer die betreffende Person Teile ihres eigenen Risikos auf andere übertragen kann.<sup>255</sup>

Eine weitere Kategorie von Risikoübertragungen manifestiert sich in Szenarien, in denen neben beteiligten Parteien auch Unbeteiligte gewissen Risiken ausgesetzt werden, wie die Fußgängerin in Beispielszenario 5 ›Unbeteiligte auf Bürgersteig‹ oder die Kinder in Beispielszenario 8 ›Herannahender LKW‹.<sup>256</sup> Eine ethisch besonders kontroverse Variante sind sogenannte Selbstopferungsszenarien, in denen das Steuerungssystem des autonomen Fahrzeugs, wie in den Beispielszenarien 6 ›Tunnel‹ und 7 ›Klippe‹, eine Kollision mit unbelebter Infrastruktur bevorzugt, in deren Folge die Insassen zu Schaden kommen, um andere Verkehrsteilnehmer zu schützen. In

255 Auch wenn diese Szenarien meist mit einer an spezifischen ethischen – meist utilitaristischen – Prinzipien ausgerichteten Programmierung der Algorithmen assoziiert sind, ist ein Ausnutzen von deren Funktionsweisen grundsätzlich unabhängig davon möglich, welche konkreten Prinzipien zugrunde gelegt werden, sofern diese transparent und allgemein bekannt sind.

256 Zur ethischen Relevanz der Differenzierung zwischen Beteiligten und Unbeteiligten wurden bereits in Kap. 4.4.1 einige Argumente aus der philosophischen Ethik erläutert. Auch in Kap. 7.3.3 wird diese Problematik nochmals risikoethisch aufgegriffen; an dieser Stelle genügt vorerst die Bemerkung, dass sich die Gesamtkonstellationen in relevanten Szenarien durch Risikoübertragungen auf Unbeteiligte verändern.

diesen Fällen wirken die Risiken, die von den Passagieren durch die Nutzung des autonomen Fahrzeugs auf andere externe Parteien übertragen wurden, gewissermaßen auf ihre Initiatoren zurück.

Im Hinblick auf die Tragweite bzw. das Ausmaß verursachter Risiken lassen sich die in Unfallszenarien relevanten Risikosituationen weiter spezifizieren. Da Dilemma-Situationen selten auftreten, sind die Eintrittswahrscheinlichkeiten entsprechend zu erwartender Schäden sehr gering. Dies wäre im Sinne spezifischer Positionen der Risikoethik als triviales Risiko zu werten. Für Thomson (1985a, S. 128–137) liegt ein solches Risiko genau dann vor, wenn die Eintrittswahrscheinlichkeit hinreichend klein ist, was sie anhand ihrer prominenten Beispiele des Gasherds einerseits und des Russischen Roulettes andererseits demonstriert. Mögliche Konsequenzen sind dabei jeweils irrelevant. Hinsichtlich der spezifischen situativen Charakteristika von Unfallszenarien erscheint es jedoch implausibel, diese als triviale Risikosituationen zu kategorisieren. Wie die in Kap. 3.2.3 aufgeführten Argumente belegen, sind Individuen im Hinblick auf mögliche Risiken bzw. Erwartungswerte in Dilemma-Situationen keineswegs indifferent, was nach Nida-Rümelin et al. (2012, S. 47) jedoch eine Voraussetzung für das Vorliegen eines trivialen Risikos sein müsste.

Andere hingegen definieren triviale Risiken vorrangig über die Tragweite eines Risikos; so betont Posner (2004, S. 141), dass ein triviales Risiko nur dann vorliegt, wenn die Konsequenzen zumindest nicht katastrophal sind, ungeachtet ihrer Eintrittswahrscheinlichkeit. Dies lässt sich auf zweierlei Weisen auf Dilemma-Szenarien übertragen: Zum einen sorgen die systematischen Effekte der in einer großen Zahl von Fahrzeugen wirksamen, identischen Steuerungssoftware dafür, dass diese nicht nur in einer konkreten Situation Anwendung findet, sondern potenziell viele Fälle betrifft, von denen nicht bekannt ist, wann, wo und wie sie auftreten werden. Die entsprechenden Risiken werden also quasi auf die Gesamtheit aller Verkehrsbeteiligten übertragen. Zum anderen sind Szenarien denkbar, in denen unmittelbare Risiken katastrophalen Ausmaßes verursacht werden, z. B. wenn es im Zuge einer Kollision zum Ausstreuen von chemischen Schadstoffen oder größeren Mengen Treibstoff kommt, was wiederum eine große Zahl von Menschen auf katastrophähnliche Weise tangiert. Auch wenn diese Szenarien berechtigterweise als unwahrscheinlich gelten können, sind Risiken

in dilemmatischen Fahrsituationen auf dem Kontinuum möglicher Risikotragweiten in jedem Fall näher am Extrempunkt der katastrophalen als an jenem der trivialen Risiken anzusiedeln; eine exakte Verortung ist im Hinblick auf pragmatische Aspekte nicht zwingend notwendig. Im nächsten Unterkapitel erfolgt nun die risikoethische Evaluation der hier vorgestellten Szenarien unter konsequentialistischen und vertragstheoretischen Gesichtspunkten.

### *7.2.3 Diskussion aus Sicht konsequentialistischer und kontraktualistischer Kriterien*

Nachdem zuvor veranschaulicht wurde, welche Konstellationen in Unfallszenarien denkbar sind, werden diese in einem zweiten Schritt aus risikoethischer Perspektive analysiert. Besonderes Augenmerk wird dabei auf die spezifische ethische Komplexität gelegt, die durch die zugrundeliegende Dilemma-Struktur begründet wird. Während es offensichtlich moralisch falsch ist, einen Fußgänger zu überfahren, sind Handlungen, die mit sehr geringer Wahrscheinlichkeit zur selben Konsequenz führen, nicht grundsätzlich verboten. Die entscheidende Frage ist, wie schwach die kausale Verknüpfung zwischen Ursache und Wirkung sein muss, um eine entsprechende Handlung zu legitimieren. Hansson (2003, S. 292–301) bezeichnet dies als das »mixture appraisal problem«, das ethische Prinzipien bzw. Theorien auf unterschiedliche Weise beantworten. Die Risikoethik stellt im Wesentlichen drei Begründungsansätze bereit, die jeweils spezifische Kriterien hinsichtlich der moralischen Zulässigkeit von Risikoübertragungen beinhalten: Der erste Ansatz umfasst quantitative Optimierungen bzw. Kosten-Nutzen-Abwägungen auf der Basis konsequentialistischer Argumente, während der zweite kontraktualistische Elemente wie die Zustimmung Betroffener in den Fokus rückt. Der dritte Ansatz schließlich argumentiert deontologisch vor dem Hintergrund individueller Rechte und allgemein anerkannter Gerechtigkeitsprinzipien. Welcher Begründungsversuch im konkreten Fall anwendbar ist, hängt von den jeweiligen situativen Umständen ab. Die nachfolgende risikoethische Analyse erfolgt daher entlang etablierter Kriterien der Risikoakzeptabilität, anhand derer sowohl die Frage nach der Zulässigkeit von Risikoübertragungen als auch distributive Aspekte erörtert werden.

## 7. Unfallalgorithmen als risikoethisches Verteilungsproblem

### 7.2.3.1 Konsequentialistische Kriterien: Grenzen quantitativer Optimierung

Konsequentialistische risikoethische Kriterien bilden den Kern rationaler Risikopraxis. Sie beantworten die Frage nach der Zulässigkeit von Risikoübertragungen mithilfe der Quantifizierung von Eintrittswahrscheinlichkeiten und Schadenshöhen in Form eines numerischen Erwartungswerts, der über alle möglichen Folgen einer Handlungsoption aggregiert (vgl. Nida-Rümelin et al., 2012, S. 36-37). Bezogen auf Unfalldilemmata des autonomen Fahrens würde dies implizieren, dass diejenige Handlungsalternative mit dem geringsten zu erwartenden Gesamtschaden gewählt werden sollte. Allerdings ist die Fokussierung auf eine rein quantitative Bestimmung von Risikowerten als Basis ethischer Bewertungen fragwürdig. An dieser Stelle sei auf die entsprechenden Ausführungen in Kap. 6.3.3 verwiesen, deren Ergebnisse hier nur in zusammengefasster Form wiedergegeben werden. Hauptkritikpunkte an konsequentialistischen Ansätzen sind, dass zum einen durch die Aggregation im Sinne der Optimierung des gesamten Risikoerwartungswerts individuelle Rechte und Risikopräferenzen systematisch missachtet werden. Zum anderen bleiben dabei auch Verteilungsaspekte und Gerechtigkeitsabwägungen unberücksichtigt; ethisch problematisch ist dabei insbesondere, dass unter dem konsequentialistischen Ansatz der Nachteil des einen stets durch einen genügend großen Vorteil eines anderen gerechtfertigt werden kann (vgl. Hansson, 2003, S. 295).

### 7.2.3.2 Kontraktualistische Kriterien: Formen der Zustimmung und ihre Relevanz für Unfallalgorithmen

Kontraktualistische Ansätze der Risikoethik legitimieren die normative Gültigkeit von Prinzipien und entsprechenden risikobehafteten Handlungen unter Bezugnahme auf das zentrale Kriterium der Zustimmung. In der Risikoethik herrscht weitgehend Einigkeit darüber, dass triviale Risiken zustimmungslos übertragen werden dürfen, da

andernfalls das soziale Leben zu stark erschwert würde.<sup>257</sup> So bewertet Thomson (1985a, S. 134–137) einen grundsätzlichen Ausschluss von Handlungen mit inakzeptablen Konsequenzen, wie ihn beispielsweise das *Maximin*-Prinzip impliziert (vgl. Rath, 2011, Fußn. 37), ungeachtet ihrer Wahrscheinlichkeit als kontraintuitiv. Auch Leonard und Zeckhauser (1986, S. 45) lehnen es ab, triviale Risiken zu regulieren, weil es ineffizient und impraktikabel ist, von jedem potenziell Betroffenen eine Zustimmung einzuholen.

Die Übertragung von Risiken, deren Ausmaß nicht als trivial gelten kann, ist dagegen kontrovers. Dies hängt teilweise damit zusammen, dass, obwohl Zustimmung in allen kontraktualistischen Entwürfen als bedeutsames Konzept erachtet wird, unterschiedliche Ansichten darüber existieren, warum dies so ist. Scheffler (1985, S. 76–83) sieht z. B. rechtebasierte Begründungen als problematisch an, in deren Rahmen Zustimmung lediglich als Mittel interpretiert wird, um praktisch notwendige Verletzungen individueller Rechte mit ihrer Geltung zu versöhnen.<sup>258</sup> Es wäre stattdessen notwendig einzuschränken, wann etwas als Rechtsverletzung gilt, und in diesem Zuge z. B. triviale Risiken zu erlauben, denn über Zustimmung allein lassen sich nicht alle denkbaren Fälle klären: »[...] consent cannot serve as the sole vehicle for the justification of social institutions and patterns of social coordination, for its own value depends in part on those very institutions and patterns.« (Ebd., S. 85) In diesem Sinne spricht er sich gegen ein absolutes Verbot von Risikoübertragungen aus.

Auch weitere Lesarten hinsichtlich der Rolle von Zustimmung in Risikofragen gemäß Scheffler legen nahe, dass es unzulässig ist, Personen ohne deren Zustimmung Risiken auszusetzen. Während die eine die instrumentelle Bedeutung von Zustimmung (»its effectiveness as an instrument for promoting good states of affairs and avoiding bad ones« (ebd., S. 77)) akzentuiert, begreift die andere Zustimmung als intrinsischen Wert bzw. Komponente eines selbstbestimmten, guten Lebens. Alle drei Interpretationsweisen begründen jedoch kein absolutes Verbot von Risikoübertragungen. Ein

257 Siehe die Diskussion des *Problem of Paralysis* in Kap. 7.1.1.

258 In diesem Zusammenhang sind Rechtsbegründungen konsequentialistisch, es fehlt eine plausible nicht-konsequentialistische Begründung (vgl. Scheffler, 1985, S. 82).

solches wäre kontraintuitiv, wenn durch Risikoübertragung ein ›beserer Zustand‹ im Sinne der jeweiligen Ansätze herbeigeführt werden könnte. Letztlich implizieren sie eine Abwägung der Vor- und Nachteile von Handlungsalternativen, die weder zwangsläufig ökonomisch noch aggregativ sein muss, sondern durchaus distributive Effekte berücksichtigen kann (vgl. ebd., S. 78–79).<sup>259</sup>

In der Risikoethik werden drei Arten der Zustimmung diskutiert, die sich hinsichtlich wesentlicher qualitativer Kriterien unterscheiden: »Es können [...] drei Bedingungen unterschieden werden, die einer faktischen Zustimmung Gültigkeit verleihen: (1) Abwesenheit von Zwang, (2) Informiertheit und (3) Kompetenz« (Rehmann-Sutter, 1998, S. 51). Für die zustimmungsbasierte Legitimation von Risiken kann grundsätzlich postuliert werden: Je höher der Partizipationsgrad betroffener Individuen an einer Entscheidung, desto höher die Legitimität der entsprechenden Risiken (vgl. Gibson, 1985, S. 153; Rath, 2011, S. 33–34). Als robusteste Form gilt die *explizite Zustimmung*, die fordert, dass dem betreffenden Individuum alle verfügbaren Informationen vorliegen und es frei von äußeren Zwängen entscheidet (vgl. Thomson, 1985a, S. 137). Das zustimmende Individuum wird in diesem Zuge selbst zum Risikourheber, die Risikosituation damit von einer sozialen zu einer individuellen (vgl. Rippe, 2013, S. 524). Im Hinblick auf die Klärung konkreter Risikoübertragungen ist diese Form jedoch in den wenigsten risikoethisch relevanten Fällen gegeben, denn in der Regel ist weder die erforderliche Informationsbasis vorhanden noch sind die praktischen Rahmenbedingungen entsprechend effizient:

---

259 Es gibt noch einen anderen Spezialfall, unter dem Zustimmung kein notwendiges Kriterium darstellt. McCarthy (1997, S. 210–217) diskutiert in diesem Kontext einen Vorschlag von Thomson, demzufolge die Zulässigkeit einer Rechtsverletzung durch eine Risikoübertragung eine Frage des Abwägens ist. Dabei ist sowohl relevant, was sich daraus für den Verursacher ergibt, als auch das Maß, in dem der Geschädigte dadurch schlechter gestellt wird: »At least roughly, a rights infringement is permissible if the good that would come of the infringement sufficiently outweighs the burden of the infringement to the bearer of the right.« (Ebd., S. 210) Dies impliziert auch, dass eine Rechtsverletzung dann unzulässig bleibt, wenn sie dem Betroffenen unverhältnismäßig schadet. Auf einen Zustimmungsvorbehalt kann demnach verzichtet werden, wenn eine Rechtsverletzung im Sinne eines Abwägens der Vorteile des Verursachers und der Nachteile des Betroffenen zulässig wäre.

Actual consent is not a realistic criterion in a complex society in which everyone performs actions with marginal but additive effects on many people's lives. According to the criterion of actual consent, you have a veto against me or anyone else who wants to drive a car in the town where you live. Similarly, I have a veto against your use of coal to heat your house, since the emissions contribute to health risks that affect me. In this way each of us can block activities by many other persons. (Hansson, 2003, S. 300)

Schwächere Formen der Zustimmung sind im Hinblick auf praktische risikoethische Probleme relevanter, jedoch hinsichtlich ihrer Legitimität diskussionswürdig. Bei der *indirekten Zustimmung* resultiert eine risikobehaftete Handlung aus einem Entscheidungsverfahren, dem explizit zugestimmt wurde; das Individuum ist hierbei von konkreten Einzelfallentscheidungen ausgeschlossen (vgl. Rath, 2011, S. 34–35). Gemäß Thomson (1985a, S. 137–139) muss in diesem Zusammenhang genau spezifiziert werden, welchen konkreten Inhalt die Zustimmung umfasst. So hat man durch die Teilnahme an einer Lotterie noch nicht einem potenziellen finanziellen Verlust zugestimmt; ebenso wenig hat man durch die Entscheidung, durch ein verrufenes Stadtviertel zu spazieren, seine Zustimmung zu einem möglichen Überfall gegeben. Prinzipiell kann eine indirekte Zustimmung auch *implizit* erfolgen. In diesem Fall lässt sich die Zustimmung eines Individuums aus »ihrer institutionellen Eingebundenheit oder ihrem beobachteten Verhalten in anderen Kontexten« (Nida-Rümelin et al., 2012, S. 196) ableiten. Es liegt nahe anzunehmen, dass diese Form der Zustimmung in vielen Fällen zu zweifelhaften Ergebnissen führt.

Die bezogen auf ihre Legitimationskraft schwächste Form stellt die *hypothetische Zustimmung* dar, die davon ausgeht, dass einem Risiko explizit zugestimmt würde, sofern alle notwendigen Informationen vorlägen (vgl. Rath, 2011, S. 35). Da hier der Risikourheber quasi stellvertretend für den Risikobetroffenen entscheidet, ist diese Form der Zustimmung in ihrer Begründung prinzipiell fragwürdig, weil sie die individuelle Autonomie und Verantwortungsfähigkeit untergräbt.<sup>260</sup>

---

260 Um dieses Problem zu vermeiden, schlägt Gibson (1985, S. 151–152) vor, die Zulässigkeit hypothetischer Zustimmung auf Fälle zu beschränken, in denen es den Betroffenen, z. B. aufgrund ihres gesundheitlichen oder kognitiven Zu-

Welche Relevanz entfalten konkrete Formen der Zustimmung nun für den in dieser Forschungsarbeit untersuchten Anwendungsfall? Wie in den meisten real-lebensweltlichen Fragen der Gestaltung sozialen Miteinanders ist auch im Hinblick auf Unfallalgorithmen eine Annahme expliziter Zustimmung zu Risikoübertragungen nicht überzeugend. Aufgrund ihres restriktiven Charakters würden Mobilitätssysteme gänzlich zum Erliegen kommen, wenn für jedes denkbare Risikoszenario eine explizite Zustimmung eingeholt werden müsste:

Würde man versuchen, das explizite Zustimmungskriterium als allgemeines Kriterium für die Risikoexposition einzuführen, würde moralisch richtiges Handeln in einer Vielzahl von Lebenssituationen nahezu unmöglich. Selbst wenn man zu Fuß durch die Stadt geht, setzt dies andere Personen Risiken aus, wie zum Beispiel andere Fußgänger oder Fahrradfahrer, denen man in den Weg laufen könnte. Jeden um Erlaubnis zu fragen, verunmöglichte die Durchführung der Handlung. (Rippe, 2013, S. 528)

Eine auf indirekter Zustimmung basierende Entscheidungsstrategie für das Anwendungsproblem erscheint auf den ersten Blick naheliegender, erweist sich bei näherer Betrachtung aber ebenfalls als diskussionswürdig. Inwiefern kann die Beteiligung am Straßenverkehr bereits als Zustimmung zu sämtlichen kontextrelevanten Risikoübertragungen gewertet werden? Rippe (2013, S. 524) schreibt dazu: »Implizite Zustimmung liegt vor, wenn eine urteilsfähige Person, obwohl sie Informationen über ein Risiko hat, der Situation der Risikoexposition nicht ausweicht oder sich bewusst in sie hineinbegibt.« Gemäß dieser Definition nähmen die Insassen eines autonomen Fahrzeugs also durch ihre Wahl des Beförderungsmittels bewusst Risiken

---

stands, tatsächlich nicht (mehr) möglich ist, selbst zu entscheiden. Auch diese Fälle sind allerdings nicht von einer kritischen Beurteilung ausgenommen. Hansson (2003, S. 300–301) bemängelt, dass Vertreter der hypothetischen Vertragstheorie bisher den Nachweis schuldig sind, wie ein hypothetischer Vertrag – also einer, der nicht tatsächlich eingegangen wird – bindend sein kann. Speziell für Fälle mit involvierten Risiken bzw. Unsicherheiten gilt zudem, dass keiner der existierenden Ansätze zeigen konnte, wieso überhaupt ein solcher Zustand angenommen werden sollte, denn auch in einem solchen kann keine für alle Beteiligten zustimmungsfähige Strategie erreicht werden: »In particular, the debate following Rawls's Theory of Justice has shown that there is no single decision-rule for risk and uncertainty that all participants in Rawls's hypothetical initial situation can be supposed to adhere to.« (Ebd., S. 301).

in eigener Verantwortung in Kauf. Doch um *welche* Risiken handelt es sich dabei? Gemäß Thomson (1985a, S. 137–139) ist es entscheidend zu spezifizieren, *wozu* inhaltlich konkret zugestimmt wurde. Es ist davon auszugehen, dass Verkehrsteilnehmer Risikoübertragungen unter der Voraussetzung zustimmen, dass sich »alle Handelnde[n] an gewisse Regeln halten, denen man wiederum zustimmt oder zumindest vernünftigerweise zustimmen kann.« (Rippe, 2013, S. 529) Risikoübertragungen, die sich aus nicht-regelkonformem Verhalten ergeben, sind demnach nicht Gegenstand der Zustimmung und somit auch nicht legitimiert; so etwa in Beispielszenario 4 ›Motorradfahrer mit/ohne Helm‹, wo für die Risikoübertragung, die daraus resultiert, dass der andere keinen Helm trägt, keine Zustimmung angenommen werden kann.

Im Kontext automatisierter Mobilität umfasst die Forderung regelkonformen Verhaltens sowohl menschliche Akteure als auch autonome Fahrzeuge. Die Zusicherung zustimmungsfähigen Verhaltens kann nur über ein ethisch rechtfertigbares Design von Steuerungsalgorithmen erfolgen. Eine entscheidende Rolle kommt hierbei den Verflechtungen zwischen ethischer Risikoakzeptabilität und soziologischer Risikoakzeptanz zu (siehe Kap. 7.1.1). Mit der Beteiligung am Straßenverkehr nehmen Individuen gewisse gesellschaftlich akzeptierte Risiken in Kauf. Im Zuge der Automatisierung des Verkehrs verändert sich jedoch die Risikostruktur relevanter Situationszusammenhänge grundlegend. Zugleich ist Mobilität weiterhin ein Grundbedürfnis, auf das nicht einfach verzichtet werden kann, auch wenn neu generierte Risikokonstellationen individuell inakzeptabel sein mögen. Im Grunde hat jemand, der den neuartigen Risiken eigentlich nicht zustimmt, keine Alternative; er entscheidet nicht frei von äußeren Zwängen, was die Plausibilität einer in diesem Kontext implizit erfolgten Zustimmung in Frage stellt. Ferner ist die Annahme, Nutzer stimmten durch ihre Nutzung eines selbstfahrenden Fahrzeugs dem implementierten ethischen Entscheidungsalgorithmus vollumfänglich zu, nur so lange plausibel, wie die Akteure auch die Möglichkeit haben, sich potenziell anders zu entscheiden. Sollten autonome Fahrzeuge eines Tages verpflichtend werden, wäre die Voraussetzung der Zwangsfreiheit nicht länger erfüllt und die Legitimierung von Risikoübertragungen müsste neu verhandelt werden.

Dennnoch kann auch für den Fall, dass keine Einschränkungen bzw. äußerer Umstände gegeben sind, die als hinreichend zwingend angesehen werden können, die Beteiligung am Verkehr noch nicht automatisch als willentlicher Akt der Zustimmung zur automatisierten Mobilität als Ganze verstanden werden. Eine indirekte Zustimmung setzt weiterhin voraus, dass die Individuen den Inhalt bzw. Gegenstandsbereich ihrer Zustimmung kennen. Hierzu ist es notwendig, dass sie über ausreichende Informationen über mögliche Auswirkungen ihrer eigenen Handlungen und der anderer Akteure verfügen. Zu diesem Zweck ist es unumgänglich, dass algorithmische Handlungsvorschriften transparent und nachvollziehbar sind. Ebenso wie Risiken menschlichen Fehlverhaltens prinzipiell erwartbar sind, muss auch die nicht vollständig eliminierbare Fehleranfälligkeit technischer Systeme in gewissem Maße akzeptiert werden, um das *Problem of Paralysis* zu vermeiden. Da jedoch der Raum möglicher Szenarien unendlich und die Informationsbasis naturgemäß unvollständig ist, kann es immer Situationen geben, über die man nicht nachgedacht und denen man folglich auch nicht zugestimmt hat.<sup>261</sup> Die Forderung einer vollumfänglichen und eindeutigen Informationslage erscheint daher vor dem Hintergrund pragmatischer Überlegungen zu stark:

Even if AVs only have a >slave morality< in which they always follow orders, and citizens implicitly consent to their use (through, say, political means), that still leaves unanswered whether the risk (of malfunction, unintended consequences, or other error) to *unintended* parties is morally permissible. After all, even if widespread consent is in some sense possible, it is completely unrealistic to believe that all humans affected by AI in AVs could give *informed* consent to their use. So does the morality of consent require adequate knowledge of what is being consented to? (Abney, 2022, S. 267, Hervorh. i. Orig.)

Ebenso wenig ist es jedoch plausibel, aus der willentlichen Teilnahme am Straßenverkehr zwingend zu folgern, dass die Individuen jeder prinzipiell denkbaren Risikoübertragung in diesem Kontext zustimmen. Gänzlich klären lässt sich die Problematik der inhaltlichen Unterbestimmtheit angenommener Zustimmungshandlungen letztlich nicht.

---

261 Dies lässt sich prinzipiell auf real-lebensweltliche Zusammenhänge verallgemeinern, denn die Zukunft ist immer unsicher.

Auch für das Konstrukt der hypothetischen Zustimmung wird eine allgemeine Zustimmungsfähigkeit der relevanten Risikoübertragungen vorausgesetzt. Die Programmierung von Unfallalgorithmen ist Teil der Regulierungsaufgabe automatisierter Mobilität. Einzelne Risikobetroffene werden zwar nicht explizit gefragt, jedoch genießt die Wahrung ihrer Autonomie hohe Priorität. Die Herausforderung besteht einerseits in der inhaltlichen Ausgestaltung eines Algorithmus, dem alle zustimmen würden, und andererseits in der ethischen Legitimierung stellvertretender Zustimmung, die unabhängig von inhaltlichen Aspekten zu begründen ist. Bisher fehlen überzeugende Ansätze, um eine hypothetische Zustimmung für praktische Probleme ernsthaft in Betracht zu ziehen.

In Anbetracht dessen lassen sich Dilemma-Szenarien nur begrenzt mithilfe der Konstrukte indirekter bzw. impliziter Zustimmung erklären. Vielmehr kann die Schaffung hinreichender Voraussetzungen für eine indirekte Zustimmung als erklärtes Ziel der Programmierung von Unfallalgorithmen verstanden werden; deren ethische Problematik lässt sich nur klären, wenn auch im Einzelfall (normativ) zustimmungsfähige Handlungsvorschriften entwickelt werden.<sup>262</sup>

Ein weiterer Typ kontraktualistischer Ansätze greift auf das Konstrukt der Risikokompensation als Zustimmungsalternative zurück. Um Fälle zu identifizieren, in denen Zustimmung nötig ist, wird zwischen kompensierbaren und nicht-kompensierbaren Risiken differenziert. Liegen Erstere vor, so kann eine Zustimmung entfallen;

---

262 Bis dato existieren in Bezug auf die ethische Problematik von Unfallalgorithmen keine systematisch ausgearbeiteten risikoethischen Entwürfe, die sich auf vertragstheoretische Argumente stützen. Der Ansatz von Derek Leben ist zwar prominent, aber nicht explizit risikoethisch. Dennoch lassen sich einige seiner Gedankengänge und Implikationen in risikoethische Argumente übertragen. Leben entwirft einen von Rawls inspirierten Algorithmus für autonome Fahrzeuge, dessen zentrale Idee auf der maschinellen Schätzung der Überlebenswahrscheinlichkeit jeder betroffenen Person bei jeder möglichen Handlung der Maschine basiert (vgl. Leben, 2017, S. 110–114). Davon ausgehend wird mittels eines iterierten *Maximin*-Prinzips berechnet, welche Option jede einzelne Person wählen würde, wenn sie sich in einer »original bargaining position of fairness« (ebd., S. 108) befände, bis hin zu einem Pareto-optimalen Zustand. Obwohl Lebens Argumentation auf Rawls' vertragstheoretischer Interpretation des *Maximin*-Prinzips beruht, betont sie damit letztlich auch die Bedeutung von Verteilungseffekten.

die Kompensation ersetzt sozusagen eine formale Zustimmung und legitimiert auf diese Weise die Risikoübertragung. Als kompensierbar gelten Risiken immer dann, wenn es sich um materielle Schäden handelt oder um solche immateriellen, die materielle Auswirkungen, z. B. in Form finanzieller Verluste, haben (vgl. Rath, 2011, S. 38; Thomson, 1986b, S. 157). Kompensation kann auf zwei Arten erfolgen: Entweder wird die betroffene Person *ex ante* auf Basis des errechneten Erwartungswerts für das Tragen des Risikos entschädigt oder die Entschädigung erfolgt *ex post* und entspricht dem tatsächlich manifestierten Schadenswert. Während eine Risikoübertragung mit Ex-Ante-Kompensation im Kern kontraktualistische Züge trägt, da sie auch abgelehnt werden kann, weist eine Kompensation *ex post* den Charakter einer Verbindlichkeit auf, die im Nachhinein beglichen wird (vgl. Rath, 2011, S. 39–40), und hat daher praktische Grenzen. So merken Leonhard und Zeckhauser (1986, S. 32–36) hier kritisch an, dass Fragen der Risikourheberschaft nicht immer eindeutig zu klären sind, woraus sich Konflikte bezüglich der Zuschreibung von Haftbarkeit und Verantwortung ergeben können. Sie plädieren daher für eine formalisierte Ex-Ante-Kompensation in Form eines Vertrags, der eine grundsätzliche Zustimmungsbereitschaft voraussetzt. Dies bedeutet, dass Kompensation nur dann eine Alternative zur Zustimmung sein kann, wenn sie *ex ante* erfolgt. Aus Sicht von Thomson (1986a, S. 66–71) hingegen kann die Ex-Post-Kompensation auch dann als gleichwertig angesehen werden, wenn sie Handlungen den Weg bereitet, bei denen *ex ante* keine Verhandlungen möglich sind. In jedem Fall lässt sich festhalten, dass Kompensation eine fehlende Zustimmung nur in sehr begrenztem Maße plausibel ersetzen kann.

Zustimmungsbedürftig sind soziale Risiken insbesondere dann, wenn sie nicht kompensierbar sind (vgl. Rath, 2011, S. 63). Dies trifft primär auf physische Schädigungen bis hin zu tödlichem Ausgang, katastrophale Schadensausmaße oder Beeinträchtigungen immaterieller Werte zu. Für diese Fälle, zu denen auch die Risiken im Kontext von Unfallalgorithmen zählen, müssen alternative Strategien entwickelt werden;<sup>263</sup> in rein kontraktualistischer Fassung sind

---

263 Der zentrale Beitrag von Raths (2011) Abhandlung besteht in der Entwicklung einer alternativen Entscheidungstheorie, die sich von den etablierten Kriterien (Bayes, *Maximin*, *Prinzip der Vorsicht*) unterscheidet, indem sie besonderen

Konzepte der Kompensation hier nicht zu begründen. Inwiefern sie dennoch im Rahmen deontologischer Ansätze integrierbar sind, wird im Folgenden gezeigt.

#### *7.2.4 Deontologische Risikoethik: Begründung, Ansätze und Konzeptionen*

Wie in Kap. 5 erörtert, sind Dilemma-Szenarien im Kontext des autonomen Fahrens durch eine Nicht-Verrechenbarkeit fundamentaler moralischer Gebote charakterisiert. Diese folgt aus der inkompatiblen Verschränkung legitimer Interessen verschiedener Personen, die sich auf incommensurable Werte zurückführen lassen. Die Ergebnisse der metaethischen Analyse implizieren, dass es durch Risikoübertragungen unvermeidbar zu moralisch unzulässigen Beeinträchtigungen von individuellen Rechten, Autonomie und Verantwortungsfähigkeit kommt. Wie zuvor dargelegt, erscheint ein radikales Verbot risikobehafteter Handlungen im Hinblick auf praktische Aspekte wenig plausibel. Es stellt sich daher die Frage, unter welchen Umständen sich spezifische, im deontologischen Sinne dilemmatische Risikoübertragungen sowohl allgemein als auch im Anwendungskontext ethisch rechtfertigen lassen. In der Grundordnung einer freiheitlichen Demokratie gelten bestimmte individuelle Rechte und Freiheiten kategorisch, d. h. sie unterliegen einem Abwägungsverbot und dürfen keinem Optimierungskalkül unterworfen werden. Diesem Umstand tragen deontologische Ansätze der Risikoethik Rechnung, die auf dem Grundgedanken basieren, dass nicht allein Schadensausmaß und Eintrittswahrscheinlichkeit für die moralische Bewertung eines Risikos relevant sind, sondern auch andere moralische Handlungsgründe Beachtung finden müssen.

Die Notwendigkeit einer deontologischen Risikoethik ergibt sich aus der Unfähigkeit konsequentialistischer Ansätze, bestimmte ethische Problematiken, die sich beispielsweise in Bezug auf Abwägungsverbote stellen, zu erfassen und normativ zu berücksichtigen. Ni-

---

Fokus auf die Entscheidungsinstanz legt. In Abgrenzung zu der Annahme traditioneller Ansätze, dass Risikoentscheidungen von einem neutralen Individuum getroffen werden, skizziert Rath im Rahmen seines sogenannten risikoethischen Kontraktualismus die Grundzüge einer prozeduralen Entscheidungsfindung durch Risikobetroffene.

da-Rümelin et al. (2012, S. 154–155) verdeutlichen dies anhand der »Tragödie der großen Zahl«: Wird eine hinreichend große Anzahl von Personen einem minimalen Risiko ausgesetzt, so führt dies trotz der geringen Eintrittswahrscheinlichkeit im Durchschnitt dennoch zu einer gewissen Anzahl an Toten. Aus konsequentialistischer Sicht wäre diese Risikoübertragung ebenso zu bewerten wie eine alternative Handlung, infolge derer die gleiche Anzahl an Personen gezielt getötet würde:

Eine konkrete Person zu töten oder ihre Tötung zuzulassen, obwohl sie vermeidbar ist, kann durch ökonomische Vorteile, Annehmlichkeiten des Lebens, den technologischen Fortschritt etc. nicht aufgewogen werden. Im Rahmen deontologischer Risikoethik können unterschiedliche Kriterien der moralischen Beurteilung eine Rolle spielen, die sich nicht auf ein fundamentales Kriterium wie das der Maximierung des Nutzenerwartungswertes reduzieren lassen. Das genannte Abwägungsverbot, das uns moralisch unverzichtbar erscheint, kann nur von deontologischen Risikoethiken integriert werden. (Ebd., S. 155)

Innerhalb der deontologischen Risikoethik existieren verschiedene Abstufungen. Eine strikt deontologische Risikoethik weist jegliche Relevanz von Handlungsfolgen für die moralische Bewertung einer Handlung zurück. Im Hinblick auf lebenspraktische moralische Dilemma-Strukturen sprechen Nida-Rümelin et al. (ebd., S. 159) von einer aporetischen Situation, die durch das absolute Abwägungsverbot im Rahmen einer strikt deontologischen Ethik zweiter Ordnung hervorgerufen wird. Um eine solche Situation vollständig aufzulösen, wäre es nötig, den Geltungsanspruch individueller Rechte zu relativieren, indem dieser an das Vorliegen gewisser Bedingungen geknüpft wird: Individuelle Rechte könnten nur dann geltend gemacht werden, wenn eine Handlungsoption existiert, bei der keine anderen deontologisch begründeten Pflichten verletzt würden. Besteht keine solche Option, werden die entsprechenden Rechtsansprüche ungültig und damit das Dilemma aufgehoben. Ist eine solche Perspektive im Fall des autonomen Fahrens denkbar? Aufgrund der Nicht-Verrechenbarkeit deontologisch begründeter Pflichten gibt es keinen moralischen Grund, der es rechtfertigen würde, das Leben einer bestimmten Person zu bevorzugen. Daher liegt es nahe anzunehmen, dass, wenn Individualrechte nicht-eliminierbare ethische Gründe darstellen, das entsprechende Dilemma bestehen bleibt und

Entscheidende zwangsläufig moralisch scheitern müssen (vgl. ebd., S. 160).<sup>264</sup>

Im Gegensatz dazu schließt eine nicht-strikt bzw. schwach deontologische Risikoethik eine konsequentialistische Optimierung nicht kategorisch aus, sondern lässt diese im Rahmen deontologischer Grenzen zu, die sich aus den normativen Merkmalen der spezifischen Entscheidungssituation ergeben. Übertragen auf den Anwendungskontext des autonomen Fahrens würde das z. B. bedeuten, dass Opferungen von Personen zum Zweck der Schadensoptimierung grundsätzlich unzulässig bleiben, weil dabei individuelle Le-

---

264 Die rechtebasierte Risikoethik diskutiert die Problematik eines kategorischen Verbots risikobehafteter Handlungen als sogenannte Risiko-These im Sinne der Prämisse eines moralisch begründeten Nicht-Schadens-Rechts. Dieses wird verletzt, wenn eine Person einem Risiko ausgesetzt wird, da dadurch möglicherweise individuelle Interessen tangiert werden, die zur Lebensführung relevant sind. Diese These wird oftmals als zu starke Einschränkung der menschlichen Handlungsfähigkeit angesehen (vgl. Hansson, 2003, S. 297–299). Darüber hinaus wird angezweifelt, ob eine auf Rechten basierende Ethik für risikoethische Fragestellungen grundsätzlich geeignet ist (vgl. Nozick, 1974, S. 73–76; Thomson, 1990, S. 243–246). Befürworter wiederum argumentieren, dass die Risiko-These nur dann *ceteris paribus* zum Problem der Handlungsunfähigkeit führen würde, wenn sie (nahezu) alle Handlungen verböte (vgl. Steigleder, 2016a, S. 254–261, 2016b, S. 440–441). Ein solches Verbot wäre allerdings nur dann notwendig, wenn Risikoübertragungen in jedem Fall eine unzulässige Rechtsverletzung darstellen würden, was nicht der Fall ist, z. B. wenn *ex ante* eine Zustimmung vorliegt, in Notwehrsituationen oder bei drohenden Schäden für Dritte (vgl. McCarthy, 1997, S. 217). Risikopraktische Fragen stehen in einem Spannungsfeld gleicher Geltungsansprüche der Rechte auf Nicht-Schaden einerseits und Verwirklichung maximaler Freiheit andererseits, das ein Abwagen von Rechtsverletzungen verschiedener Personen notwendig macht, wobei weder die eine noch die andere Seite unangemessen privilegiert werden darf (vgl. Steigleder, 2016b, S. 441). Die Bewertung der Angemessenheit orientiert sich am Kriterium der Unentbehrlichkeit der Rechte als »Voraussetzungen handelnder Selbstverwirklichung« (ebd., S. 441). Eine gewisse Risikotoleranz, die eine situativ einseitige Bevorzugung von Rechten zulässig macht, ist in diesem Sinne nicht nur pragmatisch, sondern auch normativ notwendig – sowohl auf Akteur- als auch Rezipientenseite. Straßenverkehrsrisiken stellen hierbei ein klassisches Beispiel dar, weil sie sich trotz umfanglicher Vorsichtsmaßnahmen nicht vollständig eliminieren lassen. Diese Risiken können als gering und damit kontrollierbar gelten; ein generelles Verbot würde die Rechte der Betroffenen unangemessen privilegieren und dabei die Freiheit der Akteure zu stark einschränken (vgl. Steigleder, 2016a, S. 262).

bensrechte verletzt werden.<sup>265</sup> »Risiko kann lediglich innerhalb der Grenzen optimiert werden, in denen sichergestellt ist, dass die Optimierung nicht durch die Verletzung von Rechten und anderen deontologischen Kriterien erkauft wird.« (Nida-Rümelin et al., 2012, S. 159) Mit dem Abwägungsverbot, das aus der Nicht-Verrechenbarkeit begründeter Pflichten folgt, gehen deontologische Einschränkungen einher, die bei jeglicher Risikooptimierung berücksichtigt werden müssen.<sup>266</sup> Auf diese Weise verfolgt eine deontologische Risikoethik das Ziel, einerseits Werte wie Freiheit, Autonomie und individuelle Rechte zu schützen, die durch Risikoübertragungen bedroht sind, andererseits aber auch nicht zu restriktiv zu sein, um eine völlige Paralyse sozialen Miteinanders zu vermeiden. Wie kann eine solche Risikoethik konkret aussehen? Im nachfolgenden Unterkapitel wird schrittweise ein entsprechender deontologischer Entwurf für den spezifischen Kontext von Unfallalgorithmen erarbeitet.

### 7.3 Grundzüge einer deontologischen Risikoethik für Unfallalgorithmen

#### 7.3.1 Kohärente Risikopraxis nach Julian Nida-Rümelin: Grundlinien, Ziele und Anwendung

Der Philosoph, Autor und ehemalige Politiker Julian Nida-Rümelin, der von 2020 bis 2024 Mitglied im Deutschen Ethikrat war, formuliert im Rahmen eines Kapitelbeitrags für das Handbuch *Angewandte Ethik* (1996) die Grundlagen einer deontologischen, in ihrer politischen Dimension kontraktualistisch verfassten Risikoethik. Ge-

---

265 Dies gilt, entgegen dem gleichlautenden Einwand, auch dann, wenn die Betroffenen noch gar nicht feststehen. Letztlich stehen diese in einer in gewisser Weise deterministischen Welt immer schon fest, wir können sie aufgrund unseres begrenzten epistemischen Vermögens lediglich nicht erkennen. In moralischer Hinsicht ist diese Situation demnach genauso zu bewerten wie die gezielte Schädigung von Personen (vgl. Nida-Rümelin et al., 2012, S. 152–153).

266 Das bedeutet ausdrücklich nicht, dass im Rahmen einer schwach deontologischen Risikoethik keine Abwägungen stattfinden, beispielsweise bei vorliegenden Zielkonflikten zwischen Komfort und Sicherheit; diese lassen sich allerdings kontraktualistisch plausibel begründen (vgl. Nida-Rümelin et al., 2012, S. 155).

meinsam mit Johann Schulenburg und Benjamin Rath legte er 2012 einen Band vor, der die sieben Jahre zuvor skizzierten Eckpfeiler aufgreift und weiterentwickelt. Den Ausgangspunkt dieser risikoethischen Abhandlung bildet die Feststellung, dass die zeitgenössische Risikopraxis Inkohärenzen aufweist, die sich vor allem hinsichtlich des Umgangs mit Hochtechnologien manifestieren. Zurückzuführen sind sie auf das Spannungsverhältnis zwischen konsequentialistischen und deontologischen Intuitionen, das sich in risikopraktischer Perspektive in den normativen Grundausrichtungen widerspiegelt, die Ökonomie einerseits und juridischen Normen des Rechtssystems andererseits zugrunde liegen. Diese Unterschiede sind, trotz ihrer auf den ersten Blick unvereinbar erscheinenden Grundsätze, in ein kohärentes Normensystem überführbar, welches konsequentialistische und deontologische Logik zusammenbringt (vgl. Nida-Rümelin et al., 2012, S. 161–175). Den Weg ebnet hierbei das Nutzentheorem, dessen Postulate nicht zwingend konsequentialistisch, sondern vielmehr inhaltlich neutral sind, und in das daher auch nicht-konsequentialistische Handlungsgründe wie gerechtigkeitsethische Überlegungen einfließen können.<sup>267</sup> Es »wird nicht das individuelle Wohlergehen zum Maßstab der Rationalität gemacht, sondern die Kohärenz der individuellen Präferenzen.« (Ebd., S. 165–166) Deontologische Normen sind also als Handlungsgründe zu verstehen, die sich in Form handlungsleitender Präferenzen formalisieren lassen, um im Rahmen einer quantitativen Bewertungsfunktion einzbezogen zu werden.

Ziel dieser deontologischen Konzeption ist eine kohärente Risikopraxis, die konsequentialistische und deontologische Elemente entsprechend ihrer jeweiligen Rechtfertigungsgrundlage integriert. Konkret bedeutet das: Eine (konsequentialistische) Risikooptimierung wird nicht kategorisch zurückgewiesen, sondern durch deontologische Grenzen eingeschränkt. Diese Konzeption scheint in gewissem Sinne an die in Kap. 4.4.4.5 formulierte Forderung eines pluralistischen Ansatzes anzuknüpfen, wie sie unter dominanten Forschungszugängen häufig vertreten wird. Eine kohärente Risikopraxis berücksichtigt zwar verschiedene ethische Handlungsgründe, recht-

---

267 Die vier Kohärenzpostulate lauten: Transitivität, Vollständigkeit, Monotonie und Stetigkeit der Präferenzen über Wahrscheinlichkeitsverteilungen (vgl. Nida-Rümelin et al., 2012, S. 173).

fertigt diese aber unabhängig voneinander. Die Idee, Unfallszenarien als primär utilitaristisches Optimierungsproblem zu interpretieren, das deontologische Prinzipien in Form von Nebenbedingungen (*soft constraints*) einbezieht, liegt den meisten bisher vorgeschlagenen, implementierungsnahen Ansätzen zugrunde (vgl. Geisslinger et al., 2021, 2023a; Gerdes & Thornton, 2015; Németh, 2022; Thornton et al., 2017). Im Gegensatz zu diesen und ähnlichen Ansätzen betont eine deontologische Risikoethik jedoch, dass deontologische Kriterien *harte Grenzen* (*hard constraints*) darstellen sollten, die Vorrang vor quantitativen Optimierungszielen haben: »Optimierung ist nur in den Grenzen zulässig, in denen vorrangige deontologische Kriterien nicht verletzt werden.« (Nida-Rümelin et al., 2012, S. 175)

Der Ansatz von Nida-Rümelin et al. nimmt Bezug auf die kontraktualistische Rekonstruktion einer deontologischen Ethiktheorie, die auf der Kompatibilität des Vertragsarguments von Immanuel Kant<sup>268</sup> (vgl. 1900ff., TP, AA 08) einerseits und den Grundzügen der Wesensbestimmung der Moral nach Thomas Scanlon (1982) andererseits beruht (vgl. Nida-Rümelin et al., 2012, S. 181–187).

---

268 Das weniger prominente kontraktualistische Argument von Kant lautet in konziser Fassung: »Hier ist nun ein ursprünglicher Contract, auf den allein eine bürgerliche, mithin durchgängig rechtliche Verfassung unter Menschen begründet und ein gemeines Wesen errichtet werden kann. – Allein dieser Vertrag (*contractus originarius* oder *pactum sociale* genannt), als Coalition jedes besondern und Privatwillens in einem Volk zu einem gemeinschaftlichen und öffentlichen Willen (zum Behuf einer bloß rechtlichen Gesetzgebung), ist keineswegs als ein Factum vorauszusetzen nöthig (ja als ein solches gar nicht möglich); gleichsam als ob allererst aus der Geschichte vorher bewiesen werden müßte, daß ein Volk, in dessen Rechte und Verbindlichkeiten wir als Nachkommen eingetreten sind, einmal wirklich einen solchen Actus verrichtet und eine sichere Nachricht oder ein Instrument davon uns mündlich oder schriftlich hinterlassen haben müsse, um sich an eine schon bestehende bürgerliche Verfassung für gebunden zu achten. Sondern es ist eine bloße Idee der Vernunft, die aber ihre unbezweifelte (praktische) Realität hat: nämlich jeden Gesetzgeber zu verbinden, daß er seine Gesetze so gebe, als sie aus dem vereinigten Willen eines ganzen Volks haben entspringen können, und jeden Unterthan, so fern er Bürger sein will, so anzusehen, als ob er zu einem solchen Willen mit zusammen gestimmt habe. Denn das ist der Probirstein der Rechtmäßigkeit eines jeden öffentlichen Gesetzes.« (Kant, 1900ff., TP, AA 08: 02-21, Hervorh. i. Orig.) Kants Werk *Über den Gemeinspruch: Das mag in der Theorie richtig sein, taugt aber nicht für die Praxis* wurde erstmals 1793 veröffentlicht.

Grundlage ist dabei folgender Gedankengang: Das Fundament einer deontologischen Ethik besteht in der kantianisch geprägten Konzeption individueller Autonomie, die mit Ansprüchen auf Freiheit und Gleichheit verbunden ist. Wie Scanlon im Rahmen seines ethischen Kontraktualismus erläutert, lässt sich eine auf diese Weise verstandene deontologische Ethik mit einem kontraktualistischen Verständnis dessen, was das Wesen der Moral ist, zusammenbringen. Auf dieser Basis stellt sich die Konzeption von Nida-Rümelin et al. letztlich als kontraktualistisch verfasste Risikopraxis auf der Grundlage deontologischer Begründungen dar, die ihre inhaltliche Ausgestaltung aus dem Wesen individueller Autonomie speist. Die zentralen Grenzkriterien einer solchen Risikopraxis für demokratisch und freiheitlich verfasste Gesellschaften sind daher durch individuelle Autonomie in Form von Individualrechten<sup>269</sup> einerseits und etablierte distributive

---

269 Die beiden wichtigsten Typen deontologischer Theorien sind die würdebasierte Moraltheorie, die auf Kant zurückgeht, einerseits und rechtebasierte Moraltheorien andererseits. Sie unterscheiden sich hinsichtlich dessen, was sie als normative Grundkategorie akzeptieren: Während die würdebasierte Moraltheorie moralische Pflichten aus der menschlichen Würde ableitet, legen rechtebasierte Moraltheorien moralische (oder juridische) Anspruchsrechte zugrunde. Im Allgemeinen werden rechtebasierte Ansätze kontrovers wahrgenommen; häufig werden sie als fundamentale Verkürzung des Phänomens des Moralischen betrachtet. Dies erscheint auch für das Themenfeld dieser Arbeit einschlägig. Eine Verletzung der relevanten Rechte ist nicht zuletzt deshalb moralisch unzulässig, weil Personen individuell autonom sind und daher Träger eigenständiger Entscheidungskompetenz, die ein Paternalismusverbot in Kraft setzt (vgl. Nida-Rümelin, 2005, S. 875–876; Nida-Rümelin et al., 2012, S. 57). Eine Argumentation allein auf der Basis von Anspruchsrechten würde den dahinter liegenden intrinsischen und inkommensurablen Werten der individuellen Autonomie und Freiheit nicht gerecht. Eine kritische Haltung gegenüber einem rechtebasierten Ansatz in der Risikoethik nimmt Hansson (2003, S. 289–299) ein. Er argumentiert, dass ein Nullrisiko auch bei absoluten Rechten impraktikabel ist: »Unfortunately, such a strict extension of rights and prohibitions is socially untenable. Your right not to be killed by me is certainly accompanied by a prohibition for me to perform many types of acts that involve a risk of killing you, but it does not involve a prohibition of all such acts. Such a strict interpretation would make human society impossible. I am allowed to drive a car in the town where you live, although this increases the risk of being killed by me.« (Ebd., S. 298) Zudem erscheint es implausibel, die Akzeptabilität von Risiken nur anhand ihrer Wahrscheinlichkeit zu beurteilen und mögliche durch sie hervorgerufene Vorteile außer Acht zu lassen. Aus

Gerechtigkeitsvorstellungen andererseits gegeben (vgl. Nida-Rümelin, 2005, S. 874).

Während der Entwurf von Nida-Rümelin et al. (2012) auf rein konzeptioneller Ebene verbleibt, wird die ihm zugrundeliegende Argumentation in diesem letzten Teil des Buches auf den Anwendungskontext von Unfallalgorithmen übertragen. Dabei wird der Ansatz vor dem Hintergrund der Problemspezifika inhaltlich konkretisiert und weiterentwickelt. Es werden zwei zentrale Prinzipien herausgearbeitet, die den beiden deontologischen Grenzkriterien im Hinblick auf die Realisierung einer kohärenten Risikopraxis für das Anwendungsproblem entsprechen und dessen Grundlage bilden. Diese beiden Grenzkriterien stellen nicht-verhandelbare Rahmenbedingungen dar, innerhalb derer sich die Verteilung von Vor- und Nachteilen aus Risikoübertragungen als *›neues Optimierungsproblem* bewegt.

Wie im zweiten Teil des Buches ausgeführt wurde, zeichnen sich Unfalldilemmata durch nicht-eliminierbare Individualrechte aus. Aus pragmatischer Sicht erscheint es jedoch plausibel zu fordern, dass gewisse Risiken in Kauf genommen werden sollten, ohne dass dies aus ethischer Sicht als Rechtsverletzung zu beurteilen wäre. Die Frage ist also, in welchem Rahmen Verletzungen von Individualrechten als zulässig gewertet werden können und wie diese zu begründen sind. Risikoethische Argumentationen sind dabei nicht aggregativ, sondern nehmen stets den Einzelnen in den Blick. Um die Wahrung individueller Autonomie trotz der Aufweichung des absoluten Geltungsanspruchs der entsprechenden Individualrechte zu gewährleisten, kann ein etabliertes risikoethisches Konzept zu Rate gezogen werden, das sowohl individuelle Interessen als auch soziale Erfordernisse in Betracht zieht: Risikoübertragungen sind genau dann zu akzeptieren, wenn sie den Einzelnen *zugemutet* werden dürfen. Neben ihrer Orientierung an den Interessen der Einzelnen sind Risiken zudem auch interpersonell relevant und damit beziehungskonstituierend. Für anwendungspraktische Fragestellungen ergibt sich aus ethischer Warte das Postulat einer fairen Verteilung von resultierenden Vor- und Nachteilen, die gerechtigkeitsethischen Anforderungen genügt. Wie diese beiden Forderungen vor dem Anwendungskontext automatisierter Mobilität konkret begründet und

---

den hier erörterten Gründen wird in diesem Buch auf eine tiefergehende und explizite Darstellung einer rechtebasierten Sichtweise verzichtet.

in Prinzipien umgesetzt werden können, die eine rechtfertigbare Risikopraxis konstituieren, wird im Folgenden diskutiert.

### *7.3.2 Die (absolute) Frage der Zumutbarkeit: Eine moralische Gratwanderung entlang von Risikoschwellen*

#### *7.3.2.1 Individualorientierung als Referenzpunkt*

Interaktionen im verkehrlichen Umfeld sind durch wechselseitige Risikoübertragungen charakterisiert, welche die individuelle Autonomie der Einzelnen betreffen. Aus praktischen Gründen sind gewisse Risiken zu akzeptieren, auch wenn dabei individuelle Rechte potenziell verletzt werden. Doch wo genau verläuft die Grenze zwischen dem, was akzeptiert werden muss, und dem, was inakzeptabel ist? Offensichtlich geht es darum, das rechte Maß zwischen dem (utopischen) Ideal eines Nullrisikos und einem eindeutig untragbaren Risiko zu finden. Diese Problemstellung wird in der Risikoethik unter dem Konzept der Zumutbarkeit gefasst, die das Maß bezeichnet, in dem ein Risiko aus ethischer Sicht als akzeptabel gelten kann.<sup>270</sup>

Entscheidend für die ethische Legitimierung zumutbarer Risiken ist, dass sie sich gegen eine aggregative Perspektive wendet. Zumutbarkeit nimmt als risikoethisches Kriterium stets das Individuum, seine Interessen und Ziele in den Blick. Das lässt sich zum einen auf das Bestehen individueller Abwehrrechte zurückführen (vgl. Rippe, 2013, S. 532); zum anderen ist es für eine ethische Beurteilung von hoher Relevanz, ob diejenigen, die von einem Gut oder Zustand profitieren, auch die entsprechenden Risiken tragen. Das Konzept der Zumutbarkeit tritt als Vermittler einer Verhältnismäßigkeit zwischen den individuellen Vor- und Nachteilen auf, die eine Risikoübertragung hervorbringt. Dabei werden zumeist subjektive Sicherheitsbedürfnisse tangiert, die abhängig sind von individuellen Risikopräferenzen. Jedoch sind objektive Risikoeinschätzungen vonnöten, um normative, interpersonell verbindliche Implikationen ab-

---

270 In der TFA spielt das Konzept der Zumutbarkeit ebenfalls eine zentrale Rolle, indem es eine Brücke zwischen Akzeptanz und Akzeptabilität von Technik schlägt (vgl. Grunwald, 2005).

zuleiten; sich an subjektiven Präferenzen zu orientieren ist weder praktikabel noch ethisch zu rechtfertigen. Vielmehr stellt Rippe (2013) fest, dass das ethische Problem der Zumutbarkeit von Risiken »nur dann zu klaren moralischen Antworten führen [kann], wenn es um sogenannte objektive, nicht um subjektive Risikoeinschätzungen geht.« (Ebd., S. 530)<sup>271</sup> Aus der Individualorientierung bei der Bewertung zumutbarer Risiken folgt also ausdrücklich nicht, dass sich die Zumutbarkeit eines Risikos an subjektiver Wahrnehmung bemisst. Implizit akzeptierte soziale Praktiken sind ebenfalls irrelevant. Zumutbarkeit ist kein Synonym für Sozialakzeptanz; es geht nicht darum, was *tatsächlich* sozial akzeptiert ist, sondern was *rational betrachtet* zu akzeptieren ist. Sie ist kein subjektives Empfinden, sondern nimmt individuelle Rechte bzw. Interessen ernst, indem sie mögliche Einschränkungen derselben thematisiert. Die entscheidenden Aspekte, die subjektive Risikopräferenzen in objektivierte Risikobewertungen transferieren, sind die Reziprozität in der Übertragung zumutbarer Risiken und der daraus resultierende Vorteil, keine absoluten Einschränkungen hinsichtlich der eigenen individuellen Handlungsfreiheit und der Verfolgung persönlicher Ziele hinnehmen zu müssen:

Wird von Zumutbarkeit gesprochen, ist darin enthalten, dass der Person etwas auferlegt wird, was sie vermeiden will, es aber von ihr zu teilende rationale Gründe gibt, dies dennoch anzunehmen. [...] Die zu teilenden rationalen Gründe sind hier letztlich, dass gewisse Risiken hinzunehmen sind, da man ansonsten auch selbst andere keinerlei Risiko aussetzen darf. (Ebd., S. 529)

In Kontexten mit wechselseitigen Risikoübertragungen stehen zumutbare Risiken in einem Spannungsverhältnis zwischen persönlichen Präferenzen und denen anderer. Eine interpersonelle Vergleichbarkeit zumutbarer Risiken ist jedoch nicht ohne Weiteres möglich. Grundsätzlich ist es zwar plausibel zu fordern, dass Personen diejenigen Risiken, die sie anderen zumuten, auch selbst hinnehmen müssen. Der Umkehrschluss gilt allerdings nicht: Ein Risiko, das man bereit ist für sich selbst zu akzeptieren, darf deshalb noch nicht anderen zugemutet werden (vgl. ebd., S. 531). Luhmann

---

271 Die Verwendung objektiver Risiken bezieht sich nur darauf, wie die Risiken in ihrer Höhe zu bestimmen sind, und nicht auf die Einstellungen, die einzelne Individuen diesen gegenüber haben.

(1997, S. 330) führt als Begründung hierfür mögliche Unterschiede in der persönlichen Risikopräferenz an. Bei eigenen Entscheidungen ist man tendenziell risikobereiter, darf aber anderen gleichzeitig nicht das zumuten, was man für sich selbst akzeptiert. Analog lässt sich im Sinne eines stärker rezipientenorientierten Ansatzes argumentieren, dass ausgehend von der eigenen Risikoeinstellung keine Rückschlüsse dahingehend gezogen werden dürfen, ob eine Risikoübertragung für ein anderes Individuum zumutbar ist (vgl. Birnbacher, 1996, S. 204–205).

Die grundrechtlich garantierte individuelle Autonomie gerät in praktischen Situationen regelmäßig mit risikobehaftetem Handeln in Konflikt. Basierend auf den bisherigen Ausführungen lässt sich ein (absolutes) Leitprinzip formulieren, mit dessen Hilfe unzulässige Rechtsverletzungen abgewendet werden können. Dieses stellt sicher, dass im Rahmen einer Risikooptimierung keine unzumutbaren Risiken für Einzelne entstehen, welche die individuelle Autonomie in unzulässiger Weise beeinträchtigen:

- (1) *Rechtsverletzungen, die durch Risikoübertragungen im Kontext von Unfallalgorithmen verursacht werden, können genau dann als moralisch zulässig gelten, wenn die auf die Einzelnen übertragenen Risiken unter Berücksichtigung aller ethisch relevanten Aspekte jeweils individuell zumutbar sind.*

Konkrete inhaltliche Kriterien für die Bewertung der Zumutbarkeit von Risiken im gegebenen Anwendungskontext ergeben sich im Zuge einer Auseinandersetzung mit den spezifischen Umständen relevanter Risikoübertragungen, wie nachfolgend dargelegt wird.

### 7.3.2.2 Risikoschwellen: Ansätze und Schwierigkeiten

Bereits unsere moralische Intuition legt nahe, dass gewisse Risiken zulässig sein müssen und es zu deren definitorischer Abgrenzung eines Grenzwerts bedarf. Zu den Prinzipien, die im Kontext der ethischen Zumutbarkeit von Risiken am häufigsten rezipiert werden, gehören Schwellenwert-Konzeptionen, deren Ziel es ist, eine Risikoschwelle zu bestimmen, unterhalb derer Risikoübertragungen als unproblematisch eingeschätzt werden können. In diesem Sinne liegt ein zumutbares Risiko dann vor, wenn ein nicht mehr weiter

minimierbares Restrisiko erreicht wurde, das »nur mit Maßnahmen weiter zu verringern wäre, die als unzumutbar gelten, [...] obwohl der unwahrscheinliche Schadensfall nach wie vor seinem Umfang nach missbilligt wird.« (Ropohl, 2017, S. 903–904) Eine Handlung kann somit genau dann als zumutbar gelten, wenn der Risikoerwartungswert aller Betroffenen einen bestimmten Grenzwert nicht überschreitet und dabei unabhängig von jeglicher Form der Zustimmung ist. So spricht Hansson (2003, S. 298–299) im Kontext deontologischer Theorien von einem »probability limit«; Rechte und Verbote dürfen verletzt werden, wenn die Eintrittswahrscheinlichkeit eines Schadens gering genug ist. Dies entspricht in etwa dem, was Ryazanov et al. (2023, S. 12) als »a more nuanced understanding of deontology« bezeichnen und das zwischen probabilistischen und sicheren Schäden zu differenzieren vermag: »While deontologists may not be willing to kill one to save five, they may deem it acceptable to risk a 1 % chance of harm to one to save five.« (ebd.)

Ein praktisches Beispiel für einen solchen Schwellenwert stellen Dosisgrenzwerte im Zusammenhang mit Fragen des Strahlenschutzes dar, welche die gerade noch zulässige Dosis festlegen (vgl. Hansson, 2007a, S. 152–154). Auch die von Thomson (1985a, S. 125) diskutierte Subklasse risikoethisch relevanter Handlungen, die sich durch »threshold effects« auszeichnen, ist hier einschlägig. Dabei werden zunächst triviale oder auch gar keine Risiken übertragen, jedoch entsteht infolge der wiederholten Ausführung einer Handlung durch dieselbe oder eine andere Person letztlich ein nicht-triviales Risiko, das nicht ohne Weiteres übertragen werden darf. Andere Ansätze verwenden das maximal akzeptable Risiko als zentralen Parameter zur Trajektorienplanung (vgl. Geisslinger et al., 2023a, 2023b). Sütfeld et al. (2019) plädieren ebenfalls für die Installation eines Risikoschwellenwerts, oberhalb dessen eine Risikoübertragung als unzulässiges Opfer angesehen wird, da ein absoluter Schutz von Unbeteiligten an einem gewissen Punkt nicht mehr vernünftig begründet werden kann:

The exact point at which it becomes unreasonable not to save the person at risk would need to be debated, but most will agree that at some point the absolute protection of bystanders from any risk will become unreasonable. [...] Introducing thresholds that define at which level of risk a decision is considered a sacrifice could recover some notion of

reasonableness, but only at the cost of robustness and transparency of the decisions. (Ebd., S. 10–11)

Eine der wesentlichen Anforderungen an Schwellenwert-Konzeptionen ist deren Konsistenz. Wurde ein konkretes Risiko als zumutbar eingestuft, muss dies für ein anderes, das bezüglich Schadensausmaß und Wahrscheinlichkeit gleich ist, ebenfalls gelten. Dabei muss jedoch zwischen Risikoübertragung und privatem Risiko unterschieden werden: Was eine Person sich selbst zumutet, darf ihr deshalb noch nicht durch andere zugefügt werden. So setzt sich eine Person selbst einem gewissen Risiko aus, wenn sie sich in einem autonomen Fahrzeug befördern lässt. Daraus folgt jedoch noch keine Legitimation für die Handlungen anderer, die ein Risiko mit demselben Erwartungswert generieren.

Als entscheidende Voraussetzung für die ethische Legitimation spezifischer Schwellenwerte gilt, dass alle ethisch relevanten Aspekte des Kontextes einbezogen werden. Eines der wichtigsten Probleme von Schwellenwert-Konzeptionen besteht in der numerischen Repräsentation von Risikoerwartungswerten, die mit spezifischen Problemen behaftet und deshalb nur begrenzt für risikoethische Fragen geeignet ist (siehe Kap. 6.3.3). Dies mag in Anwendungskontexten weniger problematisch sein, in denen die direkten Folgen relativ zuverlässig quantifizierbar und entsprechende Verantwortlichkeiten eindeutig sind, beispielsweise im Kontext der Emission radioaktiver Strahlung. Weitauß anspruchsvoller gestaltet es sich jedoch in Fällen, in denen nur rudimentäre Informationen über mögliche technisch induzierte Handlungsfolgen vorliegen.

Als Annäherung an die Problematik ist ein technischer Blickwinkel hilfreich. Zumutbarkeit nach technischem Verständnis beruht auf den Festschreibungen von Sicherheits- und Risikokonzepten gemäß den Vorgaben einschlägiger ISO-Normen. Die Vision des autonomen Fahrens ist eng mit dem Versprechen verbunden, die Zahl schwerer Unfälle signifikant zu reduzieren und unsere Straßen auf diese Weise sicherer zu machen. Doch wie sicher ist ›sicher genug‹? Papadimitriou et al. (2022, S. 9–11) sprechen in diesem Zusammenhang von einem Siloeffekt, da das Konzept der Sicherheit bisher aus unterschiedlichen wissenschaftlichen Perspektiven unabhängig voneinander untersucht wurde, die von ethischen bis zu rein technischen Aspekten reichen. In Forschungs- und Entwicklungsprojekten zu autonomen Fahrsystemen wird Sicherheit daher kaum als eigen-

ständiger Designwert anerkannt. Gemäß Birnbacher (1996, S. 197) ist Sicherheit »eher eine Zuschreibung als eine Beschreibung. Was sicher in diesem Sinne ist, gilt als sicher – und deswegen als akzeptabel und zumutbar.« Von Sicherheit kann im Hinblick auf technische Systeme gesprochen werden, wenn das Risiko unter dem Grenzrisiko liegt; dieses ist quasi das technische Äquivalent der ethischen Zumutbarkeitsschwelle. Gemäß ISO 26262 besteht ein zumutbares Risiko dann, wenn es unterhalb einer Schwelle liegt, die als »nicht akzeptabler Wert in einem spezifischen Kontext gemäß gesellschaftlicher, moralischer und ethischer Auffassungen« zu interpretieren ist:

Beim Betrieb eines automatisierten Fahrzeugs hängt das zumutbare Risiko von der aktuellen Situation, in der sich das Fahrzeug befindet, ab. Zur Situation gehören hier [...] alle für eine Fahrentscheidung relevanten stationären und dynamischen Objekte, die Intention der dynamischen Objekte einschließlich des autonomen Fahrzeugs, die geltenden rechtlichen Bedingungen, die Mission des autonomen Fahrzeugs, die aktuelle Leistungsfähigkeit des autonomen Fahrzeugs. (Reschka, 2015, S. 491)

Wie ist dieser Wert nun aber zu bestimmen? Geisslinger et al. (2023a, S. 140–141) empfehlen, als Initialwert vom derzeitigen Sicherheitsniveau des herkömmlichen Verkehrsgeschehens auszugehen, wobei das anvisierte Pendant für die automatisierte Mobilität zukünftig deutlich höher anzusetzen sein wird.

Schwellenwert-Konzeptionen erfordern, dass kontinuierlich in jeder Situation das Risiko für jeden beteiligten Verkehrsteilnehmer ermittelt wird und diejenigen Optionen identifiziert werden, bei denen das aktuelle und zukünftige Risiko unterhalb des Schwellenwertes bleibt. Erschwerend kommt hinzu, dass Risiken nicht nur von den Aktionen des eigenen Fahrzeugs, sondern auch vom Verhalten anderer abhängen. Dies ist technisch herausfordernd und bisher ungelöst. Potenziale bietet hier das Prinzip der funktionalen Degradation, um bei Eintreten sicherheitsrelevanter Ereignisse einen Zustand zumutbaren Risikos wiederherzustellen (vgl. Reschka, 2015, S. 505–506).

Risikoethische Ansätze greifen diese Aspekte auf und interpretieren sie unter Bezugnahme auf ethisch relevante Gesichtspunkte. Rippe (2013, S. 532) beschreibt zwei grundsätzliche Vorgehensweisen zur Bestimmung zumutbarer Risiken. Die eine stützt sich ausschließlich auf die Eintrittswahrscheinlichkeit als relevantes Kriterium und

definiert damit Ereignisse, die mit hoher Wahrscheinlichkeit eintreten, als unzulässig. Hierbei würden allerdings Risiken durch das Raster fallen, die zwar unwahrscheinlich sind, aber hohe Schäden verursachen. Um derartige Katastrophenrisiken auszuschließen, können zusätzlich die erwarteten Schadenshöhen veranschlagt werden; bei höheren Schäden wären nur sehr geringe Wahrscheinlichkeiten zugelassen. Dies entspricht z. B. dem Konzept des moralischen Erwartungswerts (*expected moral value*), den Bhargava und Kim (2017, S. 9) vorschlagen: »One cannot simply compare 0.2 and 0.8. One must consider the value of the outcomes. Of course, car designs cannot be perfect, but a 20 % probability of a life-threatening malfunction is obviously too high.« Um zwischen schweren und verhältnismäßig leichten Schäden zu differenzieren, wären hierbei je nach Schadensart unterschiedliche Schwellenwerte anzusetzen. Da eine kardinale Anordnung von Interessen hingegen kaum realisierbar erscheint, schlägt Rippe (2013) vor, Abwehrrechte gegen spezifische Risiken als Referenzpunkt zu nehmen, die sich hierarchisch anordnen lassen. Diese würden jeweils missachtet, wenn ein Risiko die für ein konkretes Recht als kritisch befundene Eintrittswahrscheinlichkeit übersteigt. Zumutbare Risiken lassen sich dann entlang von ordinalen Risikoschwellen bestimmen: »Die Zulässigkeit des Handelns lässt sich damit stets mit Blick auf jene Betroffenen prüfen, welche mit höchster Wahrscheinlichkeit in ihrem Recht gefährdet werden.« (Ebd., S. 534)

Dass individuelle Risikopräferenzen entscheidend für die Frage nach der Zumutbarkeit von Risiken sind, ist neben Eintrittswahrscheinlichkeit und Schadenshöhe ein zentrales Element des rezipientenorientierten Ansatzes, den Birnbacher (1996) vertritt. Diesem zufolge liegt der entscheidende Unterschied hinsichtlich der ethischen Frage, unter welchen Umständen wir anderen Personen Risiken anstelle sicherer Schäden zumuten dürfen, gerade darin, dass bei Ersterem die individuellen Risikoeinstellungen der potenziell Betroffenen berücksichtigt werden. Birnbachers Position stehen Ansätze gegenüber, »die von einem situations- und einstellungsinvarianten Zumutbarkeitsmaß ausgehen und entweder eine durchgängig risikoscheue oder durchgängig risikoneutrale Risikostrategie postulieren.« (Ebd., S. 205)

Daraus folgt, dass die Zulässigkeit einer Handlung stets von demjenigen Individuum abhängt, welches durch die Risikoübertragung

am schlechtesten gestellt wird. Dies gilt unabhängig davon, ob für die Bewertung der Risikoübertragung Risikopräferenzen, Eintrittswahrscheinlichkeit und Schadenshöhe oder nur die beiden Letzteren als relevant erachtet werden. In kritischen Situationen müsste für jede beteiligte Person das individuell zumutbare Risiko entsprechend dem jeweils involvierten Abwehrrecht bestimmt werden. Bei Unfalldilemmata hilft eine Hierarchisierung von Abwehrrechten hingegen nicht weiter. Es stehen sich im Hinblick auf den jeweils zu schützenden Wert inkommensurable, individuelle Abwehrrechte gegenüber. Je nach gewählter Trajektorie unterscheiden sich die übertragenen Risiken aber hinsichtlich ihrer Eintrittswahrscheinlichkeiten. So ist in Beispielszenario 2 ›Einzelperson versus Gruppe‹ das Risiko der einzelnen Person viel größer als das jedes Einzelnen in der Gruppe, wenn das Fahrzeug nach links ausweicht. Die im Verkehrskontext produzierten Risiken für Individuen sind sehr unterschiedlich und komplex, was u. a. auf bestehende Unterschiede in der individuellen Vulnerabilität<sup>272</sup> zurückzuführen ist. Da diese, neben der persönlichen physischen und psychischen Konstitution, vom gewählten Verkehrsmittel abhängt, lassen sich entsprechende Risikoübertragungen nicht allein auf Basis der Reziprozitätsprämisse legitimieren; diese ist nur dann anwendbar, wenn alle Beteiligten prinzipiell ähnliche Risiken generieren. Es kann jedoch nicht davon ausgegangen werden, dass alle Verkehrsmittel von allen Personen genutzt werden.

### 7.3.2.3 Kernkriterium moralische Sorgfaltspflichten

In bestimmten Fällen, die im Kontext autonomer Fahrsysteme nicht selten sind, bedarf es einer von der Reziprozitätsthese unabhängigen Begründung.<sup>273</sup> Besonders deutlich wird dies, wenn sich Personen bewusst manipulativ oder wenigstens unvorsichtig verhalten, indem sie Verkehrsregeln missachten. Reziprozität kann als quasi-kontraktualistisches Element nur funktionieren, wenn sich alle an gewisse Regeln halten. Unsere moralische Intuition sagt uns, dass bei bewusstem Fehlverhalten wie im Fall des Fußgängers, der in Bei-

---

272 Der Aspekt der Vulnerabilität wird in Kap. 7.3.3 genauer spezifiziert.

273 Versuche, Zumutbarkeit über (kontraktualistische) Verfahrenslösungen zu begründen, scheitern, wie die Diskussion möglicher Zustimmungsformen in Kap. 7.2.3.2 zeigt.

spielszenario 3 ›Rote Ampel‹ die Straße verkehrswidrig überquert, die Schwelle des zumutbaren Risikos für die entsprechende Person deutlich höher anzusetzen ist. Bewusste Regelmissachtungen schaffen Risiken, die anderen nicht zugemutet werden dürfen; derartige Verhaltensweisen fallen nicht mehr unter die individuelle Autonomie, denn diese endet dort, wo diejenige der anderen beginnt. Begründen lässt sich dies mithilfe eines weiteren risikoethischen Konzepts, dem Prinzip der Sorgfaltspflichten. Dieses stellt spezifische Anforderungen an sorgfältiges Handeln im Zusammenhang mit Risikoübertragungen, die sich sowohl auf die Minimierung der Eintrittswahrscheinlichkeit eines Schadens als auch des erwarteten Schadensausmaßes beziehen. Bei risikobehafteten Handlungen müssen angemessene Präventionsmaßnahmen getroffen werden, sowohl auf Seiten des Risikoverursachers als auch des Betroffenen (vgl. Rippe, 2013, S. 531).

Da Risikoübertragungen im Verkehrskontext zumindest der Art, nicht aber der Höhe nach reziprok sind, gilt diese Forderung prinzipiell für alle Verkehrsbeteiligten und äußert sich in vielfältigen Verhaltensweisen. So sollten Fußgänger sich aufmerksam und vorsichtig im Verkehr bewegen, Radfahrer auf ausreichende Abstände achten und Handzeichen für Richtungsänderungen einsetzen. In Bezug auf autonome Fahrsysteme sind hier vor allem Anforderungen an eine defensive und vorausschauende Fahrweise zu nennen:

Insofar as AVs are morally permitted to impose risks of harm to road users given their prudential goal of time efficiency, the AV must drive with the required level of caution so that it can come to a stop in what if cases that could easily occur, given the AV's evidence about relevant features of the driving environment. (Keeling, 2022, S. 44–45)

Die Passagiere eines autonomen (Privat-)Fahrzeugs müssen ihrerseits sicherstellen, dass die Sicherheitsanforderungen an die Fahrtüchtigkeit, technische Funktionsfähigkeit etc. erfüllt sind. In diesem Sinne nimmt die Person, die in Beispieldaten 5 ›Umbeteiligte auf Bürgersteig‹ die Straße vorschriftsmäßig überquert, ihre Sorgfaltspflichten ernst. Ob das auch auf die Insassen des selbstfahrenden Fahrzeugs zutrifft, dessen Bremsen unvorhergesehen versagen, kann hingegen nicht final geklärt werden. Der Motorradfahrer in Beispieldaten 4 ›Motorradfahrer mit/ohne Helm‹ generiert durch seinen wissentlichen Verzicht auf Schutzkleidung ein Risiko, das dem anderen, der seinerseits seine Sorgfaltspflichten diesbezüglich

erfüllt hat, nicht zugemutet werden darf. Auch das ungebremste Auffahren des LKW in Beispielszenario 8 ›Herannahender LKW‹ ist sicherlich unzumutbar, ebenso das verkehrswidrige Schneiden der Kurve durch den Schulbus in Beispielszenario 7 ›Klippe‹. In diesem letzten Fall wird zudem deutlich, dass auch andere ethische Probleme für die risikoethische Bewertung relevant sein können, denn Sorgfaltspflichten können nicht immer eindeutig geklärt werden. In diesem konkreten Szenario ist das Fehlverhalten dem Fahrer des Busses zuzuordnen, nicht aber den Insassen, die ebenfalls zu den Betroffenen zählen. Besonders in Szenarien, in denen Personen stellvertretend für andere risikobehaftete Entscheidungen treffen, dürfen Fragen der Verantwortlichkeit nicht außer Acht gelassen werden.

Diese beispielhaften Szenarien verdeutlichen, dass gemäß unserer ethischen Intuition gewisse Risiken bei der Bewertung regelwidrigen Verhaltens als unzumutbar einzustufen sind. Welche praktischen Implikationen sich daraus ergeben, bleibt allerdings zumindest teilweise unklar. Sind Sorgfaltspflichten erfüllt, gelten Risiken als zumutbar. Somit begründet das Grenzkriterium der Zumutbarkeit im Kern die moralische Pflicht, Sorgfaltspflichten zu ergreifen, um unzumutbare Risiken zu vermeiden und dadurch die individuelle Autonomie jedes Einzelnen zu wahren. Die entscheidende Frage ist nun, in welchem Umfang Sorgfaltspflichten im Kontext von Dilemma-Situationen gefordert werden sollten und wann diese als erfüllt anzusehen sind.

Der moralische Charakter von Entscheidungen zur Gestaltung von Unfallalgorithmen wurde bereits vielfach betont. Agieren Verkehrsbeteiligte in dem Bewusstsein, dass sie Risiken auf andere übertragen, so stellen sich sämtliche Handlungsentscheidungen im Verkehrsgeschehen als moralische dar. Es ist also zu fragen, welches Risiko angesichts der eigenen Sorgfaltspflichten, die man realisiert oder unterlassen hat, zumutbar ist. Aufgrund der moralischen Dimension von Unfalldilemmata liegt es nahe, an dieser Stelle einen moralischen Sorgfaltsbegriff zugrunde zu legen: Sorgfalt im moralischen Sinne bedeutet, unter Berücksichtigung spezifischer situativer Merkmale zu begründeten, rechtfertigbaren Entscheidungen zu gelangen. Dies umfasst eine moralische Verantwortung, der man sich nur stellen kann, indem man sich der Pflicht zur sorgfältigen Einbeziehung sämtlicher moralisch relevanter Gründe annimmt. Als moralfähige Wesen schulden wir einander eine gewisse Sorgfalt in der Begründung unserer moralischen Entscheidungen, die sich in einen

gesellschaftlichen Kontext stellen lassen (vgl. Nyholm & Smids, 2016, S. 1278–1279). Dieser Umstand begründet die Notwendigkeit, ethische Entscheidungsprozesse gewissenhaft zu durchdenken:

[...] what's important isn't just about arriving at the ›right‹ answers to difficult ethical dilemmas, as nice as that would be. But it's also about being thoughtful about your decisions and able to defend them – it's about showing your moral math. In ethics, the process of thinking through a problem is as important as the result. (Lin, 2014a, o. S.)

Das zentrale moralische Argument, das im Rahmen dieses Sorgfaltsprizes zur Anwendung kommt, besteht in der Reziprozität, durch die soziale Gefüge – oder im Fall automatisierter Mobilität: soziotechnische Gefüge – erst tragfähig werden. Die Gegenseitigkeit von Risikoübertragungen bildet die Grundlage des zweiten Grenzkriteriums, indem sie es einerseits erlaubt, die eigene Freiheit durch risikobehaftetes Handeln zu verwirklichen, andererseits aber zugleich an die Pflicht zur Wahrung der Autonomie der anderen appelliert, denen die gleiche Freiheit zusteht. Dabei bereitet das Kriterium der Zumutbarkeit in gewisser Weise die Basis für die Diskussion gerechtigkeitsbezogener Fragen, wie sie im folgenden Unterkapitel entwickelt wird. Gerechtigkeitsethische Fragen sind nicht vollständig unabhängig von jenen der Zumutbarkeit, weshalb an geeigneten Stellen Rückbezüge stattfinden bzw. das Kriterium der Zumutbarkeit nochmals aufgegriffen und erweitert wird:

Es geht daher im ›harten‹ Kern der Debatte zu Technikakzeptanz einerseits um die *unfreiwillig einzugehenden Zumutungen und ihre gesellschaftliche Verteilung, die der – im Prinzip nicht in Frage gestellte – technische Fortschritt mit sich bringt* [...]. (Grunwald, 2005, S. 58, Her vorh. i. Orig.)

### 7.3.3 Die (relative) Frage der Gerechtigkeit: Zwischen Reziprozität und Vorteilsausgleich

#### 7.3.3.1 Unfallalgorithmen vor dem Hintergrund egalitärer Gerechtigkeitskonzeptionen

Risikobehaftete Entscheidungen als eine Frage fairer Risikoverteilungen anzusehen, ist in anderen Anwendungskontexten bereits etabliert. Goodall (2017, S. 496) nennt als beispielhafte Problemstellun-

gen die Zuweisung von Organspenden im Gesundheitswesen, die Zulässigkeit von Strahlungsbelastungen, die Beibehaltung der Wehrpflicht oder Vorgaben für industrielle Sicherheitsstandards. Wie in Kap. 3.1.1 bereits gezeigt wurde, sind weder Algorithmen an sich noch ihre Effekte wertneutral. Neben der Forderung eines individuell zumutbaren Risikos wirft die Unvermeidbarkeit von Schäden, die charakteristisch für ethische Dilemmata ist, auch distributive Fragen zur moralischen Zulässigkeit von Risikoübertragungen auf; »a choice is required about how to distribute harms or risks of harm between multiple people whose interests are in conflict.« (Keeling, 2022, S. 50) Welche gerechtigkeitsethischen Kriterien sollen bei der Verteilung von Vor- und Nachteilen aus spezifischen Risikokonstellationen Anwendung finden? Da der thematische Rahmen dieses Buches keine umfassende philosophische Auseinandersetzung mit komplexen gerechtigkeitsethischen Fragen vorsieht, werden im Folgenden lediglich einige grundlegende Überlegungen zu problematischen Aspekten skizziert.

Die prominente Gerechtigkeitskonzeption von John Rawls (1971), welche in der gerechtigkeitsethischen Debatte als Dreh- und Angelpunkt gilt, fußt auf einem kantianisch geprägten Verständnis individueller Autonomie und Freiheit, die in einem Spannungsverhältnis von Individual- und Gesamtnutzen stehen. Letzterer ist dabei nicht aggregativ, sondern als Menge der von allen Individuen geteilten Vorteile zu verstehen. Durch diese wird schließlich eine Fairness konstituiert, die sich dadurch auszeichnet, dass diejenigen, die profitieren, auch die Nachteile tragen. Das Postulat einer gerechten Verteilung beinhaltet also Fairnessansprüche (*fairness claims*). Der relative Charakter des Fairnessgedankens impliziert, dass etwas immer in Bezug auf etwas anderes fair ist. Die Idee der Fairness vollzieht sich quasi erst im interpersonellen Vergleich (vgl. Broome, 1984, S. 43). Bezogen auf Unfallalgorithmen bedeutet das: Es ist nicht nur relevant im Sinne des ersten Grenzkriteriums, dass die Risiken für die Einzelnen zumutbar sind, sondern auch, wie diese im Vergleich zu denen anderer Personen ausfallen. Ergänzend zu der Forderung zumutbarer Risiken lässt sich daher im Hinblick auf eine deontologische Risikoethik für Unfallalgorithmen folgendes (relatives) Prinzip formulieren:

- (2) *Risikoübertragungen im Kontext von Unfallalgorithmen können genau dann als moralisch zulässig gelten, wenn die aus der jeweiligen Risikokonstellation resultierenden Vor- und Nachteile unter Berücksichtigung aller ethisch relevanten Aspekte auf die Einzelnen fair verteilt sind.*

Traditionell beziehen sich gerechtigkeitsethische Fragen – und damit auch die Idee der Fairness – auf die Verteilung von Vorteilen und Lasten, die aus sozialer Kooperation resultieren. Die Verkehrsinfrastruktur ist ein öffentliches Gut, durch das Bürger ihr Recht auf Mobilität verwirklichen können (vgl. Dietrich, 2021, S. 728–729). In Dilemma-Szenarien bestehen relevante Vor- bzw. Nachteile in Bezug auf Sicherheit bzw. Nicht-Schädigung: Einen Vorteil erlangt ein Individuum dann, wenn eine Trajektorie gewählt wird, die es einem geringeren Risiko aussetzt als die anderen Beteiligten des Dilemmas. Risiken für Einzelne sind demnach so zu gestalten, dass sich daraus diejenige Verteilung von Vor- und Nachteilen ergibt, die als fair angesehen werden kann.

Unter welchen Voraussetzungen können risikoinduzierte Vor- bzw. Nachteile als fair gelten? Hinsichtlich philosophischer Konzeptionen der Gerechtigkeit verläuft eine scharfe Trennlinie zwischen egalitaristischen und non-egalitaristischen Ansätzen,<sup>274</sup> die jeweils unterschiedliche Aspekte der Gerechtigkeit akzentuieren. Gerechtigkeit definiert einen Zustand, in dem jedes Individuum das hat oder erhält, worauf es einen gültigen Anspruch hat; ein solcher kann beispielsweise über ein Recht begründet sein. Egalitaristische und non-egalitaristische Konzeptionen unterscheiden sich in der Begründung von Ansprüchen und Rechten. Egalitaristische Ansätze bestimmen Gerechtigkeit über eine relationale Interpretation von Gleichheit, die sich daran bemisst, wie andere im Vergleich zu dem betrachteten Individuum dastehen. Es sind zwei Arten egalitaristischer Gerechtigkeit zu unterscheiden. Gemäß der einen wird Gerechtigkeit mit

---

274 Zu den prominentesten Vertretern egalitaristischer Gerechtigkeitskonzeptionen gehören John Rawls (1971), Amartya Sen (1980), Ronald Dworkin (1981a, 1981b, 1987a, 1987b, 2000), Thomas Nagel (1979a, insb. Kap. 8; 1991) und Richard Arneson (1989). Non-egalitaristische Positionen, die sich zumeist als kritische Antwort auf egalitaristische Grundideen verstehen, werden hingegen von Harry Frankfurt (1997), Joseph Raz (1986, insb. Kap. 9) und Derek Parfit (2003) vertreten.

einem intrinsischen Wert der Gleichheit identifiziert, der um seiner selbst willen zu beachten ist. Daraus folgt, dass ungleiche Zustände mit normativen Implikationen verbunden sind (vgl. Anderson, 2000, S. 121). Während kontingente Ungleichheiten ausgeglichen werden müssen, sind selbstverschuldete bzw. leistungsabhängige hingegen akzeptabel.<sup>275</sup>

In diesem Zusammenhang wirken egalitäre Rechte, die als ›Rechte auf das Gleiche‹, z. B. gleiche Freiheit, gleiche Partizipation, zu verstehen sind; diese Form wird daher auch als egalitäre Gerechtigkeit bezeichnet. Die entsprechenden Rechte werden dabei non-egalitaristisch begründet, denn sie haben nur in Bezug auf ihren Inhalt bzw. die Größe des geforderten Gutes relationalen Charakter. Wichtig ist nicht nur, *dass* andere etwas bekommen, sondern *wie viel*. In einer anderen Konzeption egalitaristischer Gerechtigkeit wird Gleichheit über Ansprüche definiert, die aus zwei Komponenten bestehen: zum einen deshalb, weil andere etwas auch haben; zum anderen, weil ein Recht auf Gleichbehandlung und eine damit korrespondierende Pflicht besteht. Der Gleichheitsanspruch gründet sich auf die bestehende Gleichheit von Personen; diese wiederum folgt aus der Zugehörigkeit zu einem sozialen System, aus der sich normative Implikationen ergeben – verstanden als das Recht auf Gleichbehandlung in Bezug auf alle Ansprüche, die mit der Mitgliedschaft in diesem System in Zusammenhang stehen.

Im Gegensatz dazu geht der Non-Egalitarismus davon aus, dass Menschen von Natur aus ungleich sind. Da Gleichheit hier keinen intrinsischen Wert an sich darstellt, beinhaltende Ungleichheiten in diesem Fall keine normativen Implikationen (vgl. Nida-Rümelin, 2006, S. 11–13). Gerechtigkeitsansprüche sind unabhängig davon, ob andere etwas haben; sie sind suffizienzorientiert, d. h. darauf gerichtet von etwas *genug* zu haben. Ziel ist es daher, für jeden die (Minimal-)Bedingungen zu schaffen, die ein menschenwürdiges Leben erlauben (vgl. Margalit, 1996, 1997), wobei nicht notwendigerweise die Beseitigung von Ungleichheiten impliziert ist (vgl. Nida-Rümelin, 2006, S. 9–10).

Welche Gerechtigkeitskonzeptionen erscheinen für den Anwendungskontext von Unfallalgorithmen adäquat? Plausible Gründe

---

275 Diese Unterscheidung ist beispielsweise Grundlage der Konzeption von Dworkin (2000).

sprechen dafür, dass ein egalitaristischer Ansatz dem Anwendungsproblem am besten gerecht wird. Zum einen ist eine konzeptionelle Suffizienzorientierung, wie sie non-egalitaristische Konzeptionen vorsehen, für die spezifische Problematik von Dilemmata nicht zweckmäßig. Da es sich bei Dilemmata um Situationen mit sozialer Dynamik handelt, reicht es nicht aus, jedes Individuum nur isoliert zu sehen, denn Risikoübertragungen und daraus resultierende Vor- bzw. Nachteile sind entscheidend von den Interaktionen und Handlungen Einzelner abhängig. Weiterhin ist die Risikoethik naturgemäß relational, es sind stets Beziehungen zwischen Exponenten und Exponierten relevant; risikoethische Fragen lassen sich daher unter einer non-egalitaristischen Auffassung nur begrenzt diskutieren.

Dass die Manifestation kontingenter Ungleichheiten durch Praktiken algorithmischer Verteilungsstrategien kaum zu rechtfertigen ist, kann anhand einer einfachen Anwendung des Prinzips der Schadensminimierung veranschaulicht werden. Würden sich Unfallalgorithmen an der Maßgabe orientieren, möglichst geringe persönliche Schäden zu verursachen, bestünde die Gefahr, dass Personen geopfert werden, »um die eigentlichen Verursacher des Unfallen vor den Konsequenzen ihres Fehlverhaltens zu schützen.« (Hevelke & Nida-Rümelin, 2015c, S. 19) Dies wäre im Fall des Motorradfahrers in Beispielszenario 4 ›Motorradfahrer mit/ohne Helm‹ gegeben. Auch hinsichtlich bestimmter Fahrzeugtypen käme es zu diskriminierenden Effekten, wenn leichtere Fahrzeuge mit geringeren Sicherheitsstandards möglichst als Kollisionsobjekte vermieden und auf diese Weise sicherer für ihre Insassen werden. Derartige Mechanismen setzen systematisch Fehlanreize, sich durch nicht-regelkonformes Verhalten einen persönlichen Vorteil zu verschaffen. Dies erscheint uns intuitiv unfair; es ist nicht einsichtig, weshalb andere aufgrund des Fehlverhaltens von Einzelnen Nachteile in Kauf nehmen sollten. Dieses Beispiel pointiert, was für risikoethische Situationsbewertungen generell Gültigkeit besitzt: Verursacher von (höheren) Risiken sollten auch die (größeren) Nachteile tragen.

Zum anderen ist das Bestehen kontingenter Ungleichheiten bei Risikoübertragungen prinzipiell unvereinbar mit der Nicht-Verrechenbarkeit incommensurabler Werte, wie sie im Rahmen der metaethischen Analyse in Kap. 5 begründet wurde. Die Gleichheit und der sich daraus ableitende Anspruch auf Gleichbehandlung eines jeden Verkehrsbeteiligten beruhen auf der intrinsischen Unvergleich-

barkeit des Individuums und nicht auf der Zugehörigkeit zu einem (sozialen) System. Ein solches lässt sich für den Anwendungskontext Verkehr nicht plausibel denken, denn es existiert keine spezifische Gruppendynamik und keine ethisch relevante Abgrenzung zu Personen außerhalb des Verkehrsgeschehens.

Daraus folgt, dass egalitäre Gerechtigkeitskonzeptionen im Hinblick auf die Problemstellung am besten geeignet sind. Ein anwendungsnahe Argument stützt diese These: Auch wenn Gleichheit im Sinne gleicher Rechte »auf das Gleiche« das erklärte Ziel egalitärer Gerechtigkeit ist, so lässt sich daraus nicht schließen, dass jede Form der Ungleichheit zwangsläufig ungerecht ist. Wie Nida-Rümelin et al. (2012, S. 141) konstatieren, kann auch eine Ungleichverteilung fair sein, nämlich genau dann, wenn Ansprüche auf bestimmte Vorteile anderen gegenüber bestehen.

Die entscheidende Frage ist, wie sich solche Ansprüche begründen lassen. Zuvor wurde anhand des Prinzips der Schadensminimierung erläutert, dass ein spezifisches Design von Unfallalgorithmen genau dann zulässig ist, wenn dadurch das Risiko jedes Einzelnen in gleichem Maße reduziert wird und es daher im Interesse jedes Einzelnen ist. Hevelke und Nida-Rümelin (2015c, S. 17) zeigen, dass Letzteres jedoch nicht den entscheidenden Punkt trifft; für die Zulässigkeit ist vielmehr die Erfüllung individueller, begründeter Ansprüche maßgeblich. Konsequenterweise müssen nicht nur Benachteiligungen ausgeglichen werden, sondern auch ungerechtfertigte Bevorteilungen. Gewisse Gruppen müssten im Zuge eines Benachteiligungsausgleichs ihre Sicherheitsvorteile aufgeben, deren Verlust jedoch genau dann unproblematisch im Sinne einer Instrumentalisierung der entsprechenden Individuen ist, wenn sie auf diese Vorteile gar keinen moralisch begründbaren Anspruch haben. Ein solcher ist insbesondere dann nicht vorhanden, wenn diese Vorteile im Gegenzug mit Nachteilen für andere verbunden sind. Ein Ausgleich in diesem Sinne ungerechtfertigter Vorteile kann daher mindestens als zulässig, möglicherweise sogar als wünschenswert angesehen werden.

Aus dem Postulat egalitärer Gerechtigkeit folgt somit im Hinblick auf das Ziel einer gerechten Ordnung die normative Forderung, bestehende kontingente Ungleichheiten auszugleichen. Dabei gilt, dass sich Vor- und Nachteile ausschließlich über berechtigte Ansprüche legitimieren lassen. Hier schließt sich nun die Frage an, inwiefern

berechtigte Ansprüche im Kontext von Unfalldilemmata bestehen und wie sich diese begründen lassen. Dies wird im Folgenden anhand von zwei ausgewählten Aspekten erörtert.

### 7.3.3.2 Risikoverursachung und Vulnerabilität als Begründungsversuche berechtigter Vorteilsansprüche

Vor dem Hintergrund des Fairnesspostulats sind Sicherheitsvorteile nur begründbar, wenn berechtigte Ansprüche auf eine Bevorzugung vorliegen. Eines der am häufigsten diskutierten Verteilungsprinzipien in dieser Hinsicht ist jenes, welches die Sicherheit der Insassen autonomer Fahrzeuge im Fall einer Kollision priorisiert. Das zentrale Argument, das in einschlägigen Diskussionen vorgebracht wird, lautet, dass autonome Fahrzeuge im Vergleich zum herkömmlichen motorisierten Verkehr die Verkehrssicherheit erhöhen und daher ihren Insassen ein Vorteil gegenüber den Insassen herkömmlicher Fahrzeuge im Mischverkehr zugesprochen werden sollte. Berkey (2022, S. 217–219) hebt hervor, dass dies allerdings in mancherlei Hinsicht unplausibel ist: Erstens wäre eine solche Schlussfolgerung nur unter der Einschränkung denkbar, dass alle Individuen über (annähernd) gleiche Zugangschancen zur automatisierten Mobilität verfügen, was praktisch jedoch aufgrund unterschiedlicher Wohlstands niveaus einerseits und der hohen Preisklasse autonomer Fahrzeuge andererseits nicht gegeben ist. Zweitens widerspräche eine generelle Programmierung auf Insassenpriorisierung dem Prinzip der Schadensminimierung. Die Algorithmen würden in einer Weise verzerrt, die Anreize für die Wahl von Trajektorien setzt, welche jegliche noch so geringe Schäden von Insassen autonomer Fahrzeuge abwenden und daher in überproportional hohen Schäden für Insassen herkömmlicher Fahrzeuge resultieren. Drittens würde auf diese Weise bereits die Möglichkeit einer fairen Verteilung untergraben. Da potenzielle Nutzer sich tendenziell eine Priorisierung ihrer eigenen Sicherheit wünschen (vgl. Bonnefon et al., 2015, S. 5–8), existieren hohe Anreize für Hersteller, ihre Algorithmen dementsprechend zu gestalten. Faire Verteilungen von Vor- und Nachteilen sind daher ein Problem kollektiven Handelns: Sobald ein einziger Hersteller Fahrzeuge anbietet, welche die Nutzersicherheit priorisieren, würden aufgrund des Marktdrucks alle anderen zur Anpassung

ihrer Produkte genötigt, wodurch sämtliche Fairnessbestrebungen hinfällig würden (vgl. Berkey, 2022, S. 222–223).

Auch wenn sich ein Anspruch auf Vorteile für die Insassen autonomer Fahrzeuge nicht glaubwürdig rechtfertigen lässt, bleibt der Grundgedanke, dass das Ausmaß einer Risikoverursachung ethisch relevant ist, ein prinzipiell valides Argument. Es ist offensichtlich, dass nicht-motorisierte Verkehrsteilnehmer wie Radfahrer oder Fußgänger verhältnismäßig geringere Risiken für andere herbeiführen, sodass es gute Gründe gibt, von einer ursprünglich angenommenen Gleichverteilung zu ihren Gunsten abzuweichen. Aus der geringeren Risikoübertragung lässt sich durchaus ein legitimer Anspruch auf Sicherheitsvorteile ableiten, was als Umkehrschluss der eingangs in diesem Unterkapitel festgestellten Intuition aufgefasst werden kann: Nicht nur soll gelten, dass, wer höhere Risiken bewirkt, mehr Nachteile in Kauf nehmen muss, sondern auch, dass, wer weniger Risiken verursacht, weniger Nachteile tragen sollte:

[...] it seems to me that a potentially significant deviation from an equal distribution of risks between, on the one hand, occupants of vehicles, and on the other, pedestrians and cyclists, might be required. This is because while pedestrians do not expose others to any risks at all just in virtue of using the road (or sidewalk), and cyclists expose others only to relatively small and (on average) less serious risks, those who choose to ride in vehicles expose others to (much more) substantial risks of harm, including risks of serious injury and death, just in virtue of their use of vehicles. It seems plausible that those who choose to introduce risks of this kind in order to enjoy the benefits of the activities that unavoidably involve these risks, should, where possible, at least bear a greater share of the risks than those who are not engaged in the activities that impose them. (Ebd., S. 217)

Die Prämissen, die dieser Argumentation zugrunde liegt, hat allerdings nur so lange Bestand, wie sich die betreffenden Personen an gewisse Regeln halten. Es ist einleuchtend, dass durch regelwidriges Verhalten höhere Risiken produziert werden.<sup>276</sup> Verkehrsteilnehmer müssen plausiblerweise davon ausgehen können, dass sich alle regelkonform verhalten; Hevelke und Nida-Rümelin (2015a, S. 222–224) betrachten die Vorhersehbarkeit des Verhaltens im Straßenverkehr als ausschlaggebend dafür, dass einer bestimmten Person Schutz

---

276 Siehe hierzu auch die bisher diskutierten Argumente in Kap. 4.4.1.

gewährt werden kann. Abney (2022, S. 264–272) legt dar, dass Forderungen nach einer Gleichverteilung häufig auf der Annahme einer grundsätzlichen moralischen Guttheit (*moral goodness*) aller Personen basieren. Wird diese angesichts praktischer Evidenz aufgegeben, so wird deutlich, dass moralische Urteile und damit auch Argumente in Bezug auf bestimmte risikoethische Verteilungsmechanismen wesentlich von den Motiven und Intentionen von Individuen abhängen:

If we accept that moral judgments will crucially depend on the intentions of the agents involved, then people with immoral purposes will morally require a different response from those with moral purposes—and this is hardly a novel concept in law and policy; there are commonplace ethical and legal distinctions between mere negligence versus intentional abuse. (Ebd., S. 271–272)

Anstatt über individuelles Fehlverhalten hinwegzusehen, muss dieses zur Verantwortung gezogen werden, wenn man die Idee der Gerechtigkeit ernst nimmt. Hier muss stets im Einzelfall geprüft werden, wie ein Regelbruch konkret zu bewerten ist:

If we take justice seriously, we cannot assume the moral equality of all persons in calculating consequences for moral decision-making. Some people deserve to be treated worse than others. In other words: some people, in some circumstances, deserve to be hit. In fact, they deserve to die – regardless of utility. (Ebd., S. 265)

Einen weiteren, ebenfalls häufig thematisierten Aspekt in diesem Kontext stellt die Gewährung von Sicherheitsvorteilen auf der Basis einer Kategorisierung von Verkehrsbeteiligten dar. Dieses Vorgehen entspricht der Konzeption von Theorien sozialer Gerechtigkeit, die neben dem zu verteilenden Gut, deren Empfängern und dem formalen Verteilungsprinzip auch die Eigenschaften der Empfänger, welche für die Verteilung relevant sind, als eine Hauptdimension gerechter Verteilungen begreifen (vgl. Dietrich & Weisswange, 2019, S. 229–230). Wie die Auswertung statistischer Daten zeigt, werden bestimmte Gruppen von Verkehrsteilnehmern unverhältnismäßig oft in tödliche Verkehrsunfälle mit Fahrzeugen verwickelt (vgl. Dietrich, 2021, S. 728–729; Mullen et al., 2014, S. 238). Dies ist auf diverse Gründe zurückzuführen, von denen für den Kontext dieser Forschungsarbeit vor allem zwei relevant sind: zum einen auf ein häufig besonders rücksichtsloses Verhalten von Autofahrern gegenüber

Radfahrern (vgl. Europäische Kommission, 2020, S. 31) und zum anderen darauf, dass gewisse Gruppen durch eine erhöhte Schadensneigung charakterisiert sind. So ist sich die Forschungsliteratur dahingehend einig, dass der motorisierte Verkehr gegenüber Fußgängern generell begünstigt ist (vgl. Dietrich, 2021, S. 727–728).<sup>277</sup>

Dies führt dazu, dass der erwartete Schaden für Fußgänger und Radfahrer im Fall einer Kollision mit einem Fahrzeug deutlich höher ausfällt als für die Fahrzeuginsassen. Ursächlich dafür sind primär physikalische Faktoren wie die deutlich geringere Masse, die geringere Bewegungsgeschwindigkeit und nicht vorhandene Schutzworrichtungen wie Karosserie oder Rückhalteeinrichtungen. Es ist also plausibel anzunehmen, dass zwischen Personengruppen ungleiche Ausgangsbedingungen hinsichtlich der Möglichkeiten herrschen, ihre Ansprüche auf Schutz von Leben und Gesundheit zu verwirklichen (vgl. Luetge, 2017, S. 552–553).<sup>278</sup> Im Kontext automatisierter Mobilität verschärft sich diese Problematik ungleicher Risikoexpositionen zusätzlich:

The most severe concerns are that AVs will reinforce existing inequalities and introduce new injustice aspects in the context of road traffic. An existing inequality is the unequal exposure to risks of bodily harm induced by a traffic accident between different types of road users. (Dietrich, 2021, S. 727)

Ein weit verbreiteter Terminus des einschlägigen ethischen Diskurses ist der Begriff der *Vulnerable Road Users (VRUs)* (= gefährdete Verkehrsteilnehmer)<sup>279</sup>, der als Sammelbegriff dient für »non-motorised road users, such as pedestrians and cyclists as well as motorcyclists and persons with disabilities or reduced mobility and orientation.« (European Commission Mobility und Transport, 2023) Ethisch-politische Direktiven und Empfehlungen verweisen auf die

---

277 Dies führt Dietrich (2021, S.728) auf den Umstand zurück, dass die Infrastruktur der Verkehrsstraßen in urbanen Gesellschaften motorisierten Verkehr begünstigt; siehe auch Nello-Deakin (2019, S. 699–704).

278 Auch innerhalb der Gruppe der Fahrzeuge bestehen gewichtige Unterschiede hinsichtlich möglicher Risikoexpositionen aufgrund von unterschiedlicher Masse und aktiven sowie passiven Schutzworrichtungen (vgl. Dietrich, 2021, S. 728).

279 In einschlägigen Publikationen findet sich das Akronym *VRU* sowohl mit als auch ohne Pluralendung „-s“. Aus Gründen der Einheitlichkeit wird in diesem Buch die Pluralendung verwendet.

Notwendigkeit einer Kategorisierung von Verkehrsteilnehmern, um das Sicherheitsniveau vulnerabler Gruppen zu heben. Die deutsche Ethik-Kommission beispielsweise versteht eine »Programmierung der Fahrzeuge zu defensivem und vorausschauendem, schwächere Verkehrsteilnehmer [...] schonendem Fahren« (Di Fabio et al., 2017, S. 10) als Teil einer übergeordneten Strategie zur Erhöhung der Verkehrssicherheit durch automatisierte Mobilität. Die Expertengruppe der Europäischen Kommission (2020, S. 30–32) betrachtet den Schutz von VRUs als Element einer Solidarität, die sich als Ergänzung zu gerechtigkeitsethischen Anforderungen ergibt. In diesem Sinne ist es Teil der Designaufgabe autonomer Fahrzeuge, VRUs besonders zu schützen: »CAVs should, among other things, adapt their behaviour around vulnerable road users instead of expecting these users to adapt to the (new) dangers of the road.« (Ebd., S. 7)

Ethische Begründungsgrundlagen werden im Rahmen politischer Richtlinien meist nur angedeutet. Die Expertengruppe der Europäischen Kommission nennt die Verringerung der Geschwindigkeit sobald VRUs erkannt werden und das Einhalten größerer Abstände zu diesen als mögliche praktische Maßnahmen, um VRUs zu schützen. Sie sollen allerdings nur solange in Kraft bleiben, wie sich der Gesamtschaden für andere Verkehrsteilnehmer nicht erhöht (vgl. ebd., S. 31). Wie in Fällen verfahren werden soll, in denen diese Voraussetzung nicht erfüllt ist, wird hingegen offengelassen.

Auch in der ethischen Forschungsliteratur bezüglich des Designs von Unfallalgorithmen werden Konzepte thematisiert, gefährdete Gruppen im verkehrlichen Umfeld besonders zu schützen. Der Schutz von VRUs wird dabei als Postulat egalitärer Gerechtigkeit gedeutet, das den Ausgleich der Ungleichheiten beinhaltet, die durch bestehende Unterschiede bei der Schadensneigung hervorgerufen werden. Wichtig ist zu beachten, dass den Interessen von VRUs durch den besonderen Schutz nicht mehr Gewicht gegeben, sondern lediglich ihrem legitimen Anspruch auf Ausgleich entsprochen wird:<sup>280</sup> »This programming would not amount to giving greater value to the safety of cyclists—it would rather be an attempt to correct safety inequalities, which partly result from the current behaviour of human drivers.« (Ebd., S. 31) Martínez-Buelvas et al. (2022) stellen

---

280 Dieses Argument lässt sich als Antwort auf Keelings (2018a) Kritik verstehen, dass die Ansprüche der Schlechtergestellten übermäßig gewichtet würden.

mögliche Strategien für die Gestaltung einer sicheren und gerechten Interaktion zwischen autonomen Fahrzeugen und VRUs vor, aus denen sich eine Forschungsgenda ableiten lässt. Dabei interpretieren sie Verkehrsgerechtigkeit in Bezug auf Gleichheit, Fairness und Zugang.

Auf Basis der jeweiligen Vulnerabilität ergeben sich also berechtigte Ansprüche auf Sicherheitsvorteile für bestimmte Gruppen. Nun sind allerdings nicht nur physikalische Faktoren dafür verantwortlich, wie hoch der individuelle Schaden im Fall einer Kollision ausfällt. Auch persönliche Merkmale wie das Alter, die Fitness oder der Gesundheitszustand betreffender Personen haben unter bestimmten Umständen Einfluss darauf, ob eine Person beispielsweise tödlich verletzt wird. Sollen derartige Aspekte im Rahmen von Unfallalgorithmen berücksichtigt werden, führt dies zu ernstzunehmenden ethischen Problemen. Im Kontext von Algorithmen und Künstlicher Intelligenz wird der Begriff der Fairness häufig mit der Freiheit von systematischen Verzerrungs- (*biases*) und Diskriminierungseffekten assoziiert. Der Ausschluss jeglicher ungerechtfertigten Bevorzugung bzw. Benachteiligung bestimmter Personengruppen in automatisierten Entscheidungsprozessen gilt als eine der Grundanforderungen an ein robustes Systemdesign.<sup>281</sup> Inwiefern Unfallalgorithmen – und algorithmische Entscheidungssysteme im Allgemeinen – in konkreten Anwendungsfällen diskriminierend sind, ist eine häufig thematisierte, kontroverse Fragestellung. Dies ist u. a. darauf zurückzuführen, dass einerseits die Definitionen dessen, was unter problematischer Diskriminierung verstanden wird, sehr vage sind.

Andererseits kann Diskriminierung in Entscheidungen maschinell trainierter Algorithmen auf verschiedene, oftmals subtile Weisen erzeugt werden (vgl. Zweig, 2019, Kap. 8). Eine anwendungsorientierte Sichtweise im Kontext autonomer Fahrsysteme legt Leben (2022, S. 132–138) vor. Demnach sind diskriminierende Handlungen nicht per se moralisch schlecht, sondern nur dann, wenn sie Schaden verursachen (vgl. Thomsen, 2023) bzw. wenn sie ungerechtfertigt sind. Letzteres ist genau dann der Fall, wenn einzelne Personen oder Gruppen aufgrund von Faktoren benachteiligt werden, die in Bezug auf die Aufgaben, die der Algorithmus erfüllen soll, unerheb-

---

281 Für eine Übersicht zum gegenwärtigen Forschungsstand diesbezüglich siehe Hütt und Schubert (2020).

lich sind. In anderen Worten: Ob Algorithmen in ungerechtfertigter Weise diskriminierend sind, hängt davon ab, welchem praktischen Zweck sie dienen. So besteht die primäre Aufgabe von Unfallalgorithmen in der Verteilung von Schäden; Faktoren, die für diese Zielsetzung als irrelevant gelten, sind beispielsweise sozialer Status oder Berufsgruppenzugehörigkeit. Diese sind als Grundlage für Entscheidungsalgorithmen nicht gerechtfertigt und würden daher berechtigterweise als diskriminierend angesehen. Jedoch sind auch Faktoren denkbar, die im Hinblick auf die aus einem Unfall resultierenden Personenschäden durchaus eine Relevanz besitzen. Statistisch gesehen besteht eine Korrelation zwischen bestimmten persönlichen Merkmalen wie dem Alter der Unfallopfer einerseits und deren physischer Vulnerabilität bzw. Schadensneigung andererseits; Alter konstituiert also Ungleichheiten im Sinne einer erhöhten Vulnerabilität. Gemäß Lebens Auffassung wäre demnach die Berücksichtigung des Alters als Entscheidungskriterium durchaus zu rechtfertigen.

Ein anderes hier relevantes Argument liefert Černý (2022, S. 34–38). Ihm zufolge ist ein Algorithmus, der sich in einem Dilemma für diejenige Trajektorie entscheidet, die eine jüngere Person bevorzugt, aus zwei Gründen nicht notwendigerweise diskriminierend. Zum einen ist das Vorliegen diskriminierender Handlungen an deren Unausweichlichkeit geknüpft: Ist es möglich, auch anders zu handeln? In Dilemmata ist dies naturgemäß nicht erfüllt, negative Konsequenzen sind unvermeidbar. Zum anderen ist das Alter nicht als Kriterium zu verstehen, aufgrund dessen entschieden wird, sondern lediglich als Variable einer Theorie, die Černý als »deprivation theory of the badness of death« (ebd., S. 34–35) bezeichnet. Deren Grundlage ist einerseits die These, dass der Tod auch gut für eine Person sein kann, nämlich genau dann, wenn er diese von einem schlechten Leben erlöst, und andererseits die Annahme, dass mit zunehmendem Alter die Lebensqualität abnimmt. Daraus leitet Černý die normative Implikation ab, dass es moralisch schlechter ist, eine jüngere Person zu töten:

When the AV decides between a and b, it is deciding based on age. But—and this is crucial—the age is not in fact the criterion of the choice, but merely a variable determining the extent of the badness of death. In other words, the AV decides by means of age, but not based on age: the basis of its choice is comparing the badness of death between a and b. (Ebd., S. 37)

An Černýs Ansatz lässt sich kritisieren, dass er Konflikte mit grundlegenden Normen wie der Unvergleichbarkeit menschlichen Lebens aufwirft. Sowohl Lebens als auch Černýs Ansatz liegt der Gedanke zugrunde, dass es bei Entscheidungen, die *mithilfe* des Merkmals »Alter« getroffen werden, nicht um eine Bevorzugung *aufgrund* des Alters an sich geht, sondern um die Beurteilung dessen, was in direktem Zusammenhang damit steht. Werden durch das Alter kontingente Ungleichheiten produziert – wie es im Hinblick auf die individuelle Vulnerabilität der Fall ist –, so entstehen daraus begründete Ansprüche auf Sicherheitsvorteile, die nur indirekt durch das Alter, vielmehr aber durch die daraus resultierende Vulnerabilität verursacht sind.<sup>282</sup>

Zusammenfassend ergeben sich sowohl aus einer Bewertung der Risikoverursachung als auch der individuellen Vulnerabilität bedeutende normative Implikationen, die in einem berechtigten Anspruch auf Sicherheitsvorteile bzw. eine Priorisierung nicht-motorisierter Verkehrsteilnehmer münden. Eine Gleichverteilung von Risiken ist hier nicht plausibel; Ziel ist es vielmehr, bestehende kontingente Ungleichheiten im Sinne einer ergebnisegalitaristischen Position fair zu berücksichtigen, welche die Erreichung eines Pareto-optimalen Zustands anvisiert.<sup>283</sup> Ungeachtet ihrer ethischen Relevanz stellen sich hinsichtlich der praktischen Implementierung Herausforderungen auf technischer Seite. Die Fairness risikoethischer Verteilungsstrategien hängt in letzter Konsequenz auch von der technischen Leistungs- und Perzeptionsfähigkeit der jeweiligen Systeme ab. Eine korrekte und zuverlässige perzeptive Differenzierung zwischen Kategorien von Verkehrsteilnehmern<sup>284</sup> ist Grundvoraussetzung dafür, dass VRUs ein besonderer Schutz zuteilwerden kann (vgl. Keeling, 2022, S. 47–53):

In all these examples, the recommendation to provide greater road safety to a subset of road users must always be premised on evidence

---

282 Auch in Bezug auf den Faktor »Geschlecht« bestehen möglicherweise derartige Ungleichheiten; dies ist allerdings noch nicht tiefergehend untersucht worden.

283 Der Egalitarismus kann generell auf drei Arten verstanden werden (vgl. Horn, 2003, S. 26): Verteilungsegalitarismus (gleich große Güter), Verfahrensegalitarismus (gleichmäßige Regelanwendung), Ergebnisegalitarismus (gleiche Ergebnisse).

284 Eine noch größere technische Hürde ist es freilich, die Intentionen der Handelnden korrekt zu erkennen.

that it is technically possible for a CAV to detect and respond to these road users accurately and reliably, that some users' harm-to-exposure ratio is high, that improving road safety for one subset of road users does not raise the total harm inflicted to another category of road users above its current baseline. (Europäische Kommission, 2020, S. 32)

Im Folgenden werden zwei bedeutende Konzeptionen für egalitäre risikoethische Verteilungsstrategien rekonstruiert und kritisch beleuchtet.

#### 7.3.3.3 John Rawls: Eine prioritaristische Antwort auf Ungleichheiten

Philosophische Gerechtigkeitskonzeptionen der neueren Zeit sind vornehmlich kontraktualistisch orientiert. Ankerpunkt ist John Rawls' vertragstheoretischer Ansatz, den er 1971 in seinem epochalen Werk *A Theory of Justice* vorgelegt hat und der dem bisherigen Gerechtigkeitsdiskurs eine neue Wendung fort von vornehmlich utilitaristischen Denkstrukturen gab.<sup>285</sup> Rawls begründet seine liberale Gerechtigkeitskonzeption in der Tradition hypothetischer Vertragstheorie und basierend auf dem kantianischen Verständnis von Freiheit, aus der Gleichheit (in politischer Hinsicht) folgt.<sup>286</sup> Die kantianisch geprägte Konzeption individueller Autonomie ist nicht nur mit einem Freiheits-, sondern auch mit einem Gleichheitsanspruch verbunden:

Mit dem Freiheitsanspruch wird gefordert, dass jede Person sich idealiter zugleich als Autor der Gesetze bzw. der Maximen anderer Akteure ansehen können sollte, denen sie sich ausgesetzt sieht. Mit dem Gleichheitsanspruch wiederum wird gefordert, dass sich die Handlungsbeschränkungen, die sich aus dem Freiheitsanspruch ergeben, für das Handeln einer jeden Person gegenüber allen anderen Personen gelten; dies bedeutet, dass insbesondere die situative Ungleichbehandlung einzelner Personen die Möglichkeit des Ansehens aller Personen als Gleiche nicht grundlegend erschüttern darf. (Nida-Rümelin et al., 2012, S. 181)

---

285 Für eine Übersicht zu zeitgenössischen Ansätzen sozialer Gerechtigkeit siehe Pojman (2005).

286 Ebenso wie Rawls verteidigt auch Nida-Rümelin (2006) eine egalitaristische Position, die Gleichheit im Sinne der kantianischen Verknüpfung von Freiheit und Gleichheit versteht; Freiheit ist ohne Gleichheit nicht denkbar.

Gemäß der Rawls'schen Auffassung stellt sich Gerechtigkeit als Fairness dar, die dadurch konstituiert wird, dass Akteure in einer hypothetischen Entscheidungssituation hinter einem *Schleier des Nichtwissens* unvoreingenommen über die Grundprinzipien einer gerechten Ordnung beraten. Unter den Prämissen der Gleichheit aller Personen, knapper Ressourcen und überwiegend egoistischer Motivationen würden sie sich für zwei lexikalisch angeordnete Grundsätze entscheiden: Der erste bezieht sich auf die fundamentale Grundstruktur einer Gesellschaft und postuliert eine gleiche Verteilung individueller Freiheitsrechte. Der zweite widmet sich sozio-ökonomischen und ethischen Fragen, indem er faire Chancen beim Zugang zu öffentlichen Positionen fordert. In vorläufiger Form lauten die beiden Prinzipien folgendermaßen:

First: each person is to have an equal right to the most extensive basic liberty compatible with a similar liberty for others. Second: social and economic inequalities are to be arranged to that they are both (a) reasonably expected to be to everyone's advantage, and (b) attached to positions and offices open to all. (Rawls, 1971, S. 60)

Im Rahmen seiner Diskussion des zweiten Grundsatzes formuliert Rawls schließlich das berühmte *Differenzprinzip (difference principle)*, das den Schlechtergestellten einer Gesellschaft im Hinblick auf Verteilungsfragen von Primärgütern Priorität einräumt:

Social and economic inequalities are to be arranged so that they are both (a) to the greatest benefit of the least advantaged and (b) attached to offices and positions open to all under conditions of fair equality of opportunity. (Ebd., S. 83)

Daraus folgt, dass Ungleichheiten hier nicht per se schlecht sind, sondern genau dann akzeptabel, wenn sie den Schlechtergestellten zu einem Zustand verhelfen, der sie im Vergleich zu alternativen Güterverteilungen besserstellt. Vor diesem Hintergrund lassen sich die beiden Grundsätze wie folgt reformulieren:

*First Principle[:]* Each person is to have an equal right to the most extensive total system of equal basic liberties compatible with a similar system of liberty for all. *Second Principle[:]* Social and economic inequalities are to be arranged so that they are both: (a) to the greatest benefit of the least advantaged, consistent with the just savings principle, and (b) attached to offices and positions open to all under conditions of fair equality of opportunity. (Ebd., S. 302, Hervorh. i. Orig.)

Wichtig ist anzuerkennen, dass das *Differenzprinzip* seinem Wesen nach ein Konzept gegenseitigen Vorteils ist; Ungleichheiten sind genau dann gerechtfertigt, wenn sie zum Vorteil aller dienen. Es begründet damit eine Position, die auf einer egalitären Gerechtigkeitsauffassung<sup>287</sup> fußt (vgl. Nida-Rümelin & Rechenauer, 2009, S. 304) und dabei einem prioritaristischen Grundsatz folgt:

All social primary goods—liberty and opportunity, income and wealth, and the bases of self-respect—are to be distributed equally unless an unequal distribution of any or all of these goods is to the advantage of the least favored. (Rawls, 1971, S. 303)

Die Mehrheit der Ansätze, die ausgehend von den normativen Implikationen, die sich aus den berechtigen Ansprüchen von VRUs ergeben, im Rahmen des Forschungsdiskurses entwickelt wurden, konzipieren risikobezogene Verteilungsfragen für Unfallalgorithmen auf der Basis eines Rawls'schen Prioritarismus. Vor allem in der implementierungsnahen risikoethischen Literatur findet ein von Rawls inspiriertes Design von Unfallalgorithmen Zustimmung. Übertragen auf den Kontext von Unfalldilemmata impliziert das *Differenzprinzip* spezifische Fahrentscheidungen: »[...] it will result in driving decisions in which the worst-off vehicles profit the most and it is prevented that high risk situations for certain vehicles are accepted for the benefit of all other vehicles.« (Dietrich & Weisswange, 2019, S. 233) Zudem wird betont, dass eine Priorisierung von VRUs als derjenigen Gruppe von Verkehrsteilnehmern, welche hinsichtlich ihres Schadenserwartungswerts im Rawls'schen Sinne als die am schlechtesten Gestellten gelten können, sich positiv auf das Verhalten von autonomen Fahrzeugen gegenüber VRUs auswirkt:<sup>288</sup>

It seems to be more appropriate, to always give full priority to the entity which has the highest expected harm. To formalize this, the severity of a potential encounter between two or more entities can be calculated by predicting the severity for each of the entities separately and then selecting the maximum value. In this way, the entity with highest risk has a

---

287 Siehe Arneson (1999) für eine Übersicht zu möglichen Begründungen egalitären Gerechtigkeitskonzeptionen, die von bessergestellten Individuen verlangen, den schlechtergestellten zu helfen.

288 Dietrich und Weisswange (2019, S. 233) merken dazu an, dass eine Berücksichtigung der Interessen der Schwächsten auch die Wahrscheinlichkeit von Diskriminierung reduziert.

strong effect on the risk-based behavior planning—the consequence is that cautious behaviors around vulnerable road users will become more likely. (Dietrich, 2021, S. 735)

Auf der formalen Ebene lässt sich das *Differenzprinzip* mithilfe der *Maximin*-Regel ausdrücken; das Ergebnis ist Pareto-optimal, d. h. es existiert keine andere Schadensverteilung, die ein Individuum besserstellt, ohne gleichzeitig mindestens ein anderes schlechterzustellen. Der prominente, von Leben (2017) entwickelte formale Algorithmus basiert auf einer Schätzung der Überlebenswahrscheinlichkeit für jedes involvierte Individuum bei jeder möglichen Trajektorie und der Auswahl derjenigen, bei der die Schwächsten mit der höchsten Wahrscheinlichkeit überleben. Anstelle der einfachen *Maximin*-Regel verwendet er eine iterierte Variante, das sogenannte *Leximin*-Prinzip.<sup>289</sup> Dieses unterscheidet sich vom *Maximin*-Prinzip dahingehend, dass für den Fall, dass zwei Alternativen zum gleichen Ergebnis führen, die Wahrscheinlichkeit desjenigen Individuums den Ausschlag gibt, das am zweitschlechtesten gestellt ist. So soll verhindert werden, dass die alleinige Orientierung an demjenigen Individuum, welches am schlechtesten gestellt ist, zur Wahl einer insgesamt schlechteren Verteilung führt.

Lebens Entwurf ist im Forschungsdiskurs kontrovers rezipiert worden. Eine prominente Kritik stammt von Keeling (2018a, S. 264–270), der einwendet, dass Lebens Ansatz auf der Basis Rawls'scher Gründe nur schwer zu rechtfertigen ist und einer unabhängigen Begründung bedarf. Er stellt u. a. klar, dass die Individuen bei Rawls nicht zwischen konkreten Güterverteilungen anhand des *Maximin*-Prinzips wählen, sondern zwischen Gerechtigkeitsprinzipien, wobei die Wahl aus rationalen Gründen auf die *Maximin*-Regel fällt. In Lebens Konzeption hingegen wählen die Individuen zwischen konkreten Verteilungen mithilfe des bereits akzeptierten *Maximin*-Prinzips. Zudem legt Leben den Individuen in der Ausgangsposition des *Schleiers des Nichtwissens* Restriktionen hinsichtlich der verfügbaren Informationen auf, die aus Rawls'scher Sicht nicht gerechtfertigt sind.

---

289 Das *Leximin*-Prinzip stammt aus der Wohlfahrtsökonomik und wurde von Amartya Sen (1976, 1980) in Anlehnung an das Rawls'sche *Differenzprinzip* entworfen, um eine etablierte Konvention innerhalb der Sozialwahltheorie zu beschreiben.

Ferner ist die an Rawls angelehnte Konzeption aufgrund der Spezifika des Anwendungskontextes in mancherlei pragmatischer Hinsicht ethisch problematisch. Für das Entscheidungsproblem, welches sich im Kontext von Unfallalgorithmen stellt, gilt, dass die Identität der betroffenen Parteien zum Zeitpunkt der Implementierung unbekannt ist. Gemäß Hevelke und Nida-Rümelin (2015c, S. 11–14) folgt aus dieser fehlenden Determiniertheit der Identität potenzieller Opfer, dass eine auf Schadensminderung ausgerichtete Programmierung nicht grundsätzlich unvereinbar mit einer deontologischen Ethik sein muss. Denn auch wenn Schadensminimierung letztlich immer ein Abwägen erfordert, muss dieses nicht zwangsläufig ethisch problematisch sein. Wenn die konkreten Vor- und Nachteile einer Person nicht bestimmbar sind, weil schon ihre Identität nicht bestimmbar ist, so lässt sich lediglich untersuchen, welche Handlung im Sinne aller wäre. Hevelke und Nida-Rümelin (2015c, S. 11) folgern daher, »dass eine auf die Minimierung der Opfer ausgelegte Programmierung durchaus im Interesse jedes Einzelnen sein kann – nämlich genau dann, wenn diese Programmierung das Risiko eines jeden Einzelnen reduziert bzw. minimiert.« Dadurch, dass es grundsätzlich jeden treffen könnte, würde eine Programmierung auf Schadensminimierung niemanden im kantischen Sinne bloß als Mittel gebrauchen. Analog konstatiert die Ethik-Kommission des BMVI in ihrem Bericht, dass eine solche Programmierung keinen Verstoß gegen Art. 1 Abs. 1 GG darstellen würde, sofern die Voraussetzung erfüllt ist, dass »die Programmierung das Risiko eines jeden einzelnen Verkehrsteilnehmers in gleichem Maße reduziert. Solange nämlich die vorherige Programmierung für alle die Risiken in gleicher Weise minimiert, war sie auch im Interesse der Geopferten, bevor sie situativ als solche identifizierbar waren.« (Di Fabio et al., 2017, S. 18)

Diese Konstellation wird in der Literatur häufig als heuristisch interpretierte Instanz des Rawls'schen *Schleier des Nichtwissens* diskutiert (siehe auch Kap. 4.4.4.2). Dadurch, dass zum Zeitpunkt des Designs von Unfallalgorithmen die Identität späterer Opfer unbekannt ist, werden die Individuen der Möglichkeit beraubt, sich selbst taktische Vorteile zu verschaffen, indem sie sozialschädliche Regeln installieren. So ist es unter der Annahme weitgehend egoistischer Individuen plausibel zu erwarten, dass diese sich für faire Verteilungsprinzipien entscheiden würden, um auf diese Weise den

Nutzen aller – und damit auch ihren eigenen – zu steigern. Wenn jede Person jede mögliche Rolle in einem Dilemma mit gleicher Wahrscheinlichkeit einnimmt, dann verhalten sich rationale Individuen risikoavers (vgl. Rath, 2011, S. 77) und entscheiden sich für eine Gleichverteilung von Risiken; ebendas ist im Kontext autonomen Fahrens jedoch fragwürdig. Auch wenn die Individuen nicht sicher wissen, wie mögliche Dilemmata für sie ausgehen würden, so haben sie doch Anhaltspunkte dafür, mit welcher Wahrscheinlichkeit sie z. B. die Position eines VRU oder eines Fahrzeuginsassen innehätten. Wer niemals Rad fährt, für den sind die Interessen von Radfahrern unerheblich, und es fehlen Anreize, eine faire Behandlung dieser Gruppe und damit eine egalitäre Gerechtigkeitskonzeption zu fordern. Sind Risiken dagegen ungleich verteilt, fehlen die Voraussetzungen dafür, dass eine Einigung auf eine entsprechende Konstellation erzielt werden würde; eine Ungleichverteilung wäre genau dann nicht im Interesse aller, wenn dadurch die Sicherheit einer Gruppe von Personen systematisch priorisiert würde (vgl. Hevelke & Nida-Rümelin, 2015c, S. 13–14), unabhängig davon, ob dies moralisch zulässig ist oder nicht. Aus (straf-)rechtlicher Sicht ist die Anonymität möglicher Opfer zum Tatzeitpunkt unerheblich für die juristische Bewertung:

Moralisch und rechtlich gesehen bedeutet es nach den bisher akzeptierten Standards für die Unrechtsbewertung einer Tat aber keinen wesentlichen Unterschied, ob man das Opfer schon persönlich identifiziert hat oder ob die Identität des Opfers vom Zufall abhängt bzw. von Umständen, die z. Z. der Tötungshandlung noch nicht bekannt waren. [...] Das Unrecht der Tat liegt in der Opferung des Menschen (als solchem), auf irgendwelche Identitätsmerkmale kommt es nicht an. (Hilgendorf, 2018a, S. 693)

Mithilfe eines modifizierten *Schleier des Nichtwissens* lassen sich berechtigte Ansprüche also nicht begründen. Zudem haften dem Prioritarismus in seinem Wesen als utilitaristische Variante in gewisser Hinsicht auch die Schwächen des Standardutilitarismus an. So ist ein prioritaristischer Ansatz nur anwendbar auf Fälle, in denen sich die ethisch relevanten Merkmale in Form von gerechtigkeitsethischen Abwägungen zwischen Schlechter- und Bessergestellten ausdrücken lassen. Hingegen können Fälle nicht erfasst werden, in denen sich Individuen willentlich selbst einem Risiko aussetzen, dieses deshalb aber noch nicht auf eine andere Person übertragen dürfen; der Prio-

ritarismus geht fälschlicherweise davon aus, dass Risiken interpersonell austauschbar sind:

Just like utilitarianism, prioritarianism treats risks as entities that can be unproblematically transferred between persons, albeit with a change in magnitude if the persons in question are not equally well-off. Furthermore, like utilitarianism, it always allows us to justify a disadvantage affecting one person by a sufficiently large advantage to some other person. (Hansson, 2003, S. 297)

#### 7.3.3.4 Sven Ove Hansson: Gerechtigkeit durch Reziprozität

Hansson (2003) entwirft im Rahmen seines deontologisch-risikoethischen Ansatzes, der auf der Grundlage individueller Rechte und Pflichten fußt, eine alternative Möglichkeit, berechtigte Vorteile für VRUs zu begründen. Er entwickelt sein Kriterium der Risikoakzeptabilität ausgehend von der Prämisse, dass das Recht, keinem Risiko ausgesetzt zu werden, als ein *Prima-Facie*-Recht zu verstehen ist, das absoluten Geltungsansprüchen individueller Rechte eine Absage erteilt. Dieses steht grundsätzlich jedem Individuum zu; jedoch können Risikoübertragungen und damit einhergehende Verletzungen dieses Rechts aus lebensweltlich-pragmatischen Gründen erlaubt sein.<sup>290</sup> Diesen Umstand beschreibt er als das sogenannte »exemption problem«:

It is a prima facie moral right not to be exposed to risk of negative impact, such as damage to one's health or one's property, through the actions of others. What are the conditions under which this right is overridden, so that someone is allowed to expose other persons to risk? (Ebd., S. 303)

Unter welchen Bedingungen darf ein Recht nun überschrieben und eine Person einem Risiko ausgesetzt werden? Hanssons zentrale These lautet: Das *Prima-Facie*-Recht, keinen Risiken exponiert zu

---

290 Es ist anzumerken, dass Hanssons Verwendung der Bezeichnung *prima facie* von derjenigen abweicht, wie sie in Kap. 5.3.1 skizziert wurde. Die risikoethische Verwendungsweise, die Hansson impliziert, beruht auf der Einsicht, dass Individualrechte zwar absolut gelten, Risiken, die diese zu verletzen drohen, sich aber nicht zwangsläufig manifestieren müssen. Risikoexpositionen sind damit noch nicht als Rechtsverletzungen zu werten und tangieren den Status von Individualrechten nicht in unzulässiger Weise.

werden, darf genau dann überschrieben werden, wenn die Risikoübertragung Teil eines gerechten Systems von Risiken ist, welches dem Einzelnen Vorteile bringt.

Schrittweise entwickelt Hansson ein soziales System zur reziproken Verteilung von Vor- und Nachteilen aus Risikoübertragungen, das allen zum wechselseitigen Vorteil gereicht: »Exposure of a person to a risk is acceptable if and only if this exposure is part of an equitable social system of risk-taking that works to her advantage.« (Ebd., S. 305) Dabei installiert er die Reziprozität von Risikoübertragungen als zentrales Element distributiver Gerechtigkeit, indem er betont, dass bei der Frage nach der Zulässigkeit einer Risikoübertragung zwei zentrale Komponenten zusammenwirken: auf der einen Seite deontologisch begründete Individualrechte, auf der anderen ein reziprokes soziales System zur Verteilung von risikoinduzierten Vor- und Nachteilen. Letzteres kann beispielsweise dadurch gewährleistet sein, dass Vor- und Nachteile zwischen den Betroffenen ausgetauscht werden, sofern sich diese Verteilung insgesamt betrachtet als für alle vorteilhaft erweist: »Each of us takes risks in order to obtain benefits for ourselves. It is beneficial for all of us to extend this practice to mutual exchanges of risks and benefits.« (2007b, S. 31) Hansson zufolge ist es für die Problemstellung zwar unerheblich, ob ein überschriebenes Recht vollständig eliminiert wird, nicht aber für die praktischen Konsequenzen, die sich daraus ergeben. Bleiben alle Rechte in ihrer Geltungskraft erhalten, so verbleiben wiederum Pflichten für den Handelnden, die z. B. in Form von Kompensationen oder verstärkten Maßnahmen zur Risikosenkung bestehen können (vgl. 2003, S. 303).

Die von Hansson vorgeschlagene soziale Praxis des wechselseitigen Austausches von Vor- und Nachteilen erhält ihre Legitimation dadurch, dass auf diese Weise die Handlungsmöglichkeiten aller ausgeweitet werden. Die Einführung einer Technologie ist also zulässig, wenn insgesamt gesehen jeder von ihr profitiert, auch wenn im konkreten Einzelfall gegebenenfalls Nachteile in Kauf genommen werden müssen. Eine Rechtsverletzung ist nur dann zulässig, wenn den Betroffenen durch Teilhabe am gerechten, sozialen System grundsätzlich dessen Vorteile offenstehen. Genau darin besteht der kritische Punkt von Hanssons Ansatz: Insbesondere im Kontext von Technologien bleibt fraglich, ob wechselseitige Vorteile immer gegeben sind. Innovative, auf Endkunden bezogene Technologien

sind bei der Markteinführung häufig mit hohen Kosten verbunden, die bestimmte Bevölkerungsgruppen zumindest zunächst von deren Nutzung und damit auch Vorteilen ausschließen, nicht aber ihren Risiken. Daraus müsste im Sinne von Hanssons Zulässigkeitskriterium folgen, dass Risikoübertragungen auf diese Personen nicht legitimiert wären, was wiederum die Rechtfertigbarkeit der Technologie an sich in Frage stellen würde.

Als Einwand auf diese Schlussfolgerung lässt sich argumentieren, dass eine Technologie meist verschiedene Vorteile bietet. Auch wenn eine Person von den (direkten) Vorteilen des autonomen Fahrens ausgeschlossen wäre, welche sich u. a. in einem gesteigerten Fahrkomfort äußern, so könnte sie dennoch indirekt davon profitieren, z. B. von einer allgemein erhöhten Straßenverkehrssicherheit und verringerten Schadstoffemissionen. Müller und Gogoll (2020) erweitern den Horizont der Teilhabe an möglichen Vorteilen auf die lokale Güterversorgung entsprechend:

The reason why we generally accept that other people expose us to risk by driving, we submit, is that there is a wide consensus that our traffic system works to the advantage of everybody. Everybody who uses vehicles as a means of transportation to get from one point to another benefits by it. However, it is not only people who drive that benefit from traffic but also—for instance—people who, even though they never drive, rely on their local supermarket to be fully stocked with all sorts of products that would never be there if it was not for a (international) system of traffic that relied on some mode of transportation that is associated with a tiny risk of hurting people. (Ebd., S. 1554)

Die zentrale Frage ist hierbei, ob das Risiko, dem diese Personen ausgesetzt sind, durch indirekte Vorteile gerechtfertigt werden kann. Zur Feststellung der Gleichwertigkeit notwendige Bewertungsskalen sind angesichts der Tatsache, dass sich der Wert eines menschlichen Lebens und seine Facetten grundsätzlich nicht ethisch vertretbar quantifizieren lassen, problematisch. Darüber hinaus können ideo-logische Gründe Einzelne dazu bewegen, eine Technologie nicht zu nutzen. Ob ein solcher bewusster Verzicht auf die Vorteile jedoch als Zustimmung zu entsprechenden Risikoübertragungen gewertet werden kann, ist fraglich. Das risikoethische Konzept der Zustimmung ist als Kriterium für zulässige Risikoübertragungen nur begrenzt sinnvoll und wird überdies kontrovers beurteilt, denn Handlungen können auch dann als ethisch fragwürdig gelten, wenn Betroffene

zugestimmt haben. In der Konsequenz erscheint Hanssons Ansatz zwar grundsätzlich plausibel, aber unvollständig. Er greift risikoethisch relevante Gerechtigkeitsethische Grundfragen auf, spezifiziert seinen Entwurf aber nicht konkret genug, um für eine direkte Anwendung zu taugen; die Integration anderer risikoethischer Konzepte ist notwendig.

Aufgrund des reziproken Designs kann es dazu kommen, dass Unterschiede hinsichtlich erlangter Vorteile nicht direkt ausgeglichen werden können, sondern beispielsweise über Kompensationen (vgl. McCarthy, 1997, S. 219–220) entschädigt werden müssen. Ein Nachteil in einer Situation wäre dann durch einen Vorteil in einer anderen abgegolten. Was für den Fall nicht-dilemmatischer, alltäglicher Fahrsituationen noch legitim sein mag, ist es für Dilemma-Situationen hingegen nicht. In Letzteren können mögliche Nachteile so gravierend sein, dass spätere Vorteile irreversible Schäden nicht mehr kompensieren können. Zudem sind Konzeptionen einer möglichen Kompensation aufgrund fehlender Zustimmung generell risikoethisch fragwürdig; aus praktischer Sicht kommen auch hohe Implementierungskosten hinzu (vgl. Dietrich & Weisswange, 2019, S. 233). So ist es beispielsweise sinnvoll, ein bestimmtes Maß an grundsätzlich erforderlicher Risikotoleranz zugrunde zu legen, um das System reziproker Vor- und Nachteile dahingehend zu entlasten, bei jedem noch so geringen Risiko für Vorteilsausgleich sorgen zu müssen. Um zu verhindern, dass Einzelnen sehr große Nachteile aufgebürdet werden, bietet es sich an, eine Zumutbarkeitsgrenze festzulegen, wie sie in Kap. 7.3.2 vor dem Hintergrund eines moralischen Sorgfaltsbegriffs diskutiert wurde. Auf diese Weise lassen sich die beiden Grenzkriterien integrieren und als deontologische Rahmenbedingungen eines Optimierungsproblems formulieren, das zumutbare und faire Risiken in den Fokus rückt.