自動運転車におけるトロッコ問題の実装の研究
Study of the Implementation of the Trolley Problem in Autonomous Vehicles

山中　豊 [1]
Yutaka　Yamanaka[1]

[1] 日本電気株式会社[1]　サポートサービス事業部門長　マネージングディレクター

E-mail: y-yamanakaab@nec.com

E-mail: forcenine540@gmail.com

自動運転車の開発が進んでおり，プログラムが主体となって運転制御を行う自動運転レベル 3 以上の技術が登場した．これに伴い，事故発生時の責任所在などの議論が行われている．

一方で自動運転車におけるトロッコ問題が注目されている．トロッコ問題は命の選択をテーマとする哲学的な問いである．人は何をもって正しいと判断するのかという正義の価値観の議論はいまだ結論が定まっていない．しかし自動運転車の制御にはプログラムの定義が必要である．では正義の価値観をどのように定義すればプログラムとして実装できるのか，それは自動運転車の社会受容性においてどのような意味を持つのか．本研究ではトロッコ問題を功利主義，義務論，利他主義，二重結果論の四つの類型に分類し，技術，倫理，法律の観点から自動運転車における実現性の検討を行った．

Key Words: *自動運転, トロッコ問題, Autonomous driving, Trolley problem*

1.　自動運転の動向と倫理問題

2020 年 4 月道路交通法と道路運送車両法が改正され自動運転レベル 3 が規定された．これにより高速道路渋滞時など一定の条件下で，システムがドライバーに代わって運転操作を行うことが可能となった．さらに 2022 年 4 月には自動運転レベル 4 にむけた法改正が成立し 2022 年度内にはレベル 4 による公道での走行が可能となる見通しである．レベル 4 の自動運転は人の手を介さず，緊急時の対応も含めてすべての動作をシステムが自動で行う．このようにシステムが主体となって操作を行う場合「もし自動運転車が事故を起こしたら誰が責任を取るのか？」という責任所在問題があり法整備の議論が行われているが，現行の法律では対応できない場合がある．このためレベル 4 の法改正では公共交通などの移動サービスの社会実装を想定した許可制度が新設され「特定自動運行」という定義を行い，「特定自動運行計画」の策定や「特定自動運行主任者」の指定を求めるなどが規定され，自動運転の範囲や責任を明確にしている．

(1)　自動運転車の倫理問題

法整備が進む中で，自動運転車の倫理の問題も議論されている．例えば自動運転車が前方の危険を感知しこれを回避しようとして対向車線にでる，など一時的に道路交通法に反するような挙動をとることが許されるのか，といった法規制とプログラム倫理の議論である．

一方でこの回避により別の人と衝突する可能性を認識した，などのケースも想定される．これはトロッコ問題と呼ばれる倫理上のジレンマをテーマとした問題として知られているが，このような場合に自動運転車はどのように挙動すべきかといった指針はまだ定まっていない．

本研究ではこの倫理的ジレンマの問題を扱う．トロッコ問題を自動運転車という現実の事象に当てはめたときに，どのような原理を用いて決めるべきか，そしてそれは社会実装が可能なのかを考察する．

[1] 本論文は個人の研究によるものであり，日本電気（株）の意見を代表するものではありません

2. トロッコ問題の類型整理

トロッコ問題は現在では「ある人を助けるために他の人を犠牲にすることは許されるか？」といった倫理・道徳的なジレンマを扱った命題であると受け取られている．まずこの類型を整理する．

(1) トロッコ問題とは

トロッコ問題はイギリスの哲学者フィリッパ・フットの論文に端を発している [1]．フットは二重結果論を用いてカトリックにおける妊娠中絶問題を検討したが，その中に以下のようなケースがある（表 1）．

表 1 暴走する路面電車の運転手のケース

(A)	暴走するトラム（路面電車）の運転手が別の線路へと切り替えることしかできない状況になった．狭い線路上で 5 人の男が働いている．もう一方の線路では 1 人が働いている．いずれかの線路の上にいる人は必ず殺される．

このケースでは，運転手はそのまま進み 5 人を殺すか，または進行方向を変えてもう一方の線路にいる 1 人を殺すかの選択を暗に迫られている．これに対してジュディス・トムソンは，さらにいくつかの条件を追加したケースを提示した [2]．その一つに次のようなケースがある（表 2）

表 2 暴走する路面電車と太った男のケース

(B)	Georgeはトロリー（路面電車）の線路の上の歩道橋にいる．暴走したトロリーが近づいてきて，歩道橋の先には5人の人がいてすぐには逃げられない状況である．暴走するトロリーを止める唯一の方法は何か重たいものを歩道橋の上から落とすしかない．しかし唯一の重たいものとは同じ歩道橋の上からトロリーを観ている太った男しかいない．Georgeは太った男を押して線路に突き落とし（太った男を殺して）トロリーを止めるか，またはそのようなことはやめて5人の人が死ぬに任せるかの選択を迫られている

今日トロッコ問題はこの(A)と(B)の対比として紹介されることが多い（図 1, 図 2）．その後，この倫理的ジレンマの問題は，この他にもさまざまな条件が付加されたケースが考案され，また問題のケースもトロッコのみならず各分野において多様なパターンが作られ議論されている．これらの論争では様々な主義・主張が持ち出されているが，人は何をもって正しいと判断するのかという正義の価値観の議論はいまだ結論が定まっていない．本研究は「正義とは何か」を論ずるものではないが，これらの価値観を分類し，自動運転車への実装の可能性を検討するため「正義」という呼称を用いる．

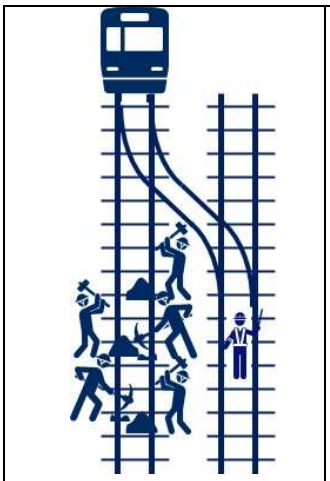

図 1 ケース(A)

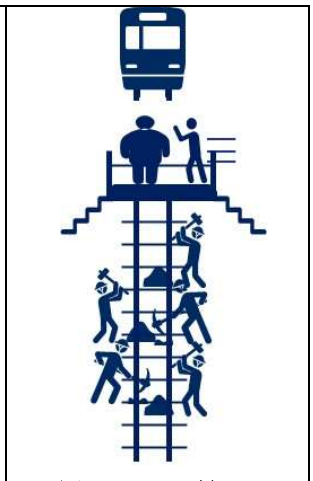

図 2 ケース(B)

(2) トロッコ問題の類型整理

トロッコ問題には様々な議論があるが，これらを整理したモデルを作成した．まず図 1，図 2 で示したケースは功利主義対義務論として，トロッコ問題の議論で語られる典型的な例であるが，これは「正義」の判断を数量で決めていいのか，という対立軸である．次に事件の当事者が誰であるかによって「正義」の判断が変わるのかという議論の軸があり，これは利他主義対利己主義として議論され，自動運転車における主体と自己犠牲問題に深く関わりがある．最後にトロッコ問題の発端となった二重結果論があるが，これは上記を部分的に包含し，個別の事象を判断することにより帰結を導くものである．トロッコ問題の命題は「人は何をもって『良いこと・正しいこと』であると考えるのか」という哲学的な問いである．この議論には他にも様々な主義・主張があるが，このいずれかに依拠または類似するものが多く，「正義」の議論としては概ねカバーできているものと考える．以下これを「正義の分類モデル」と呼ぶ（図 3）．

本研究ではこの正義の分類モデルに基づき①功利主義，②義務論，③利他主義，④二重結果論について，自動運転車との関係性を明らかにし，これをプログラムとして実装する場合の実現可能性を検討する．（利己主義の検討は利他主義の考察に含むものとする）

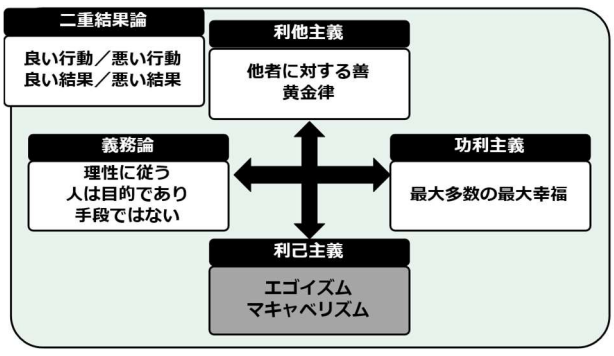

図3 正義の分類モデル

3. トロッコ問題のプログラムの定義

図 3 で提示したトロッコ問題の類型整理に従い，4 つの原理の概要を示しこれをプログラムとして実装するための定義を検討する．

(1) 功利主義 (Utilitarianism)

功利主義は，ジェレミー・ベンサム，ジョン・スチュアート・ミルらが唱えた原理であり，その中心概念は「最大多数の最大幸福」である．功利主義による道徳の至高の原理は幸福であり，これを最大化するものである．ベンサムは「効用」という言葉を用いて，快楽や幸福を生むもの，または苦痛や苦難を取り除くものを表し，比較換算できるようにした．そして効用の最大化とは個人だけではなく，コミュニティ全体の幸福を最大化するためにあらゆる手段をとるべきであるとする．トロッコ問題では図 1 の例で示したように，暴走するトロッコの犠牲になる人数が 5 対 1 という分量（人数）の比較で語られるような類型である．このケースにおいてはスイッチを切り替え，トロッコを引き込み線に誘導し，5 人の命を救う（引き込み線上にいる 1 人を犠牲にする）ことを是とする．マーク・ハウザーらが実施したモラルジレンマにおけるアンケート調査では，5000 人以上が回答したうち 89%の人がこの選択を「許される」と回答した[3]．

これを元に功利主義を実装する方法を考察してみる．ここでは事故を起こす可能性のある現場で，関係者を直接的な被害に関わる利害関係者に限定して検討する．この条件下で自動運転車がジレンマ状況に至るプロセスを検討する．一般的に走行中の自動運転車の前に，人が同時に飛び出してきて，かつ自動運転車がこれを全く同時に認識したという事態は考えにくい．より現実的な想定としては次のようなプロセスが考えられる（表 3）．

表 3 自動運転車がジレンマ状況に陥るプロセス

①	走行中の車両の前に人 A が飛び出してきて，プログラムが衝突可能性を認知した
②	次にプログラムが人 A を回避しようとした場合，新たな人 B に衝突してしまう可能性を認知した．または予め進行方向外に人 B の存在を既に認知していたが，人 A を回避しようとすると人 B に衝突してしまうという計算結果を認識した

このケースの場合，人 A の認知→人 B の認知（あるいはその逆）という時系列の流れはあるものの，どちらかを選ばなければならないというジレンマの状況になっている．このとき功利主義の原理では人 A と人 B のいずれかまたは両方が複数人であった場合，その量の比較を行う．功利主義では多数の幸福を道徳的正義とするので，人数が多い方を救うことを選択する．もし人 A と人 B が同数であると判断された場合は回避行動をとり，人 B を犠牲にして人 A を救ったとしても，より価値の大きい効用を生み出したとは言えず，この回避行動は功利主義の原理に合致しないこととなる．従ってプログラムはジレンマ状況に至るプロセスの中で，比較する人数が同数であった場合には回避行動の処理を中断する．

以上により本研究においては功利主義の実装を，**「走行中にいずれかの犠牲者を出すことが避けられないようなジレンマ状況に陥った場合，自動運転車はより人数が多い方を救う回避行動を行う（より人数が少ない方を犠牲者として選択する）．ただし同数であった場合は回避行動を行わない」**と定義する．なお人 A の認知から人 B の認知までは極めて短い時間であり，仮に人 B の存在がなければ，人 A の回避行動が十分に可能な時間の範囲内で行われたものであるとする．

(2) 義務論 (Deontology)

義務論は，エマニュエル・カントが主張する原理であり，トロッコ問題の議論においては功利主義との対比としてよく用いられる．図 2 で示すように，暴走するトロッコの前方にいる 5 人の作業者を救うために，橋の上から 1 人の太った男を突き落とす（他にトロッコを止める手段がないと仮定されている）といった例が提示されている．義務論ではこのように人を手段として用いることは許されないとする．人々の反応も前述のアンケートでは，これを「許される」と回答した人は 11%であった．

以上のことから，義務論を実装する方法について検討する．ジレンマの状況に至るプロセスは同様に表 3 の通りであるとする．どちらかの命を救うために，他方の命を犠牲にすることは義務論の原理に合致しない．このジレンマの状況において義務論の原理では人 A の命を救うために人 B の命を犠牲にしてはならない．すなわちプログラムは人 B の存在を認知した時点，または予めその存在を認知しており，回避行動により人 B に衝突してしまうことを認識した時点で回避行動の処理を中断する．また義務論においては人 A または人 B のいずれか，あるいは両方が複数であってもその判断は変わらない．

以上により本研究においては義務論の実装を，**「走行中に人 A との衝突可能性を認知しこれを回避しようとしたときに，新たな人 B との衝突可能性を認知した場合，回避行動を行わない」**と定義する．

(3) 利他主義 (Altruism)

利他主義はその行為の目的を他人に対する善におく倫理学上の学説であり，利己主義（エゴイズム）の対比として用いられる．他人の幸せに関心を払う主義やそのための行動は宗教学上の教義の中にもみられる．多くの宗教で「自分にしてもらいたいように人に対してせよ」と

いう文脈で語られている教えであり，黄金律（Golden Rule）として知られている．トロッコ問題の議論では他の原理に対する批判的なケースとして用いられることが多い．例えば図 1 で示したケースにおいて功利主義の立場では，5 人の命を救い 1 人を犠牲にすることが許されるとしたのに対して，もしこの犠牲になる 1 人が自分自身あるいは自分に近しい家族であっても同じ選択をするか？という問いが投げかけられる．ドイツの哲学者フリードリヒ・ニーチェは反キリスト教的立場からこのような宗教的犠牲の精神を批判し，自己犠牲的な「善」と利己的な「悪」という構図は誤っており，「健全」対「病的」が自然であるとしている．自分を犠牲にしてまで他人を救うことが善であるとする考えは，宗教的に押し付けられた弱者の論理であり病的である．本来は自分や家族を救いたいと思う方が健全である，というのである．

自動運転車へ利他主義の実装を検討する際に注意すべき点がある．それは自己と他の定義である．もし現在の車を運転している人が，飛び出してきた人を回避するために，自らの危険を顧みず回避行動をとり何かに衝突した，という事態を想定すればこの運転手の行為は利他的であると言えるだろう．これがレベル 4 以上の自動運転車で搭乗者が運転に関与しなくなった場合でも回避行動をとり，搭乗者を犠牲にすることがあっても利他的と言えるのか．またタクシーやバスのように多くの乗客を乗せているような自動運転車であっても，その犠牲を顧みず歩行者の命を優先するような回避行動が認められるだろうか．この問題については倫理面の実装の項で詳しく検討することとし，ここでは自動運転車の主体を車両とプログラムが一体になったものであるとしこれを自己側，衝突可能性のある人を他側であるとする．すなわち，本研究においては利他主義の実装を**「自己の危険性を顧みず，他の安全を優先し回避行動を行う」**ものと定義する．なおこの定義では自己側（一体となった車両とプログラム）にはそこに搭乗する人も含まれることになる．ちなみに前述の表 3 でジレンマ状況に至るプロセスをこの利他主義の定義に当てはめると，人 A を回避しようとすると人 B に衝突してしまう結果になるという計算結果を認識した場合，これは他対他の対立となりこの利他主義の定義の範疇では解決できない．このようなケースでは他の原理と組み合わせて検討を行う必要がある．

(4) 二重結果論 (Principle of double effect)

二重結果論はイタリアの神学者トマス・アクイナスによって提唱されたものである．ある行為が良い結果と悪い結果を生じるような状況において，そのような行為を行ってもよいのか，という問いに対してその要件を二重結果の原則としてまとめたものである．その要件とは次の 4 点を満たす必要がある（表 4）．

表 4　二重結果の原則の判定要件

①	行為それ自体が悪いものであってはならない．道徳的に良いものか，少なくとも中立でなければならない
②	よい結果が悪い手段によって獲得されてはならない．よい結果は少なくとも悪い結果と同程度には，行動の直接的結果でなければならない
③	行為者は積極的に悪い結果を望んでいない．悪い結果は「予見されて」はいても，「意図されて」いてはならない
④	よい結果は悪い結果を相殺するに足るだけの効果をもたなければならない．悪い結果を許容するに相当する重要な理由がなければならない

これら 4 要件のうち①～③は義務論的な要請である．④は帰結主義的（功利主義的）な要請である．この「許容するに相当する理由」には帰結以外の事柄も同様に考慮することが求められる．このことは刑法三十七条の緊急避難において，その要件の一つである法益権衡の原則にも同様の帰結主義的要請が見て取れる．緊急避難については法律面での実装の項で詳しく検討する．

これらの要件をトロッコ問題の議論でみると，図 1 の 5 人の命を救うためにスイッチを切り替え，その結果 1 人を犠牲にしたケースでは，①スイッチを操作し進路を切り替える行為自体は道徳的には中立である．悪い行為ではない．②操作者は 1 人を犠牲にすることによって 5 人を救ったのではない．5 人はトロッコの進行方向が変わったことによる直接的な結果として救われた．③操作者は引き込み線にいた 1 人の死を積極的に望んでいたわけではない．④1 人の命が犠牲になったが 5 人の命が救われた．よって良い結果の方が大きい．以上により 4 要件を満たしており，このケースは二重結果論で見ると許容される．一方図 2 の暴走するトロッコを止めるために橋の上から太った男を突き落としたケースでみると，①橋の上から人を突き落とすのは道徳的に悪い行為である．②5 人の命を救ったという結果は人を突き落とすという悪い手段によって獲得された．③行為者は死を招くことを知りながら，太った男を橋の上から突き落とした．これは積極的に殺したのと同等であり要件③に反する．要件④については満たしていると言える．以上により，このケースでは二重結果論で見ると否定される．

この原則によると道徳的によい行為がたまたま悪い副作用を生むことは仕方ないが，良い結果を引き起こそうとしてわざわざ悪い行為をするべきではない，となる．二重結果論はフットとトムソンによるトロッコ問題の議論の発端となった原理であるが，これをプログラムすることは難しい．まず行為それ自体の良い／悪いという定義が困難である．また，二重結果論は功利主義や義務論やなど複数の倫理の原理を複合的に援用し，一つのケースに対して複数の根拠を併用して判断を行う．このため状況次第では個別の主義が対立するような場合があり，

論理的にプログラムすることが困難となる．以上のことから本研究ではいくつかの箇所で二重結果の原則を用いた考察を行うが，これ自体をプログラムとして定義し実装することは行わない．

(5) 正義の分類モデルにおけるプログラムの定義

本研究では何が「正義」であるかという点は論じない．トロッコ問題は長らく，現実的には起こる可能性の低い机上の空論であるなどの批判も浴びてきた．しかし自動運転車の前に子供が飛び出してくるといったようなケースは現実的に容易に想像しうる事象である．この時に自動運転車はどのように挙動すべきか，またそれはいかなる理由により判断されるのか，といった議論を進める必要がある．

自動運転車には何らかの判断ルールをプログラムしなければならない．AI は意思を持たないため何が「正しい」ことなのかを自ら判断できない．判断ルールは予め人が決め，プログラムしておく必要がある．ルールの実装には，シンプルでパワフルな要件定義が必要となる．

表5 正義の分類モデルにおけるプログラムの定義

分類	プログラムの定義
功利主義	走行中にいずれかの犠牲者を出すことが避けられないようなジレンマ状況に陥った場合，より人数が多い方を救う回避行動を行う
義務論	走行中に人 A との衝突可能性を認知しこれを回避しようとしたときに，新たな人 B との衝突可能性を認知した場合，回避行動を行わない
利他主義	自己の危険性を顧みず，他の安全を優先し回避行動を行う
二重結果論	この原理自体をプログラムとして定義し実装することはできない

4. 正義の分類の実装方法と実現可能性の検討

トロッコ問題が現実社会に実装されるためには，多角的に検討を行う必要がある．本研究では正義の分類モデルのプログラムの定義を実装するにあたり技術，倫理，法律の各々の面においてその実現性を検討する．

(1) 技術面での実装の検討

この項では現状の技術水準を前提に正義の分類モデルにおけるプログラムの定義の実装可能性を検討する．

a) 功利主義の技術的実装の検討

功利主義を技術的に実装する場合には表 5 に示した定義「走行中にいずれかの犠牲者を出すことが避けられないようなジレンマ状況に陥った場合，より人数が多い方を救う回避行動を行う」をどの程度確からしく実現できるかを検討する必要がある．まずこの定義では人数の多少を正確に判別する必要がある．現在の技術では LiDAR などのセンサの性能向上により群衆はかなりの精度で検知できる．しかし自動運転車のプログラムとして実装する場合には，判別精度をより厳密に設計する必要がある．例えば人の群を複数対複数で比較できるか，人と人の重なりを正確に判別できるか，などの課題が考えられる．LiDAR などのセンサー以外にカメラ画像を AI 技術で判別する方式も考えられる．画像認識の AI 技術は急速に開発が進んでおり，人の認識においては顔認識や骨格検知，群衆を認識しカウントする AI が開発されている．これらを用いれば人数の多少を判別できる可能性がある．しかし画像認識の場合，入力情報となるカメラが取得する映像に大きく影響を受ける．高速走行中でかつ天候の影響を受けるなど，自動運転車におけるカメラ画像は条件が厳しい．LiDAR と同様にどこまで正確に判別できるかが課題である．さらに人数の検知だけではなく，その物体が「人か物であるか」という点も正確に判別する必要がある．人以外の動物である場合や，人を模した看板などがあった場合これを正確に識別できる必要がある．国土交通省のまとめによると主要高速道路における落下物で最も多いものは，プラスチック・布・ビニール類であり年間約 10 万件あるという．これらは風によって動いたり変形したりするので，LiDAR や画像認識技術でも正確な判別は困難を伴う．一般道路においてはさらに多くの情報を判別しなければならない．

以上のことから功利主義をプログラムとして実装する場合には，表 3 で示したプロセスでは人 A と人 B の人数に明確な差がある場合これを検知できる可能性が高い．飛び出してきた人 A を避けて，歩道にいる多人数 B の群れに突っ込むというような事態は防止できそうである．人 A と人 B が単数であった場合は回避行動を行わないこととなるが，厳密に判別できない場合，または人 A と人 B がいずれも複数であるがその正確な量比較が難しい場合は，功利主義の原理は適用できない．功利主義の実装に関しては限定的な条件，環境下において人数の多少を判断するような実装とならざるを得ないであろう．

b) 義務論の技術的実装の検討

義務論を実装するには表 5 に示した定義「走行中に人 A との衝突可能性を認知しこれを回避しようとしたときに，新たな人 B との衝突可能性を認知した場合，回避行動を行わない」をどの程度確からしく実現できるかを検討する必要がある．LiDAR やミリ波レーダーにより前方に飛び出した人 A は高い確率で検知できそうである．またこれを回避しようとしたときに新たに衝突可能性のある人 B も検知できる可能性が高い．義務論は人 A と人 B

両方を認識した時点で回避処理を棄却するので，これは実現できそうである．

義務論の実装は一見簡単そうに見える．しかしここでは補足的に表 3 で示したジレンマ状況に至るプロセスを拡張して検討してみる．まず自動運転車は飛び出してきた人Aを検知した際に，緊急ブレーキを作動させるなどの初期的な安全対処を行うはずである．しかしそれでも衝突回避が間に合わなくなったと判断される場合，回避行動を検討する．ここからが厳密にいう義務論の定義の範囲である．そして人Aを回避しようとすると人Bに衝突してしまうという倫理的ジレンマの状況を認識し，回避行動を中止する．この場合自動運転車は次の処置としてプログラムされている更なる安全処置や別の回避措置へと進むことになる．例えばもはやそれでも避けられない事態なったと判断した場合は，少しでも被害を軽減させるような衝突緩和装置を作動させるなどの対処が考えられる．義務論のプログラムの実装においては，その前後に実施される安全措置の部分を除き，倫理的ジレンマの状況における命の選択の部分に限って言えば，「現状からの変更を行わない」ということになる．プログラムを何も実装しないということと，判断した結果として回避処理を行わないということは異なる．前述の通り義務論の定義を実装するためには，車両の前方および周囲の危険の検知，初期的な安全動作を行った上での判断，回避行動の検討と新たな衝突の危険の検知，といった複数の認識，判断を正確に行った上で「進路方向の変更を行わない」決定をするということになる．

義務論の実装における技術的な課題としては，人の検知の確からしさであろう．ミリ波レーダーや LiDAR といったセンサーには検知できる物体の材質に得手不得手がある．風に舞って目の前に飛び出してきたビニールや，道路わきの人型の看板を人と誤認してしまうとジレンマ状況の判断の前提が崩れてしまう．現在のセンサー技術による人検知の性能はかなり向上しているが，天候などの環境要因により精度が落ちるなどの課題もある．技術的に実装するためにはセンサーの検知感度を安全側に倒す必要があるが，誤認の確率を合理的な水準まで低くできない限り，義務論の原理を達成したことにはならない．

c) 利他主義の技術的実装の検討

利他主義の実装においては「自己の危険性を顧みず，他の安全を優先し回避行動を行う」と定義した．この定義においては表 3 で示した人Aに衝突することを回避しようとした場合に人Bに衝突してしまうという倫理的ジレンマの状況への対応は実装できない．これは自動運転車側から見た場合人Aと人Bという他者同士の対立構造となっており，いずれを選択しても利他主義の定義に反するためである．利他主義が適用できるケースは限定的であるが．これについては倫理面の実装の項で詳しく検討する．

d) 二重結果論の技術的実装の検討

二重結果論は本研究では実装せず考察のみ行うとしたが，仮に実装するとした場合どのような課題があるのだろうか．表 4 において二重結果の原則の判定要件を示したが，これについて技術的実装を検討してみる．表 3 の倫理的ジレンマの状況に至るプロセスでは，①行為それ自体が悪いものであってはならないという要件は，人Aが飛び出してきたことを検知し，これを避けようとする行為自体は悪いことではない．具体的には人Aを検知した事実と回避行動を判断したことを運行記録装置やドライブレコーダーの画像に残すことにより要件①は立証されるであろう．②の良い結果が悪い手段によって獲得されてはならない，という要件は回避行動により人が救われるという良い結果と，他方の人Bが犠牲になるという悪い結果が同時に存在するが，少なくとも回避行動そのものは悪い手段ではない．よって要件②も立証可能であろう．③の行為者は積極的に悪い結果を望んでいないは，プログラムは積極的に悪い結果を望んでいるわけではないのでこれを立証することになる．仮にプログラムが審査を受けるとしたら，その設計ロジックにおいて人を検知し避ける機能を実装していることを説明できればよい．④の良い結果は悪い結果を相殺するに足るだけの効果を持たなければならない，については実装が難しい．仮に④に功利主義的観点を持ち込むと人数が多い方を救うほうが，良い結果が大きくなる．人数が同じ場合でも人Aと人Bは少なくとも同等の価値をもつので，結果の効果としても支持されうる．しかし倫理的ジレンマの状況に既に陥っている状況においては人Aに加えて人Bも認知されている．そして前述の説明とは逆に，「人Bを救うため」として回避を行わなかったとしても①～③は成り立つ．そして④の結果も同様の論理により成立する．これではプログラムは回避しても回避しなくてもよいことになってしまうため定義できない．

二重結果論は一つの行動が良い結果と悪い結果を同時にもたらすときにこれを判断するための原理である．仮に「道徳的によい行為がたまたま悪い副作用を生むことは仕方ない」という主義だけを切り出して定義しようとした場合，①～③の要件に従って，良い行為を行おうとしたことだけを立証すればよいようにも思えるが，④の要件はあくまで結果としての判断されるものであり，もしここに功利主義的な人数優先の概念を取り入れた場合には，③の要件である，悪い結果が予見されていたものから意図されたものに変わる可能性があり成り立たなくなる．従って全体としてプログラムで実装することはできない．

(2) 倫理面での実装の検討

この項では，倫理的な観点から自動運転車が「危険を回避する」とはどのような意味を持つのかを考察し，その実現可能性を検討する．

トロッコ問題で議論されているケースではどちらの命を救うべきかという設問に対して，回答者は第三者的な「観察者」の立場である場合が多く見受けられる．一部のケースでは事案に関与する当事者として設定されているものもあるが，それでも自身の直接的な被害を想起することは少なく，自己の価値観のみで回答を行う場合が多い．従来の人が運転する車は人との事故を起こした場合「加害者」として扱われる立場である．物理的な重量と強大な運動エネルギーを有する車からみれば歩行者の方が弱者であることは明白である．しかしレベル 4 の自動運転車が登場し，人が運転者から単なる搭乗者や利用者の立場となった場合でも同様の位置づけであり得るのか．自動運転車のトロッコ問題を考察する際には搭乗者にも及ぶ被害を考慮することが欠かせない．

自動運転車の前に人が飛び出して来た場合どうすべきかという問いを倫理的な観点でみると，人的被害が発生しないことが最善であることは論を俟たない．誰にも害を及ぼすことなく安全に避けることが可能であれば避けるべきであると多くの人が思うであろう．この安全に避けるという機能は操舵回避支援システムなどの名称で一部の車種に実装され始めている．現状の操舵回避支援システムでは二次衝突の危険性が少しでもあった場合には回避しないようになっている．この二次衝突の危険性とは，回避するための十分なスペースがなかったり，回避先にある別の障害物と衝突してしまうことを指す．

これは将来的にレベル 4 自動運転車が登場した際にも同様の原則が引き継がれるものと思われる．自動運転車の緊急回避は最初の危険を安全に避けうることが明確な場合にのみ遂行される．二次衝突の危険が既にあったり，最初に認識された危険と回避行動をとった結果としての二次衝突の危険可能性の相対的な大小がよくわからない場合には，危険可能性を予見できない制御を行うのは適切ではなく，緊急回避行動は棄却される可能性がある．実際にはこの危険可能性のリスク評価は様々な外部要素（複数のセンサーからの入力や周辺の状況，走行速度など）をパラメータとして入力しこれを統合して評価される．回避行動後のリスク評価がどのぐらい高ければ危険と判断し緊急回避を行わない判断をするのか，あるいはどのぐらいまで低ければ安全と判断し緊急回避を行うのか，これらは自動車メーカー・設計者によって異なり，アルゴリズムによって積極的危険回避型や消極的危険回避型などの差が生じる可能性がある．

以上のことより，自動運転車の緊急回避行動は現状の操舵回避支援システムの延長線上にあるとの前提のもとに，自動運転車に搭乗者が存在し，前方に障害物が発生し倫理的ジレンマの状況に陥った場合どう対処すべきかを検討してみる．この場合プログラムへの実装としては**「二次衝突の危険性の評価を行うこと」**と定義できる．この場合の障害物は物または人である場合があるが，これをケースとして整理すると次のようになる（表 6）．

表 6　二次衝突の危険性の評価対象の分類

検知する障害物の種類		回避先の障害物	
		物	人
前方の障害物	物	①	②
	人	③	④

①	前方に落石などの危険を感知し，緊急回避を行おうとするとき，障害物等に衝突する可能性を認識する
②	前方に落石などの危険を感知し，緊急回避を行おうとするとき，別の人に衝突する可能性を認識する
③	前方に飛び出してきた人を感知し，緊急回避を行おうとするとき，障害物等に衝突する可能性を認識する
④	前方に飛び出してきた人を感知し，緊急回避を行おうとするとき，別の人に衝突する可能性を認識する

①は本研究で扱う倫理的ジレンマの状況ではない．これは現行の操舵回避支援システムの延長線上として，技術的に緊急制動が可能かを検討するものである．

これに対して②，③は自動運転車の搭乗者対外部の人というケースである．緊急回避を検討した結果，②は外部の人に被害をもたらす可能性を，③は自動運転車の搭乗者に対して被害をもたらす可能性を認識したものであり，いずれも倫理的ジレンマの状況に陥っている．これらのケースに対して二次衝突の危険性の大小と表 5 で定義した正義の分類モデルにおけるプログラムの定義を合わせて検討してみる．なお④は表 3 で示した人対人の倫理的ジレンマに陥るプロセスに新たに自車の搭乗者が加わったものである．これについては後述する．

a) 功利主義の倫理的実装の検討

功利主義の定義では人数が多い方を救う選択を行う．自動運転車においては車内に設置されたセンサーなどにより自車の搭乗者の人数を把握することは比較的容易であると考えられる．二次衝突の危険性がある場合は回避行動を行うべきではないという前提条件がある場合，回避行動をとることによって搭乗者の危険可能性が増減するかを評価しなければならない．もし自車内の人数＞外部の人数であった場合は②，③いずれのケースにおいても回避行動をとることによって搭乗者の危険可能性が下がるとは限らない場合には，功利主義の価値観に合致しない．二次衝突の被害リスクが相当量低いと予見される場合以外は回避を行わない消極的回避型の車となる可能性がある．さらにケース②では回避行動をとることは別の人に衝突することを意味する．これは自車の搭乗者の

危険可能性に加えさらに回避先の人を巻き込んだリスクの総体を増加させることになる．功利主義の原理にも反するこのような回避行動は許容されないだろう．

逆に自車内の人数＜外部の人数であった場合は②のケースでは当然ながら回避行動は行われないが，③のケースでは功利主義の定義では回避行動を行うべきとなる．この場合前提条件である二次衝突の被害リスクをどこまで許容するかにより，積極的危険回避型や消極的危険回避型などに分かれる可能性がある．これは自己犠牲型や自己防御型といった自動運転車の性格を決めるものであるが，回避行動を行うことにより搭乗者の危険可能性が増加しないことが条件となる．

b) 義務論の倫理的実装の検討

義務論においてはジレンマの状況において（人数にかかわらず）回避行動を行わないことと定義されている．すなわち②，③のケースともに回避先に二次衝突の危険性があった場合は回避しない．これは現行の操舵回避支援システムで既に実現できていると言え前提条件とも合致する．義務論では回避行動によりいずれかのリスクが増加するような行動は棄却され，回避先に少しでも二次衝突の危険性があった場合は回避しない消極的回避型の自動運転車となる可能性がある．

c) 利他主義の倫理的実装の検討

利他主義の定義では他の安全を優先する回避行動を行うとされている．従ってケース②では二次衝突の対象が人であるため回避行動を行わないのに対し，ケース③では前方に飛び出して来た人を回避する行動を優先する．この時，前提条件である二次衝突の危険性がある場合は回避しない，という条件をおきつつもその危険性の評価において相当量まで許容される積極的回避型の車になる可能性がある．

d) 3 者の利害関係者がある場合の考察

④のケースでは飛び出して来た人，回避先にいる人に加え自車の搭乗者が加わり 3 者の関係性となっている．この場合表 5 で定義した 2 者間の対立構造におけるプログラムの定義の範疇を超えるので新たな定義が必要となるが，ここでは考察のみ行う．功利主義の原理では最大多数の幸福を求めるため，すべての利害関係者の被害リスクを判定しこれを最小化するような動作が求められる．このようなプログラム定義を行うことは相当の困難を伴うであろう．なお，この被害リスクの最小化については法律面での実装検討でさらに詳しく述べる．義務論の原理では倫理的ジレンマの状況において区別を行わない．これは利害関係者が 2 者から 3 者になっても同様である．従って④のケースにおいても回避行動を行わないこととなる．利他主義は前述のとおり人対人の対立構造においては実装できない．

利害関係者が増えその関係性が複雑になるとプログラムの定義はより一層複雑になる．現実世界に起こり得るすべての事象を想定することは不可能であり，プログラムの定義においては極力シンプルな判定根拠により判断を行うようなルールが必要である．

(3) 法律面での実装の検討

この項では法律面での検討を行う．まず現行の法体系において自動車が許可を得て走行するに至るまでのプロセスを，そして事故を起こした際の責任問題に深く関わる緊急避難について確認する．法的観点から人を避けるようにプログラムを行うとはどのような意味を持つのかを考察し，その実装可能性を検討する．

まず関連法規制の整理を行うと，自動車に関わる法律としては道路交通法や道路運送車両法がある．道路運送車両法では保安基準が定められており，自動車はこの技術基準に適合するものでなければならない．技術基準とは保安基準で定義される車両の規格や各装置の基準である．すなわち自動車は保安基準に規定された技術基準に適合し，試験機関で審査を受け，これに適合していることの証明（適合証明）を受けて走行することができるようになる．自動運転車のプログラミングもこれらの法規制で規定しこれに適合させる必要がある．

一方事故を起こした場合の製造者の責任に関する法律に目を向けると，まず犯罪の違法性の認定についてトロッコ問題の議論では，いずれかの人の命を犠牲にすることによって他方を救うという前提がある．目的は別として自動運転車が人の命を犠牲にしたという結果を生じている以上，これは犯罪の構成要件に該当すると考えられる．次に犯罪の認定手順としては違法阻却性事由がないかが検討される．違法阻却性事由とは正当防衛（刑法三十五条）および緊急避難（同三十七条）に該当する事由のことであり，本研究ではこの緊急避難を検討する．緊急避難が成立するための要件としては以下の 4 点を満たす必要がある（表 7）．

表 7 緊急避難の成立要件

①	現在の危機	危険が実際に存在するか間近に迫っている状態のこと
②	避難の意思	危険を避けようという意図で行ったということ
③	補充の原則	危険を避けるために，他に方法がないこと
④	法益権衡の原則	生じた害が避けようとした害の程度を超えないこと

これらの要件を表 3 で示したプロセスで検討すると，①現在の危難（飛び出した人に衝突する危難がある），②避難の意思（人と衝突するのを避けようという意思がある），③補充の原則（衝突を避けるためには，人を回

避するようにプログラミングするしかない），となりこれらは要件を満たしている．④法益権衡の原則は飛び出してきた人，回避した先の人，および自動運転車の搭乗者の負傷した人数やけがの程度により認められるかどうかが判断される．緊急避難が成立すれば製造者は免責となるが，もし認められない場合でも過剰避難として刑の減軽または免除が認められる場合がある．

もう一点，自動運転車が死傷事故を起こした場合は刑法の業務上過失致傷罪（第二百十一条）にも問われる可能性がある．この場合の過失は注意義務に違反する状態や不注意をいい，システム設計者が注意義務を果たしていたかどうかは，設計したシステムが保安基準を満たしていたか否かによる．すなわち事故を回避するために人を回避するというシステム設計が保安基準に適合しているかどうか，というということになる．ここで注意義務違反とは結果の予見可能性と結果の回避可能性があることである．保安基準に適合しているということは，そのような「プログラミングを行うこと」を意味する．これを「回避する」ことは出来ずそれによって結果が発生した以上は結果の回避可能性がないこととなる．事故を起こすという結果の予見可能性はあったとしても，結果の回避の可能性がないのであれば，これは注意義務違反に当たらない．システムは緊急避難を意図するようにプログラムされており，事故が起きたとしても緊急避難の条件が成立すれば，これは保安基準に適合していることとなり，システム設計における過失はなくなる．すなわち業務上過失致死傷罪にはあたらない，ということになる．

以上により法的観点から人を避けるようにプログラムの実装を行う場合には，道路運送車両法の適合証明を受けつつ，かつ**「緊急避難が成立するような状況において動作すること」**と定義することができる．この定義を正義の分類モデルにおいて検討してみる．

a) 功利主義の法的実装の検討

功利主義の実装の定義である「より人数が多い方を救う」は緊急避難の成立要件である法益権衡の原則とも合致している．現行法の緊急避難は，功利主義の定義を支持していると考えられる．これまでの検討においては人数が同数または多少の判別が難しい場合には緊急回避を行わないという条件が付与されてきた．しかし，刑法三十七条の緊急避難ではもし衝突を避けようとして，結果として生じた害の方が大きくなってしまった場合でも任意的減免が認められる場合がありうる．これは二重結果論にも通じているとも考えられる．これにより法的観点での功利主義の実装は現行法への親和性が高いと言える．

b) 義務論の法的実装の検討

義務論は回避先に別の人を検知した場合，たとえ回避により救われる側の人数（法益）が明らかに多い場合であっても回避を行わない．これは法益権衡の原則とは合致しないが，そもそも緊急回避行動を行わないため，緊急避難には当たらない．義務論の実装により回避行動を行わなかった場合には結果として救われた人数（法益）が多かったとしても，緊急避難を行うように意図された実装を立証する根拠がなければ違法阻却性事由に該当しない．結果として犠牲者を出している以上，何らかの法的責任は免れない．義務論の実装は現行法においては支持されない．

c) 利他主義の法的実装の検討

利他主義は表 6 で示したような自車対他の構図になっている時に実装可能性がある．表 6 のケース③の場合，飛び出して来た人を回避する行動を優先するが，二次衝突の危険性がある場合において，たとえ飛び出してきた人よりも自車の搭乗者の数が多いような場合にも回避行動をとってしまう行為は法益権衡の原則に合致しない．回避行動そのものは緊急避難であるが良い結果を伴う場合と伴わない場合があり，法的判断としては二重結果論の判断に近いものになるともいえる．

5. まとめ

本研究ではトロッコ問題を正義の分類モデルとして整理し，技術的，倫理的，法的観点から考察してきた．その結果，正義は分類可能であること，また各原理の実装可能性には差があることが分かった．これらは倫理的ジレンマの状況をプログラムする際に一定の指針となり得る．これまでの検討結果を以下にまとめる（表 8）．

表 8　正義の分類モデルの実装検討結果

分類	プログラムの定義の実装検討結果
功利主義	現行の法律と整合性がよいが技術的な実現性は限定的である
義務論	現行技術の延長線上として技術的・倫理的観点での実現性が高い．現行法における法的支持は得られない
利他主義	適用できるケースが限られており，これ単独では実装できない．現行法による法的支持も得られない場合がある

まず功利主義対義務論では義務論の方が実装可能性が高い結果となった．義務論は，技術的，倫理的観点での実装可能性が高く，現行の操舵回避支援システムなどの延長線上として期待できる．人の命を区別しないという原理は，二次衝突も含めたリスクの最小化において，よりシンプルな実装の可能性を提示するものである．一方で緊急避難が成立するような状況には至らず，法的根拠に基づく支持は得られない．今後の法整備が待たれると

ころである．

利他主義においては実装が困難であることがわかった．まず人対人という倫理的ジレンマの状況を解決できないという定義上の限界があり，またこれを自己（自動運転車）対他とした場合においても，法的支持の得られない挙動をとってしまう場合があるなどの課題があった．ただし利他主義が実装できない＝利己主義（自己防御型の車）を推奨するものではない点に注意が必要である．

最後に二重結果論はプログラム自体の定義が困難であるため，実装できない結果となった（図 4）．

いずれの分類においても人検知の確からしさや制御可能性といった技術的課題が残る．特に後者は，現行の操舵回避支援システムの延長線上として考えた場合，より短い時間の中で正確な判断を行う必要があり，厳密な信頼性評価が必要となる． これらの原理を実装するためには，今後のさらなる技術開発が必要となる．

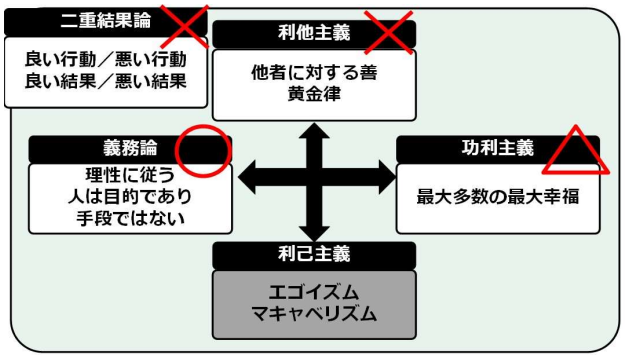

図 1　正義の分類モデルの検討結果

自動運転車の開発が急速に進む中で AI 技術に対する過度な期待から事故発生時の責任論ばかりに目が向きがちである．しかし技術には常に限界がある．本研究においてもトロッコ問題のプログラムの定義を極力シンプル化したが，これが複雑な現実世界において全てに通用するものではない．本研究で示したように正義の価値観は一つではなく，複数を組合わせるなどの応用も必要となるであろう．このため現実的には緊急時にはまずブレーキをかけること，などの議論もなされている．しかし自動運転車が社会的に受け入れられるためには，技術的な限界点だけではなく，倫理的観点，法的観点を統合して議論を行ない人々の理解を深める必要がある．今後も各分野での議論が深まることが期待される．

謝辞：本研究を進めるにあたり専門的見地から貴重なご助言を頂いた，小川 仁志 山口大学国際総合科学部教授（哲学），樋笠 尭士 多摩大学経営情報学部専任講師（法学），青木 啓二 先進モビリティ株式会社 代表取締役社長（自動運転技術開発），雨宮 秀樹 日本電気株式会社 上席技術主幹：（自動運転技術・ITS 技術），河合 英直 独立行政法人交通安全研究所 主幹研究員（安全技術），他同席諸氏に心より感謝致します．

参考文献

1) Foot, P.: The Problem of Abortion and the Doctrine of the Double Effect, *Oxford Review*, No. 5, 1967
2) Thomson, J.: Killing, Letting Die, and The Trolley Problem, *The Monist*, Vol. 59, Pages 204-217, 1976
3) Hauser, M. D et al.: A dissociation between moral judgments and justifications, *Mind & Language* 22(1):1-21, 2007
4) 樋笠尭士：AI と自動運転車に関する刑法上の諸問題～ドイツ倫理規則と許された危険の法理～， 嘉悦大学研究論集第 62 巻 2 号 p21-33，2020
5) 佐藤英明：自動運転車とトロリー問題，中央学院大学人間・自然論叢　vol.48, p.21-54，2020
6) 笠木雅史：自動運転の応用倫理学の現状と課題：自動運転とトロリー問題， 日本ロボット学会誌 Vol.39 No.1 p22-27，2021

(Received ???)
(Accepted ???)

Study of the Implementation of the Trolley Problem in Autonomous Vehicles

Yutaka Yamanaka

The development of autonomous vehicles is progressing, and technologies for automatic driving level 3 or higher, in which the program takes the initiative in controlling driving, have emerged. With this development, there has been a debate on the responsibility in the event of an accident.

On the other hand, the TROLLEY PROBLEM in autonomous vehicles is attracting attention. The trolley problem is a philosophical question about the choice of life. The debate on the value of justice, what people judge to be right, has not yet been settled. However, the control of autonomous vehicles requires the definition of a program. How can we define the values of justice so that they can be implemented as a program, and what does this mean for the social acceptability of autonomous vehicles? This study classifies the trolley problem into four types: Utilitarianism, Deontology, Altruism, and Principle of double effect, and examines its feasibility in autonomous driving from the viewpoints of technology, ethics, and law.