

[論文]

# 道徳的行為者としての自動運転車

佐藤英明

〈目次〉はじめに

- 1 自動運転車の責任問題
  - 2 人工的道徳的行為者の責任
    - 2-1 人工的道徳的行為者 (AMA)
    - 2-2 責任なき道徳的行為者
    - 2-3 AMA 批判
  - 3 「責任」概念の分析
  - 4 自動運転車の事故の責任
- おわりに

## はじめに

自動運転車に関する大きな課題の一つとして、交通事故が生じた場合の法律的責任の問題が指摘されている。民事責任については、現行の事故補償の制度が自動運転車の事故においても対応可能なものであるのか否かが問題となる。また、人間が運転していない状態で事故が生じたとき誰が処罰されるのかという刑事責任の問題もある。人間の運転者がAI (Artificial Intelligence) に置きかえられるという観点からみれば、それはAIへの帰責が可能かという問題とみることもできる。しかし、この問題に関連する「責任」「自律」「行為者」といったことばは多義的でその意味も曖昧であり、議論の混乱もみられる。本稿では、これらの概念の検討を通じて、自動運転車の事故における帰責の問題について考察する。

## 1 自動運転車の責任問題

自動車の自動走行技術には、多くの期待が寄せられている。交通事故の原因の約9割を占めるとされる人間の運転ミスによる事故がなくなれば死者数を大幅に減らすことができるし、円滑な運転により交通渋滞も緩和され環境負荷も軽減することができる（より安全かつ円滑な道路交通）。また、高齢者や過疎地域における移動支援を実現でき、ドライバー等の労働力不足にも対応できるうえ、運転の快適性も向上される（より多くの人々が快適に移動できる社会）。さらに、自動車関連産業の国際競争力強化をもたらす新たな関連産業も創出できるうえ、運輸・物流業の効率化をはかることができる（産業競争力の向上・関連産業の効率化）<sup>(1)</sup>。

清水和夫は、このような大きな期待を担う自動走行を実現するにあたって課題となることを①技術的な課題、②社会受容性、③法律問題の三つに分けたとき、「法律問題が一番難しい」のではないかと述べている。自動運転車

の事故が起こったときのメーカーの責任、販売者の責任、使用者の責任などの「責任問題」について論じる際に、「今までの考え方とは違う新しい思考回路」「新しい社会の倫理や常識」「法律」が必要になると考えられるからである。<sup>(2)</sup>

交通事故の加害者の責任については、刑事責任（刑罰）、民事責任（損害賠償）、行政責任（行政処分）、社会的責任などがある。これらの責任については、これまで人間の運転者が車両の運転操作をおこなうことを前提と考えられてきた。自動運転車が普及したとしても、交通事故がまったくなくなるというわけではない。自動運転車の絡む交通事故について、そうした責任のあり方をどう考えるべきか、どう考え直すべきかを検討しておく必要がある。

自動走行に関する議論においては、自動運転の技術レベルの違いが前提とされている。Society of Automotive Engineers (SAE) Internationalが作成した基準 J3016が国際的に用いられている。高度情報通信ネットワーク社会推進戦略本部・官民データ活用推進戦略会議が公表している「官民 ITS 構想・ロードマップ2021」によれば、その概要は、以下の通りである。

### ③ 運転自動化レベルの定義

SAE レベル	概 要	操縦の主体
レベル0 運転自動化なし	・運転者が全ての動的運転タスクを実行	運転者
レベル1 運転支援	・システムが縦方向又は横方向のいずれかの車両運動制御のサブタスクを限定領域において実行	運転者
レベル2 部分運転自動化	・システムが縦方向及び横方向両方の車両運動制御のサブタスクを限定領域において実行	運転者
レベル3 条件付運転自動化	・システムが全ての動的運転タスクを限定領域において実行 ・作動継続が困難な場合は、システムの介入要求等に適切に応答	システム (作動継続が困難な場合は運転者)
レベル4 高度運転自動化	・システムが全ての動的運転タスク及び作動継続が困難な場合への応答を限定領域において実行	システム
レベル5 完全運転自動化	・システムが全ての動的運転タスク及び作動継続が困難な場合への応答を無制限に（すなわち、限定領域内ではない）実行	システム

このうちレベル2とレベル3との間には大きな違いがあるとされる。レベル2以下においては、「運転者が一部又は全ての動的運転タスクを実行」しなければならないのに対し、レベル3以上においては「自動運転システムが（作動時は）全ての動的運転タスクを実行」することになる。

ただし、レベル3の「条件付運転自動化」では、自動運転システムのモニタリングに限界が生じると、運転者に対応が委ねられ、運転者が適切に対応（オーバーライド）しなければならない。これに対し、レベル4以上になると、人間の運転者によるオーバーライドは想定されていない。この点で、レベル3とレベル4以上では、緊急時等に人間の運転者による対応が必要となるか否かで大きな差がある。<sup>(4)</sup>

自動運転技術のレベルが異なると、運転者が車両のコントロールに関わる度合いが違ってくる。運転者の車両制御への関与が低下すると、事故に対して責任を負うべき主体も変わる。場合によっては、現行法の枠組み自体が維持できるかどうかの問題になるとも考えられる。

刑事責任については、運転者にとって事故の回避が可能であったことが前提とされる。レベル3以下で人間が運転者となっている場合の事故については、現行法の規定が適用されると考えられる。しかし、レベル4以上で事故が生じた場合には、乗員が事故回避措置をとることが不可能なため、乗員の刑事責任を問うことはできなくなる。自動車メーカーや自動走行システムの設計者の刑事責任を問うことは不可能ではないが、現実に刑事責任が肯定される可能性は低いと考えられている。<sup>(5)</sup>

民事責任については、運転者の一般不法行為責任（民法709条）、使用者責任（民法715条）、運行供用者責任（自動車損害賠償保障法3条）などに加えて、自動車メーカーに対し製造物責任（製造物責任法3条）が問われる可能性もある。<sup>(6)</sup> 自動運転技術により人間の運転ミスによる事故が減ると、責任が運転者から自動車メーカーへと移行し、メーカーの責任比率が高まるとも予想されている。

そうしたなかで自動走行システムをシステムたらしめている AI そのもの

の責任を問えるのかということも一つの問題となっている。今井猛嘉によれば、民事責任に関しては、「システムを責任主体とする発想」が、「一定程度、妥当する」。「自動運転の市場投入に関連したさまざまなもの」を一括して「システム」とし、その責任を問うことは、「現在の民事ないし行政法上の責任追及システム（たとえば、共同不法行為論）とも親和的だからである<sup>(7)</sup>」。しかし、システムそのものの刑事責任を問い、システムに刑罰を課すことができるとは考えにくい。

自動運転車はAIにより自律的に動くロボットの種類とも考えられ、ロボットカーとも呼ばれる。自律型機械の道徳性についてはこれまでさまざまな考察がおこなわれてきたが、そのなかの一つが、機械を「道徳的行為者」とみなすことは可能かというテーマである。「人工的道徳的行為者」の可能性をめぐる議論である。以下では、この「人工的道徳的行為者」の「責任」に関する考察を通じて、自動運転車に関する「責任問題」を考える。

## 2 人工的道徳的行為者の責任

### 2-1 人工的道徳的行為者（AMA）

自律型機械の道徳性の考察においては、ロボットをたんなる機械として扱うのか道徳的行為者として扱うのが問題となる。むろん、ロボットが人間とまったく同じ意味で「道徳的行為者」であるとは考えられない。人間とまったく同等の道徳的行為者とみなされるには、心をもち自由意志をもつ必要があると考えられるからである。だが、機械が心をもちうるか否かは、心の哲学における永遠の難問である。

それでも今後、自律型機械の振る舞いが人間にますます重大な倫理的影響をおよぼすものとなることはたしかであろう。その際、ロボットを従来どおりたんなる「道具」として扱うべきか、それとも人間とは異なる意味において「道徳的行為者」として扱うべきかということが問題となる。「道徳的行

為者」という概念を機械にまで拡張する場合、それらの人工物は、「自由意志や精神状態や責任を示す必要はないが、善悪のために行動することが可能な存在者」として「人工的道德的行為者 (Artificial Moral Agent = AMA)」<sup>(8)</sup>と呼ばれる。

人間同士の場合、相手と自分の間に特定の道德的問題に関して意見の不一致があったとしても、相手がそのような結論にいたった理由を説明できるのであれば、その人物を道德的行為者と認めることができる。機械であっても、ある状況下で一定の行為を選択したとき、そのような選択にいたった理由を人間並みに説明できるのであれば、道德的行為者とみなしてもよいのではないか。そう考えれば、選択した行為を正当化する理由を与える能力に関して人間と見分けがつかない機械は、AMA と判断してよいことになる。チューリング・テストにもとづくこのような判定基準は、「道德的チューリング・テスト (Moral Turing Test = MTT)」<sup>(9)</sup>と呼ばれる。しかし、チューリング・テストのように人間か機械かという識別可能性を「道德的行為者」の判定基準とするのは適切とはいえない。人間の場合、自分の行為に理由を与える能力をもたない小さな子どもであっても道德的行為者とみなされるし、道德的に不正と判断されるような行為を選択する人間も少なくない。より不適切な行為を選択する方が人間だと考えれば、機械を識別するのは容易だろう。

それでは、識別可能性ではなく、人間以上に一貫して適切な行為の選択ができることを基準としてどうか。人間と比較してより道德的であることを判定基準とするのである。これは「比較道德的チューリング・テスト (comparative MTT)」と呼ばれる。しかし、人間よりも道德的であるという基準は低すぎるであろう。人間が道德的過ちを犯すことめずらしくないし、人々はそれに対して寛容である。しかし、機械が道德的過ちを犯すことは許容されないであろう。機械にはたんに人間を上回る適切な道德的意思決定が求められるだけでなく、完璧な選択が期待されることになる。他人に害をもたらすような行為の選択が機械によってなされた場合、同じ選択が人間に

よってなされた場合以上に、人々は不寛容になるだろう。<sup>(10)</sup>

チューリング・テストのように人間との比較によって「道徳的行為者」か否かを判断するのは妥当とはいえない。しかし、推論によって選択された行為を正当化する理由を与える能力が「道徳的行為者」に求められることはたしかであろう。AMAには、適切な道徳的選択がおこなわれるような規範を実装しなければならない。ウォラックとアレンは、その方法として、特定の倫理理論を実装するトップダウン・アプローチ、学習、発達、進化の方法を適用してAMAが適切な道徳的決定をおこなえるようにしていこうとするボトムアップ・アプローチ、その二つを組み合わせたハイブリッド型のアプローチについて論じている。

## 2-2 責任なき道徳的行為者

ルチアーノ・フロリディは、情報倫理学において採用される「抽象化レベル (level of abstraction)」は、生命倫理学や環境倫理学とは異なるという。理論は少なくとも一つの抽象化レベルを用いて、それが存在論的に関与する部分を明らかにする。ある理論において何が「道徳的行為者」とみなされるかについても、抽象化レベルに依存する。<sup>(11)</sup>

ある行為は、それが道徳的善か道徳的悪をもたらしうるとき、道徳的性質をもつ。そのような性質をもつ行為をなすものは「道徳的行為者」である。では、そもそも「行為者」とはいかなる存在か。フロリディによれば、情報倫理学における「行為者」とは「相互的、自律的、かつ適応的な遷移システム」のことである。初期状態から状態を変化させるのが遷移システムであるが、抽象化レベルが低くなるほど変化を捉えるための状態要素の数は増える。それゆえ抽象化レベルを変えることで、遷移システムとみなされる範囲は拡大する。遷移システムが環境と相互に作用しあうとき、それは「相互的」である。遷移システムが、直接の相互作用なしに内部的遷移によって状態を変化させるとき、それは「自律的」である。環境との相互作用によって状態変化の遷移ルールが変化するとき、遷移システムは「適応的」である。

このような相互性、自律性、適応性をそなえた遷移システムが「行為者」である。行為者は、受容者と同時に関わり合う相互的存在であり、環境から独立して変化する自律的存在であり、相互作用から作動方法を学習する存在である。<sup>(12)</sup>

行為者のなす行為が、道徳的善か道徳的悪をもたらすとき、その行為者は「道徳的行為者」とみなされる。それゆえ、情報倫理学の抽象化レベルでは、道徳的行為者の範囲は拡大されることになる。しかし、一般に倫理学では道徳性は責任を含意するとされる。自らの行為について責任をとることができなければ、たとえ行為者とはみなされても「道徳的行為者」とみなすことはできない。情報倫理学においては、AMAのような人工的行為者も「道徳的行為者」とみなされるのに対し、責任をとることができないAMAを「道徳的行為者」とすることは、責任をとれないものに責任を負わせるという概念的誤謬を犯すことになる。批判されることになる。

このような批判に対して、フロリディは、道徳的行為者 (moral agent) であることと道徳的に責任をとれる行為者 (morally responsible agent) であることは区別されなければならないとする。何かを道徳的行為の道徳的源泉あるいは原因として「同定する (identify)」ことは、そのように同定されたものがその行為に道徳的責任をも負うか否か、どの程度負うのかと「評価する (evaluate)」こととは異なる。フロリディは「すべての規範的言説を責任分析に還元すべきである」という前提は受け入れがたいものであり、「法的誤謬」ですらあるという。責任の帰属という問題とは独立に道徳的行為者を同定する必要がある規範的言説というものがある。<sup>(13)</sup>

フロリディは、搜索救助犬という例をあげる。行方不明者の人命救助に貢献した犬は、道徳的に善い行為をなしたものと同定することができる。しかし、犬がその行為に対して道徳的責任があると評価することはできない。まだ責任ある行為者とはなっていない幼い子どもも、道徳的行為者と同定すべきであろう。犬や幼児は、道徳に関連するゲームに参加しており、ゲーム内の行為によって生じた事態の原因となりうる (accountable)。道徳的善もし

くは悪は、それらの行為者のためにおこった (account for) とみなされる。道徳的善悪は、責任をとることのできない「道徳的行為者」によってもたらされることがありうるのである。道徳的行為者は、道徳的事態の源泉となる可能性もしくは道徳的善悪の源泉と説明される可能性という意味においてアカウントビリティ (accountability) をもつ。そして、さらに行為の選択を正当化する理由を与える能力をもつという意味においてもアカウントビリティをもつ。このアカウントビリティは、説明する義務や責任という意味の「説明責任」のことではなく、道徳的善悪をもたらすことができるという生成可能性のことであり、なぜそのような事態が生じたかという理由を明らかにするものとなりうるという意味における説明可能性のことである。行為者として道徳的善悪を生成しその理由を明らかにできる能力 (accountability) は責任 (responsibility) とは区別されなければならない。フロリディによれば、責任がなく、道徳的アカウントビリティ (moral accountability) と道徳的行為能力 (capacity for moral action) だけがある場合でも、規範的行為を促進することはまったく理にかなっている。

しかし、道徳的行為者性を責任から切り離し、道徳的行為者の範囲を拡大することにはいかなる利点があるのか。道徳的善悪の源泉を人間に限定しないことによって、何か悪しき事態が生じたときに責任のある個人を探し出すための遡及が不要になることをフロリディはあげている。責任を帰属させるために、いかなる対価を払おうとも人間の道徳的行為者を同定するという必要はなくなる。人間以外の責任のない行為者を道徳的行為者として扱うことで、人間以外の文脈でよりよく道徳を論じることができる点に、このような考え方の大きな長所があるとフロリディはいう。それは、「責任」という概念がよけいなものであるということの意味しているのではない。むしろ、責任を切り離して定式化することで、道徳における責任を十分に明確にできるため、責任概念をさらに詳細に分析する必要性を明らかにするものであるという。

一般には「責任を問うことができる道徳的行為者がいて、規範的行為が存

在する」のであり、「責任を問うことができなければ、道徳的行為者ではなく、規範的行為は存在しない」というふうに分される。フロリディは、このような責任の有無による二分法を克服すべきであるとしている<sup>(14)</sup>。

### 2-3 AMA 批判

これに対し、西垣通はこうした二分法を堅持すべきであるとしてフロリディを批判する。西垣によれば、「自由意思 (free will)」をもっていることが人間が「道徳的主体」であるための「大前提」である。自由意思をもつことによって、人間は道徳的主体として判断をくだし行為を選択する。そして「その結果に『責任 (responsibility)』をとらなくてはならない<sup>(15)</sup>」。

では自由意思の有無はどのように確認すればよいか。それは「理論的自律性 (theoretical autonomy)」が認められるかどうかによってである。理論的自律性は、下等な生物でももっているような特性である。生物はオートポイエティック (自己-創出的) な存在であり、ある環境条件のもとで実行される行為の内部ルールを自分でつくりあげる。環境条件の変化により内部ルールは変化することもあるが、どのようなルールが出現するかは、「当の生物をふくめ誰にもわからない」。この「不可知性 (絶対的な予測困難性)」が、生物の場合には成立している。そのため生物は「他者の指令をまったく受けずに行動する」。これが「理論的自律性」である。これと区別されるのが「実践的自律性 (practical autonomy)」である。これは、道徳的判断をくだし「社会的な責任をとれる主体のもつ特性」とであるとされる。

西垣は、これらの概念の論理的包含関係を次のようにまとめている (A ⊃ B は A が B の必要条件であることを示す)。

予測困難性 ⊃ 理論的自律性 ⊃ 自由意思 ⊃ 責任 ⊃ 実践的自律性

AI の場合、どのような入力に対してどのような出力 (行動) が現れるのかが、設計者には基本的に分かっている。細部は不明でも、原則として行動が予測可能である。AI は、生物とはちがひ、人間が設計するアロポイエティック (他者-創出的) な存在であり、予測困難性が成立しない。それゆえ、

AIには理論的自律性も自由意思もなく、責任を問うこともできない。AIは、実践的自律性とは無縁であり、AIを「道徳的行為者」とみなすような考え方は認められないことになる。<sup>(16)</sup>

「フロリディはAIをふくめコンピュータ処理を責任という概念と短絡してしまう」と西垣は批判する。<sup>(17)</sup> 人工的行為者を道徳的行為者とみなす際に、「責任という概念の粗っぽいとらえ直しがおこなわれている」という。通常、規範に反した「行為者は裁判にかけられるなど、社会的制裁をうける恐れがある」が、「機械であるAIエージェントを牢屋に入れることはできない」。それゆえ人工的行為者は、責任とは無関係である。フロリディは accountability と responsibility を切り離し、「前者だけをとれるAIエージェントを責任主体と見なす」が、西垣は、それによって責任という概念が曖昧になり「下手をすると、AI活用とともに無責任行為が横行することになりはしないか」という懸念を示している。<sup>(18)</sup>

デボラ・ジョンソンもまた、機械を道徳的行為者とみなすフロリディの考え方を批判している。<sup>(19)</sup> AIエージェントが道徳的行為者とされると、設計者の責任が軽んじられることになるというのである。

ジョンソンによれば、コンピュータ・システムが作動しているときには、志向性の三要素がはたっている。設計者の志向性、使用者の志向性とコンピュータ・システムそのものの志向性である。システムは心的状態をもたないが、それを作り出した設計者の志向的作用をとおして志向性をもつようになる。三要素はすべて志向性と効果をもっており、使用者は行為を開始し、システムはそれを実行し、設計者はシステムのもたらす効果をつくりだす。

ジョンソンは、独立して自動的に作動する地雷の例をあげる。地雷は、入力があれば爆発し、なければそのままの状態を保つようにするという単純な志向性をもつ。戦争中に設置され、戦後何年も経ってから、この地雷を子どもが踏んで殺されてしまったとしよう。爆発による子どもの死という効果は、設置した使用者（兵士）の志向性からは時間的に隔たっており、設計者の志向性からは時間的にも空間的にも隔たっているが、長い時間が経過して

も持続している地雷の志向性によってもたらされたものである。兵士も設計者も子どもを殺すことは意図していなかった。このとき、フロリディアのように、地雷を帰責の対象ではないにせよ道徳的行為者であるとみなしてしまうと、誰が責任を負うべきかという問題が考慮されなくなるおそれがある。

地雷は自然物ではなく、道徳的に中立的ではない。子どもの死は、地雷の志向性の効果ではあるが、人間の志向性や行為と無関係なものではない。たとえ、兵士も設計者も子どもを殺すことを意図していなかったとしても、地雷の爆発によって子どもの死がもたらされた理由を説明するのは、彼らの志向性である。地雷は、子どもの死をもたらした行為の一部ではあるが、それだけで道徳的行為者とみなすことは適切ではない。人工物のもたらす効果は、人工物の志向性だけでなく、設計者の志向性、使用者の志向性を含む三要素によって説明されなければならない。ジョンソンは、これら三要素が一つになって行為をおこなうと考えるべきだとする。

それゆえ、コンピュータ・システムそのものは道徳的行為者とはなりえない。ただし、人間の道徳的行為者に何らかの効果をもたらし、人間の道徳的行為者性の結果となりうるかぎりにおいて、人間の道徳的行為者性の一部とはなりうる。その意味では、コンピュータ・システムは道徳的存在者 (moral entity) でありうる。しかし、単独では道徳的行為者とはなりえないとされる。<sup>(20)</sup>

岡本慎平は、「自律機械の行為を、『設計者—人工物—使用者』という三者の協同的な行為とみなすことは、我々の直観にも合致する」とし、「現実問題としての自律機械の問題に対応することが可能」な考え方としてジョンソンの見解を評価している。<sup>(21)</sup>

だが、フロリディアは、道徳的行為者性を責任から切り離し善悪の源泉を人間に限定しないことによって、何か悪しき事態が生じたときに責任のある個人を探し出すための遡及が不要になることを、むしろ「利点」であるとしていた。そして、「すべての規範的言説を責任分析に還元すべきである」という前提は「法的誤謬」であるとも述べていた。以下では、こうしたフロリデ

イの主張の意味を明らかにするために、瀧川裕英による「責任」概念の分析を参考としたい。

### 3 「責任」概念の分析

H. L. A. ハートは、「責任」という語の多義性を示すために次のような例をあげている。この例のなかに出てくる「責任」という語は、すべてが同じ意味で用いられているわけではない。<sup>(22)</sup>

船長として、Xはその乗客と乗組員の安全に対して責任があった。しかし、前回の航海で彼は毎晩酔いつぶれており、定員いっぱいの乗客を載せた船の遭難に責任があった。彼は正気ではないという噂が流れたが、医師は彼が自らの行為に対して責任があると判断した。航海の間中、彼はまったく無責任に行動していたし、これまでのさまざまな経歴からも彼が責任ある人物ではないということが明らかだった。船長は、遭難の責任は異常な冬の嵐にあると繰り返し主張したが、刑事訴訟において、彼は自らの過失行為に対して刑事責任があるとされた。それとは別に、民事訴訟において、生命と財産の損失に対して法的責任を負わされた。彼は今でも生きており、多くの女性と子どもの死に対して道徳的に責任がある。

法学において「責任」という語が多義的に用いられている原因として、瀧川は「継受元のドイツにおいて複数の語で言い表されていた概念を、日本では『責任』の一語に担わせたこと」をあげている。「日本語の『責任』に対応するドイツ語としては Schuld; Haftung; Zurechnung; Last; Verantwortung; Verantwortlichkeitなどを挙げるができる。同様に英語においても responsibility; liability; guilt; accountability; answerability; burden; imputationなどを挙げる<sup>(23)</sup>ことができる」。

ハートは、このような「責任」の多様な意味を、役割責任、因果責任、負担責任、能力責任の四つに分類することを提唱した。また、瀧川によれば、ケルゼンの責任概念に関する言明は、制裁、義務、責任という三概念に再構

成できる。瀧川は、ハートやケルゼンのそうした分類に検討を加えたうえで、責任概念を関与責任、負担責任、責務責任の三類型に分け、以下のよう<sup>(24)</sup>に分析している。

### (1) 関与責任

「関与責任」は、「過去の出来事に対する何らかの作用・生成・連関・関与を意味する概念」と定義される。関与責任を認めるということは、出来事の原因をある存在者に帰属させるということである。関与責任は「負わされたり引き受けられたり果たされたりする」ものではなく、あるかないかを認定されるものであるため「ある責任」とも呼ばれる。

さらに関与責任は、たんに「ある出来事の生成原因を意味する場合」と「行為者が一応の非難対象者であることを意味する場合」とに分けることができる。前者は「生成責任」、後者は「有責責任」と呼ばれる。ハートの例のなかの「船の遭難の責任は異常な冬の嵐にある」という船長の主張は、生成責任に関するものである。嵐に遭難の生成責任が認められる可能性はあるが、嵐が非難対象となることはありえない（生成責任があっても有責責任はない）。「船の遭難の責任は船長の行為にある」という言明もまた生成責任に関するものであるが、この場合には、船長は非難対象者となる可能性がある（生成責任があり有責責任もある）。

有責責任の有無と密接に関係するのが、生成責任が帰属されるのが「出来事」か「行為」かという違いである。「自動車事故の責任はブレーキの故障にある」という場合には、ブレーキの故障という「出来事」に生成責任が認められるが、「自動車事故の責任は運転者の前方不注意にある」という場合には、運転者の「行為」に生成責任があるとされる。前者は「出来事生成責任」、後者は「行為生成責任」と呼ばれる。出来事生成責任の探究は、原因の探求であり、原理上、無限に遡ることが可能である。事故原因はブレーキ故障、その原因は部品の劣化、その原因は劣悪な使用環境……というように限りがない。それに対して、行為生成責任の追求は無限に遡ることはなく、

その帰着点は「行為者」である。

ある行為者に生成責任（行為生成責任）が認められた場合、その行為者に有責任が帰属させられる可能性が生じる。出来事生成責任から行為生成責任が区別されるのは、それが有責任の帰属の前提として要請されるからである。しかし、有責任の帰属にはさまざまなレベルがあり、帰属条件によって有責任の有無は異なる。生成責任があれば有責任もある場合、预见可能性が必要とされる場合、意図が必要とされる場合、さらに責任能力が必要とされる場合など、その条件によって有責任の帰属は異なる。

このように関与責任は、生成責任と有責任に区別され、さらに生成責任は出来事生成責任と行為生成責任に区別される。そして、ある存在者がある事態に対して行為生成責任があると認められた場合に、その存在者に有責任が帰属される可能性が生じることになる。

## （2）負担責任

以上の関与責任に対して、「規範違反の結果として発生する負担あるいは不利益」を意味するのが「負担責任」である。負担責任については「責任を負う」という表現が用いられる。関与責任を「ある責任」と呼ぶのに対し、負担責任は「負う責任」と呼ぶことができる。法的には刑罰や損害賠償が、道徳的・社会的には非難や配慮・辞任・解任などが、負担責任の内容となる。ハートの例では、船長の刑事責任、民事責任、道徳的責任が、これにあたる。

有責任が規範に違反した行為者に帰属され、行為者に固着するのに対し、負担責任は、行為者に固着せず「転嫁可能」とであるという特徴をもつ。この「転嫁可能性」という特徴が最も明確に見られるのが損害賠償の場合である。自動車運転者の過誤による人身事故の有責任は、行為生成責任のある運転者に帰属させられる。しかし、この運転者が運送業者に雇われ業務を遂行中であつた場合には、負担責任である損害賠償責任は運行供用者である運送業者に帰属させられる。有責任は行為者である運転者に固着するが、

負担責任は運送業者に転嫁される。年少者が不法行為をおこなった場合、有責責任は行為者にあるが、その行為に起因する損害賠償責任は監督義務者である親に転嫁することができる。

同じ負担責任でも損害賠償については転嫁可能だが、刑罰は規範に違反した当の行為者以外には転嫁できないと思われる。しかし、負担責任は負担や不利益であって、概念上は行為者に固着しない。少なくとも近代以前においては規範違反者以外に刑罰が課されることがあったし、江戸時代まで存続していた「縁座」などの連帯責任は刑罰の転嫁可能性を示している。

ファインバーグは、転嫁可能性に関連して次のように述べている。<sup>(25)</sup>

負担は転嫁可能であるが、行為者性・因果性・過失は転嫁可能でなく、おそらく罪も転嫁可能ではない。

「負担」としての刑罰は転嫁可能である。しかし、「刑罰の一条件としての『罪』は有責責任であり、概念上転嫁不可能である」<sup>(26)</sup>。罪人であっても罰を受けないことも、罪人でなくても罰を受けることも、概念上は可能ということである。

負担責任として転嫁可能性をもつという点においては、損害賠償と刑罰の間に違いはない。しかし、未成年者の監督義務者に損害賠償させることは正当でも、刑罰を課すことは不当と考えられる。自動車事故を起こした運転者が賠償責任を負う場合の損害の補填を目的とした「責任保険」は現に存在している。自賠責保険がその典型である。しかし、運転者が刑事責任を問われた場合に、他者が代わりに刑事責任を負うような「非難保険」は存在しない。そのような制度は不当と考えられる。瀧川によれば、両者の違いは「直接責任」と「間接責任」の区別に対応する。直接責任は「有責責任のある者が負担することによって初めて意味を持つ負担責任」であり、非難や謝罪のように答責者と問責者の「直接的対面性の下でのみ意味を持つ」。それに対して、間接責任は、被害救済のための補償のように、答責者と問責者の「直

接的対面性を離れても意味を持つ負担責任」である。刑事責任は直接責任であり、加害者を離れると意味を失う。それゆえ、他者に転嫁するのは不当とされ、「非難保険」は存在しない。「社会的非難・道徳的責任・刑事責任」など負担責任は、「概念上」は転嫁可能である。しかし、「実践上」は転嫁することが不当とされるのである。<sup>(27)</sup> 責任が「ある」のに責任を「負わない」というのは矛盾していると受けとられがちだが、関与責任は「ある」が負担責任は「負わない」ということは、概念上は可能なのである。

### (3) 責務責任

「責務責任」は、「人がある立場・地位・役割を占めることで発生する何らかの責務」を意味する。「親は子供の養育に責任がある」といった場合であり、ハートの例では「船長は乗客と乗員の安全に責任がある」というのが、責務責任である。自動車の運転者の「安全運転義務」も責務責任とみなすことができる。

関与責任や負担責任が、すでに何らかの事態が生じた「過去に関する責任状況」において問題となるのに対し、責務責任は「未来に関する責任状況」において果たされるべき責務である。損害賠償は、負担責任と責務責任の両方の側面から把握する必要がある。すでに起こった交通事故に対する損害賠償責任は、負担責任である。損害賠償責任を負うことになった者は、それを履行する義務を課される。この損害賠償義務は責務責任である。この責務を果たさなければ、強制執行により負担責任を負うことになる。負担責任が、

<sup>(28)</sup>  
責任の概念

概 念		意 味
関与責任 (ある責任)	生成責任	出来事生成責任 行為生成責任
	有責責任	行為者が一応の非難対象であること
	負担責任 (負う責任)	負担 (あるいは不利益) …… 事後責任
責務責任	責務	…… 事前責任

あくまでも「事後責任」であるのに対して、責務責任は「事前責任」であり、何かが起こったときに責任を負うべき者をあらかじめ指示し、そのようなことが起こらないように予防措置をとることを期待するものである。

## 4 自動運転車の事故の責任

以上のような瀧川の責任概念の分析にもとづいて、以下では自動運転車の事故の責任問題について検討していく。

ある出来事の生成責任が帰属されるのが「出来事」か「行為」かという違いは、生成責任の遡及の問題と関係している。自動車事故が運転者の「行為」によって生じたとみなされれば、生成責任の追求はそれで終わる。しかし、事故がブレーキ故障という「出来事」によって生じたとされれば、何らかの「行為」に帰着するまで、その原因を遡らなければならない。自動車の保守・管理に問題があったことが判明すれば、保守・管理について責務責任のある人間の「行為」に生成責任が帰属させられる。そこに問題がなければ、製造上の欠陥という「出来事」に遡り、それをもたらした人間の「行為」に遡らなければならない。

製造物責任法（PL法）第3条により、製品の欠陥によって生じた損害については製造業者等に損害賠償を請求することができる。しかし、被害者には、欠陥と損害との因果関係を明らかにすることが求められる。自動車事故の被害者が、自動車メーカーの損害賠償責任を追及しようとするれば、損害の原因を自動車の欠陥にまで遡れることを立証し、製造者の「行為」に生成責任を帰属させなければならないのである。

しかし、自動運転車の事故原因の究明は、きわめて難しいものとなると考えられる。センサーなどのハードウェアの故障であれば特定可能かもしれないが、プログラムに設計上の欠陥があったとしても、それを特定するにはかなりの困難が予想される。さらに、走行時の深層学習などの処理を詳細に分析することはほぼ不可能とも思われる。<sup>(29)</sup> そのような「欠陥」が事故原因であ

ることを被害者に立証させるのは、妥当とはいえないであろう。

PL法は、損害という事態に対して行為生成責任があると認められた者に有責責任を帰属させ、賠償という負担責任を負わせるかたちをとっている。それに対し、自動車損害賠償保障法（自賠法）は、これとは異なるかたちで負担責任を負わせるものとなっていると考えられる。自賠法第3条は以下のように規定している。

自己のために自動車を運行の用に供する者は、その運行によって他人の生命又は身体を害したときは、これによって生じた損害を賠償する責に任ずる。

浦川道太郎によれば、このような責任規定のルーツは1909年に制定されたドイツの自動車交通法であり、これは運転者の運転過誤と自動車の構造上の欠陥・機能上の障害を原因とする事故のいずれに関しても、原則として自動車保有者（運行供用者）が責任を負うという考え方に立脚している<sup>(30)</sup>。自動車の運行の危険を支配でき自動車の運行から利益を得ている「運行供用者」が、交通事故が生じた場合に損害賠償義務を負うことになるという責務責任が定められているのである。事故が起こった場合に賠償責任を負うべき者をあらかじめ指示しているということになる。被害者は、事故の原因を立証する必要はない。事故原因が、運転者の過失でも自動車の欠陥でもなく他にあった場合には、免責されるが、それは加害者が立証しなければならない。これにより、交通事故被害者の救済・保護がはかられているのである。

自動車は馬車の発展形態であるが、馬車における事故の原因としては御者の操作の過誤と馬の暴走が考えられる。いずれにせよ、事故が起こった場合に馬車の保有者が賠償責任を負うという責務責任を定めておけば、被害者は損害賠償を請求できる<sup>(31)</sup>。御者の操作過誤は「行為」であり、生成責任をそれ以上遡ることはない。有責責任は御者にあるが、負担責任は所有者に転嫁される。

他方、馬の暴走は「出来事」とも「行為」とも解釈できる。黒田亘は「出

来事の因果連鎖」と「行為の因果連鎖」を対比し、両者は「同じ一つの世界についての二つの語りかた」であるとした。<sup>(32)</sup> ある人の「右手が上がる」という出来事は、その人が「右手を上げる」という行為でもある。生理現象としてみれば、その原因は腕の筋肉の収縮、中枢からの神経伝達、それをもたらした感覚刺激、さらにその原因と、「どこまでも、いわば不定無限に遡ることができる」。しかし、同じ動作を行為としてみれば、その因果連鎖は行為者自身から発現し、「別の行為者には結びつかない」。行為の因果連鎖を原因の方向にむかってたどれば、主体となる存在によって区切られるが、出来事の原因の連鎖は無際限に続いている。<sup>(33)</sup>

馬の暴走を出来事とみなしたうえで、暴走によってもたらされた損害の責任を追究しようとするれば、その原因を誰かの行為にまで遡って特定しなければならぬ。馬に薬物が投与されたことが原因であったと判明すれば、投与した人物が責任を負うことになるだろうが、そのように何らかの行為に帰着させることができるケースは稀であろう。それに対し、暴走を行為とみなせば、馬に行為生成責任を帰属させることができる。行為生成責任さえあれば有責責任を帰属させることができるという帰属条件であれば、予見可能性や意図や責任能力の有無とは無関係に、馬に有責責任があることになる。「暴走してはならない」という規範に違反した馬に有責責任は固着する。しかし、負担責任は転嫁可能であるから、馬の暴走行為に起因する損害賠償責任は、馬車の所有者に転嫁されることになる。馬の暴走を出来事ではなく行為とみなすことによって、「責任のある個人を探し出すための遡及が不要になる」ため、被害者の救済が容易になる。自賠法の責任規定のルーツをたどれば、このように解釈することができるだろう。

レベル4以上の自動運転車は、御者のいない馬車のようなものである。きわめて高度な調教を受けた馬が、御者がいなくても、乗客を乗せた馬車を目的地まで運行できるとしよう。フロリディは、人命救助をおこなった犬を善い行為をなしたものと同定できるとしているが、この馬も善い行為をなす「行為者」と同定することができるだろう。この馬が暴走し事故を起こせば、

悪をなした有責任はこの馬にある。フロリディのいうアカウントビリティは、道徳的事態の原因となりうるということであり、規範にしたがったものか否か説明されうるということである。その意味で、この馬はアカウントビリティをもつ。しかし、馬に損害賠償義務を課したり損害賠償責任を負わせたりすることはできない。そうした責務責任や負担責任は所有者に転嫁されることになる。アカウントビリティ (accountability) はあるが責任 (responsibility) のない道徳的行為者は、有責任はあるが負担責任を負わせることのできない存在と言いかえることができる。これらの責任概念を区別すれば、責任をとれないものに責任を認めることは概念的誤謬とはいえない。

自動車は馬とは違い、製造物である。事故の原因をその欠陥にまで遡ることができれば、製造者に生成責任を帰属させ、負担責任を負わせることができる。しかし、自動運転車のように、欠陥を立証することがきわめて困難となることが予測される場合、被害者に立証責任を負わせることはできないだろう。自動運転車に「行為者性」を認めれば、フロリディがいうように「何か悪しき事態が生じたときに責任のある個人を探し出すための遡及が不要になる」。自賠法が、事故について生成責任のある運転者とは別に運行供用者を賠償責任を負うべき者と定めているように、生成責任が自動運転車にあるとしたうえで、製造者の責務責任を定めることは可能ではないだろうか。その意味で、道徳的行為者の範囲を拡大することは、人間の道徳的行為者にまで遡らなくても人間に責任を負わせることができるという利点があるといえるだろう。AIの「暴走」による事故が起こったときに発生する損害に対する責務責任を定めておくことで、その原因究明がいかに困難であったとしても、被害者の救済・保護がはかれることになるのである。

## おわりに

トマス・ネーゲルは、「行為とはなにか」という問いは「クモの活動や、無意識であったり意図的でなかったりする人間の行動にもあてはまる」とい

<sup>(34)</sup>う。それゆえ、この問いは、自由意志の問題よりもはるかに多くのものに関わる。行為を自由意志のような内面からとらえるのではなく、主体と外部との関係において行為の正当化可能性を問うとすれば、行為は理由という観点からとらえられることになる。その場合、選択された行為が理由にもとづいて正当化できるかどうかの問題となる。フロリディがアカウントビリティと呼ぶのは、そうした正当化可能性のことであろう。相互性、自律性、適応性をそなえた遷移システムが道徳的善あるいは道徳的悪をもたらしたとき、それを、理由にもとづく正当化という観点からみることができ、システムの遷移を「理由の空間」において「行為」として把握できるのである。

他方で、人間の意図的な行為であっても、「因果の空間」において把握されれば、それは「出来事」である。「世界を客観的に思い描くとき、ある行為をわたしがすること——あるいは他人がある行為をすること——は消えてなくなるように思われる」とネーゲルはいう。「神経の刺激や化学反応、筋骨の動きの世界には、行為の居場所はない」し、そこに「存在するのは出来事だけである」。道徳的評価や帰責の対象とされる行為も、因果の空間では、たんなる身体運動にすぎない。人間であれコンピュータ・システムであれ、因果の空間においては、行為の主体とはなり得ない。因果の空間にあるのは出来事の因果連鎖だけであり、「そこには、『する』も見当たらなければ、いわんや『する主体』の影も匂いもない」<sup>(35)</sup>。

一つの世界について、理由の空間において行為の因果連鎖を描くか、それとも因果の空間において出来事の因果連鎖を描くか、二つの描き方がある。大森荘蔵の言葉をかりれば、一方には「行為の言葉」である「生の言葉」があり、他方には「物理学の言葉」である「事物の言葉」がある。そして「世界はこの二つの言葉によって『重ね描き』されるのである」<sup>(36)</sup>。AIをあくまでも因果の空間においてのみとらえ「事物の言葉」で語るのであれば、AIの責任が問題となることはありえない。しかし、理由の空間にAIを位置づけたとき、「生の言葉」によって、その行為や責任について論じることが可能となる。フロリディが指摘しているように、それによって責任概念をさらに

詳細に分析することが必要となるのである。

〔注〕

- (1) 高畑敬信「『社会受容性』の意味、本事業の目的、体制、活動成果」NBL (New Business Law) 1099号 (2017年), 13-14頁.
- (2) 清水和夫「『自動走行技術』の社会的導入の嬉しさと導入のロードマップ」NBL (New Business Law) 1099号 (2017年), 10頁.
- (3) 高度情報通信ネットワーク社会推進戦略本部・官民データ活用推進戦略会議『官民ITS構想・ロードマップ2021』, 35頁.
- (4) 深町晋也『緊急避難の理論とアクチュアリティ』(弘文堂・2018年), 243頁以下.
- (5) 今井猛嘉「自動走行に関与する者の刑事責任」NBL (New Business Law) 1099号 (2017年), 25-29頁.
- (6) 浦川道太郎「自動走行と民事責任」NBL (New Business Law) 1099号 (2017年), 30頁以下.
- (7) 今井・前掲注(5), 28頁.
- (8) Glanmarco Veruggio, Fiorella Operto and George Bekey, '80 Roboethics: Social and Ethical Implication of Robotics,' *Springer handbook of robotics* 2nd ed. (Springer, 2016).
- (9) Wendell Wallach and Colin Allen, *Moral Machines : Teaching Robots Right from Wrong* (Oxford UP, 2009), pp.69-71. 岡本慎平, 久木田水生訳『ロボットに倫理を教える』(名古屋大学出版会・2019年), 95-98頁.
- (10) *ibid.*, pp.206-207. 邦訳291-293頁.
- (11) Luciano Floridi and J. W. Sanders, 'On the Morality of Artificial Agents,' *Minds and Machines*, Vol.14, No.3, 2004, pp.349-379.  
Luciano Floridi, 'Information Ethics: Its Nature and Scope,' Jeroen van den Hoven and John Weckert (eds.), *Information Technology and Moral Philosophy* (Cambridge UP, 2008). ルチアーノ・フロリディ (西垣通訳)「情報倫理の本質と範囲」, 西垣通・竹之内禎編『情報倫理の思想』(NTT出版・2007年), 47-98頁.
- (12) フロリディ・前掲注(11), 邦訳69頁以下.
- (13) 同書, 73頁以下.
- (14) 同書, 75頁以下.
- (15) 西垣通, 河島茂生『AI倫理—人工知能は「責任」をとれるのか』(中央公

- 論新社・2019年), 61頁.
- (16) 西垣, 河島・前掲注 (15), 55頁以下.
- (17) 同書, 123頁.
- (18) 同書, 124頁.
- (19) Deborah G. Johnson, 'Computer systems: Moral entities but not moral agents,' *Ethics and Information Technology*, No.8, 2006, pp.195-204.
- (20) *ibid.*, pp.202-203.
- (21) 岡本慎平「人工的道徳的行為者への近年のアプローチの検討」西日本応用倫理学研究会『HABITUS』第16巻 (2012年), 71頁.
- (22) H. L. A. Hart, *Punishment and Responsibility*, (Oxford UP, 1968), p.211. 'The exceptional winter storms were responsible for the loss of the ship.' は, ふつうは「船の遭難の責任は異常な冬の嵐にあった」とは訳されず, 「責任」という語は用いずに「船の遭難は異常な嵐のせいであった」と訳されるだろう. 英語の 'responsibility' と日本語の「責任」との間には, ずれがある. 「責任」という語は, おもに「非難対象」に対して使用され, たんなる「原因」にはあまり用いられないが, 「せい (所為)」という語は, たんなる「原因」をさす場合だけでなく「非難対象」をさす場合にも使用される.
- (23) 瀧川裕英『責任の意味と制度—負担から応答へ』(勁草書房・2013年), 16-17頁.
- (24) 同書, 26-46頁.
- (25) Joel Feinberg, *Doing and Deserving: Essays in the Theory of Responsibility*, (Princeton UP, 1974), p.233.
- (26) 瀧川・前掲注 (23), 40頁.
- (27) 同書, 45-46頁.
- (28) 同書, 38頁.
- (29) 西垣, 河島・前掲注 (15), 196-197頁.
- (30) 浦川・前掲注 (6), 30頁以下.
- (31) 同書, 31頁.
- (32) 黒田亘『行為と規範』(勁草書房・1992年), 69頁.
- (33) 同書, 65-68頁.
- (34) Thomas Nagel, *The View from Nowhere*: (Oxford UP, 1986), p.111. 中村昇ほか訳『どこでもないところからの眺め』(春秋社・2009年), 181頁.
- (35) 大庭健「乖離していく主体—行為の因果説の帰趨」『専修人文論集』第93巻 (2013年), 94頁.
- (36) 大森荘蔵『新視覚新論』(講談社学術文庫・2021年), 271頁.

【付記】

本研究は JSPS 科研費 JP21K00015 の助成を受けたものである。