

エリック・ヒルゲンドルフ
「ロボットは有責に行為することができるか？
—規範的な基本語彙の機械への
転用可能性について—」

Eric Hilgendorf, Können Roboter schuldhaft handeln? Zur Übertragbarkeit unseres normativen Grundvokabulars auf Maschinen.

伊藤嘉亮

- I. 導入
- II. 文献紹介
 - 1. はじめに
 - 2. 概念の意味の規定方法
 - 3. ロボット工学の規範的理解
 - 4. 結論
- III. コメント
 - 1. フィクションとしての自由意思
 - 2. ロボットに対する刑罰
 - 3. ロボットの人格性

I. 導入

現代社会においては、自律的ないし準自律的に周囲の環境を認識・判断し、それに応じて作動するロボットが我々の日常生活に浸透し始め、利便性が飛躍的に向上する一方、技術的發展に伴う新たな問題も生じるようになった。例えば、2016年6月30日にアメリカのテスラ社製の自動車が自動走行中に起こした死亡事故(運転手の死亡)をその一例として挙げることができよう⁽¹⁾。それでは、法は、自律的に作動する機械が関係するこういった事案にどう対応すべきであろうか。この問題に(解釈論

として、あるいは立法論として)結論を下すには文理の垣根を超えた学際的な共同研究が不可欠になるが、その前提作業として、まずは法律学の枠内で議論し、問題を整理・検討しておく必要がある。

ここで考察すべき法的問題は、自律的・準自律的ロボットを製造、販売、所有、使用した自然人(あるいは法人)の法的責任⁽²⁾とロボット自体の法的責任に区分できる。そのうちの後者を検討したのが、ここで紹介するエリック・ヒルゲンドルフの論文である(『ロボットと法(Robotik und Recht)』の第1巻である『人間と機械の超越(Jenseits von Mensch und Maschine)』119頁以下に所収⁽³⁾)。この問題は、少なくとも今現在は、喫緊の検討課題とはいえないし、そもそもロボットに法的責任を認める可能性ない意義に疑問を呈する論者も多いところであろう⁽⁴⁾。しかし、法人の場合も現在はその犯罪能力を肯定するのが通説の立場であるし、また、(我が国の刑法は法人処罰を特定の犯罪類型に限定しているが、)故殺罪にまでその射程範囲を拡張する立法例(イギリスにおける法人故殺法⁽⁵⁾)もあるくらいだが、こうした状況は、法人の犯罪能力を否定するのが当然

-
- (1) 朝日新聞平成28年7月2日朝刊。もっとも、本件事故車が、必要に応じて運転手が自ら運転操作を行わなければならない「レベル2」に分類されるものであった点には留意すべきである。
 - (2) これについては、*Eric Hilgendorf*, *Recht und autonome Maschinen - ein Problemaufriß*, in: *Eric Hilgendorf/Sven Hotitzsch* (Hrsg.), *Das Recht vor den Herausforderungen der modernen Technik*, 2013, S. 11 ff. 本論文を紹介するものとして、富川雅満「エリック・ヒルゲンドルフ『法と自律型機械—問題概説』」千葉大学法学論集31巻2号(2016年)135頁以下)。
 - (3) *Eric Hilgendorf*, *Können Roboter schuldhaft handeln?: Zur Übertragbarkeit unseres normativen Grundvokabulars auf Maschinen*, in: *Susanne Beck* (Hrsg.), *Jenseits von Mensch und Maschine*, 2012, S. 119 ff.
 - (4) ロボットの刑事責任を否定するものとして、*Jan C. Joerden*, *Strafrechtliche Perspektiven der Robotik*, in: *Eric Hilgendorf/Jan-Philipp Günther* (Hrsg.), *Robotik und Gesetzgebung*, 2013, S. 195 ff. (本論文については今井康介「ヤン・C・イエルデン『ロボット工学の刑法的諸観点』」千葉大学法学論集31巻2号(2016年)111頁以下)。
 - (5) 今井猛嘉「イギリスにおける法人処罰——2007年法人故殺法」志林106巻3号(2009年)145頁以下参照。

であった時代⁽⁶⁾には想像もできなかったのではないだろうか。そうであるならば、近い将来、ロボット自体の犯罪能力・受刑能力を肯定する時代が到来することも考えられないではない。それゆえ、本論文の問題提起は、我が国にとっても決して他人事ではなく、今後の刑法学における重要な一領域を形成するものと思われる。

II. 文献紹介

1. はじめに

行為(Handlung)、帰属(Zurechnung)、答責(Verantwortung)、あるいは責任(Schuld)といった概念は、不可侵かつ不可変であって、ロボットに適用されることはないと考えられてきた。たとえロボットが自律的(あるいは準自律的)に作動するとしても、行為することはないし、ロボットにある事象が帰属され、あるいは答責・責任が認められることはない。このように考える立場は、ロボットが関係する問題に法が対応し、ロボットが惹き起こした損害についてロボット自体に責任を問う可能性をア・プリオリに排除してしまうものである。しかし、人間とロボットの間になんかそのような違いを認めることに理由はあるであろうか。人間への適用を予定した概念をロボットに適用することに躊躇するのは、「我々人間は唯一無二の存在である」という考えに固執するからに過ぎないのではないだろうか。本稿では、上記の諸概念の意味内容が不可変のものであるのか、それとも拡張的に解釈され、あるいは新たな意味が付与されうるものなのかを問うことにする。

2. 概念の意味の規定方法

概念の意味は語法(Sprachgebrauch)を通じて規定される場合と、定義(Definition)を通じて規定される場合がある。

一般的な語法において、記号(Zeichen)の意味は、使用者が一方的に

(6) 法人処罰をめぐる議論については、樋口亮介『法人処罰と刑法理論』(東京大学出版会、2009年)1頁以下。

定めうるものではなく、名宛人がそれをどのように理解するかが重要となる。つまり、言語共同体が正しいと判断した解釈が記号の意味になるのである。しかし、同一の言語共同体の中においても同じ表現が常に同じ意味で用いられているわけではないし、文言の用法が不明確である場合も稀ではない。それゆえ、概念を明確にする方法を模索しなければならないわけだが、概念に明白な意味を与える第一の方法が注釈(Explikation)である。注釈とは、ある概念の意味を具体的に摘示し、当該概念の個々のメルクマールを分析することである。例えば、「奪取(Wegnahme)」という概念は、「他人の占有の侵害および新たな占有の確立」と注釈されている。これに対して、第二の方法が定義である。定義は、不明確な文言を使用する場合にも、あるいは定義によってはじめて意味が付与される新たな概念を定立する場合にも用いられる。定義することで「代用語(Platzhalter)」が設定され、冗長な表現が簡略化されることもある(省略機能)。その例としては、「この法律の意味において、1. 親族とは、a)直系血族、直系姻族、配偶者、婚約者、兄弟姉妹、兄弟姉妹の配偶者又は配偶者の兄弟姉妹をいうが、その関係の基礎となる婚姻がもはや存在しない場合、又は、血族関係若しくは姻族関係が解消している場合も親族とする、b)扶養親及び扶養子をいう」とするドイツ刑法11条1項1号を挙げることができる。

定義は、意味を記述するのではなく、取り決めによって言語の新たな意味を確定するものである。それゆえ、定義を正しいものと誤っているものに区分することはできず、相当性(合目的性)の程度に差があるに過ぎない。そうすると、言語の意味は、不可変のものでないことになる。つまり、言語は、歴史的過程の中で変化するものであって、定義を通じて意味を変えうるし、まったく新しい意味が与えられることもあるのである。また、まさに法という領域においては、新たな解釈を通じて、概念をニーズの変化に適合させ、断続的に変化させていくのである。

3. ロボット工学の規範的理解

かくして、行為、答責、あるいは責任といった基本概念のロボットへの適用は否定されないことになる。現在の語法は決して不可侵ではな

く、変更されうるのである。一般的な語法が歴史的な流れの中で変わっていくのに対して、専門分野における語法は、専門用語を用いる目的に応じて創設される定義を通じて修正されていく。

(1) 機械の行為について

「行為(Handeln)」を意思によって制御された態度(Verhalten)と理解する場合、少なくとも、ロボットが外観的に認識可能な身体動作という意味での「態度」を取ることは想定できる。しかし、確かにロボットが四肢を動かしたり、言語音を発したり、周囲の環境に影響を与えることはあるが、そうした態度を「意思に制御されたもの」といえるかは明らかでない。機械をプログラミングした人間、または使用した人間の意思に着目することも可能であるが、このように考えると機械ではなく「背後に存在する」人間に機械の態度を帰属させることになる。機械が行為するというには、機械それ自体の中に人間の意思に相当するものを見出す必要がある。ここでは機械を制御するプログラムが考えられうる。最も単純なプログラムは、「～の場合」→「～する」の組み合わせにより構築され、ロボットは、状況Aの場合にはA'という態様で作動し、状況Bの場合にはB'という態様で作動することになる。このことは、意思により制御された態度の可能性を妨げるものであろうか。人間も、また、原則としては一定のルールに沿って、あるいはそれに導かれながら態度を取っていることに鑑みると、人間と機械を類比し、機械の場合にはプログラムによる「意思制御」と表現することも正当化されると思われる。

しかし、この類比は「自由意思の問題」を考慮していない。自由意思においては「人間は彼の態度について(完全には)決定されておらず、自由に決定しうる」ことが前提にされているが、この命題はとりわけ刑法においてこそ重要になってくる。もっとも、実務は、意思の自由に関して結論を留保しており、意思制御を否定する要件(抵抗不能な強制)に着目するだけである。そうであればロボットの行為も同様に考えることができる。つまり、機械も行為能力を有するとして、抵抗不能な強制のような要因が認められる場合にのみ行為は否定されると考えるのである。

(2) 機械の答責について

「負責(Haftung)」とも表現される答責のうち、実務上最も重要なのは

民法上の責任(例えば、損害賠償)であるが、刑法上・警察法上の責任もある。

まずは、自然人だけが責任主体になりうるかが問われる。民法上企業に責任が肯定されうることは一般に認められており、また、国によっては企業の刑事責任さえ認められていることからすると、自然人以外に責任を認めることは可能である。原理的には人間以外のアクター(nicht-menschliche Akteure)に責任を肯定することは法の自由である。それゆえ、少なくとも概念的にはロボットの「責任」を肯定することに問題はない。

他方で、責任を認めた場合の法効果をどうするかが問題となる。ロボットは財産を所有できないため、ロボットに損害賠償を言い渡すことはできないとも考えられる。しかし、この問題は、ロボットの使用者に対して、ロボットが損害賠償を義務付けられる場合のために保険契約をするよう法律上義務付けたり、あるいは法律により使用するそれぞれのロボットにつき一定額を銀行口座に払い込まなければならない、万が一の場合のロボットの損害賠償はそこから支払われると規定することで解決できる。結局のところ、民法上の責任をロボットに認めることに原理的な問題はないといえる。

(3) 機械の責任について

刑法上の責任を認めるには、行為能力だけでなく責任も前提とされる。責任は一般に「非難可能性(Vorwerfbarkeit)」と解されているが、1952年のBGH判例では意思の自由に関する大刑事部の所見が示されている。

「責任という無価値評価によって、行為者は、彼が法に従った態度を取ることができ、適法を選択するよう決定できたにもかかわらず、法に従った態度を取らず、不法を選択するよう決定したことを非難される。責任非難の内在的根拠は、人は道徳的に成熟するや否や、自由かつ道徳的に自己決定する能力が病的現象により一時的に鈍麻していない、あるいは長期間阻害されていない限り、自由、責任的、かつ道徳的に自己決定する能力を有し、それゆえ適法に従い、不法に抗う決定をし、彼の態度を法的当為の規範に適合させ、法

的に禁止されたことを回避することができる、という点にある」⁽⁷⁾。

それゆえ、BGHは、意思の自由について、「人は、病的変化に妨げられない限り、適法と不法、善と悪の間で自由に決定することができる」という立場を示しており、そのような意思の自由は、責任刑法の本質的要素であって、ドイツ刑法において今日まで広く受容されている。そして、こうした状況を踏まえると、一見したところ、ロボットの責任は認められないようにも思われる。

しかし、21世紀の脳研究が意思の自由に疑問を示したことで、法律学は、意思の自由をフィクションと考える立場に立ち返ることになった。意思の自由がフィクションに過ぎないのであれば、フィクションをロボットにまで拡張することも可能なのではないだろうか。フィクションは現実を反映するものではない。時に現実と相反することすらある。フィクションは合目的性という観点から定立されるのだから、ロボットの自由意思をフィクションとして導入することも、相応の必要性が実際に認められる限りは可能になろう。しかし、現時点ではそのような必要性は認められないと思われる。人間の処罰との類比でロボットの「処罰」を考えることは、法学というよりはSFのテーマである。ロボットの態度を変更させるにはプログラムを書き直すのが最も簡単な方法であるが、例えば自己学習システムが故にプログラムへの介入が功を奏さない場合には、機械の態度を変更させうる「害悪の付加」が検討されることになる。しかしながら、そのような害悪の付加のために幾重もの意味を持つ「刑罰」を用いる必要はない。

ところで、法律学上、実は意思の自由にはそれほどの意味はない。行為者の「責任」は、①責任能力(ドイツ刑法20条、21条)、②特別の責任要素(例えば、ドイツ刑法315条cの「無謀」要素)、③責任過失・故意、④不法の意識、⑤免責事由(例えば、免責的緊急避難)の不存在により肯定されるが、この中に意思の自由が問題とされる場面はない。これは、ドイツ刑法が即物主義的に考え、意思の自由のような形而上学的な問題を考慮の外に置いているからである。したがって、機械が以上の5つの要

(7) BGHSt 2, 194 [200 f.].

件を充足するのであれば、ロボットに責任を認めることも可能であると思われる。このような考え方は、機械は所為の不法を理解する能力を「実際には」有していない、「実際に」無謀に行為するわけでない」と批判されるかもしれない。しかし、これらの批判に対しては、第一に、人間の場合もその「実際には」が何を意味するか明らかでない、第二に、上述の語彙を適用する際には単にフィクションが問われているだけであるが、それは、実務上の理由により導入されたものであって、そこで示された内的・心理的事象を「実際に」正しく描写するよう要求するものではない、と反論することができよう。

4. 結論

本稿は、「行為」や「答責」といった概念を自律的な機械に転用しうると結論付けた。これらの概念をロボットに適合させる必要がある限り、法の枠中であれば概念の拡張解釈により必要な修正を施すことは可能である。形而上学的命題が関係する場面では難問も生じうるが、意思の自由という問題は法実務上ほとんど顧慮されていないし、また、人間の自由意思も(国家にとって必要な)フィクションであることを指摘すれば十分であろう。以上のように考えると、ロボットの行為、答責、責任を認めることも可能になる。

Ⅲ. コメント

本論文は、結論として、ロボットの刑事責任の可能性を肯定するものである。この結論を支える論拠は、①「行為」や「責任」といった刑法上の基本概念がその時々¹の社会的ニーズに応じて変化・修正されうること、②自然人の場合であっても、責任の前提である「自由意思」がフィクションである以上、ロボットの責任をフィクションとして認めることも可能であることの二つに大別できる。以下では②を検討し、ロボットの刑事責任の可能性について考察することにする。

1. フィクションとしての自由意思

自然人の自由意思もフィクションに過ぎないとする立場は、我が国においても見られるところである。例えば、「自由意思と他行為可能性を完全には証明できず、さらに、裁判の場で確定できないという事実を否定することはできない」が、「刑法を保安処分法に原理的に解消させないため」に自由意思を「規範的要請」と理解する内藤謙⁽⁸⁾、「責任の有無と程度を決めるための基準としての自由と可能性は、経験的事実ではなく、規範的要請ないし仮設(仮『設』であって、仮『説』ではない)として前提に置かれる」とする井田良⁽⁹⁾、「責任は、実体としての責任ではなく、社会的虚構とさえいえるのであ」って、「このような社会的虚構がなければ社会は存立することはできず、その結果、責任および刑罰は、自由意思を仮設として前提としなければ、存立不可能な制度と理解せざるを得ない」とする高橋則夫⁽¹⁰⁾の主張がそれである⁽¹¹⁾。しかし、仮に自然人の自由意思が規範的要請に基づく法的フィクションであったとしても、そこから直ちにロボットの意思(あるいはプログラム)の自由をも法的フィクションとして肯定してよいことにはならない。既に佐伯千仞が指摘していたように、刑法上の責任および自由意思が一種の擬制であったとしても、それは、「社会生活の現実とその基礎を有するもので決して架空的なものではない」からである⁽¹²⁾。自然人の場合も、何らかの社会的基礎が存在するからこそ、自由意思の仮定が許されるようになると考えるのであれば、ロボットの責任および自由意思を仮定する場合にも同様の社会的基礎が要求されなければならないことになる。

しかし、そもそも自然人の自由意思を許容する社会的基礎とは何であ

(8) 内藤謙『刑法講義総論(下) I』(有斐閣、1991年)786頁。

(9) 井田良『講義刑法学・総論』(有斐閣、2008年)358頁。

(10) 高橋則夫『刑法総論〔第2版〕』(成文堂、2013年)334頁。

(11) その他、斉藤誠二「刑罰の正当化の根拠をめぐって」『宮澤浩一先生古稀祝賀論文集第二巻』(成文堂、2000年)25頁以下、安田拓人『刑事責任能力の本質とその判断』(弘文堂、2006年)1頁以下、山中敬一『刑法総論〔第3版〕』(成文堂、2015年)626頁も参照。

(12) 佐伯千仞『刑法における期待可能性の思想(下巻)』(有斐閣、1949年)614頁。

ろうか。この点、増田豊は、「国家的なフィクションとしての自由を通じて行為者個人に責任非難を帰属し、この者に刑罰を科すといった構想は、およそわれわれの相互主観的な承認を得られるものではなく、正統化されるものではない」として批判し⁽¹³⁾、行為者への責任帰属を認めるには「責任を問うための相互主観的・コミュニケーション的条件が整備されていなければならない、つまり「責任非難の基礎となる自由も、たとえコンヴェンショナルなものであり、社会的・文化的構成物であるとしても、国家的な擬制に先行して相互主観的に構成された現実でなければならない」と説いている⁽¹⁴⁾。ここでは、要するに、責任を問う側と問われる側が相互に対等な立場にあること、つまりは互いの立場の交換可能性が責任(および自由意思)の前提として要請されているものと思われる。更に言えば、対等な立場にある者同士がその共通認識として——それ自体は証明することのできない——自由意思を指定するからこそ、それをフィクションとして仮設することが許されるのであろう。それではロボットの責任・自由意思を論じる前提条件、すなわち責任を問う側(自然人)と問われる側(ロボット)の対等性は認められるであろうか。少なくとも現時点では、我々人間とロボットの間に関主観的な議論を展開するための基盤は存在しないと言わざるを得ない。したがって、自然人の場合に自由意思の存在をフィクションとして認めるからといって、この論理をそのままロボットの場合に導入することはできないと思われる。

2. ロボットに対する刑罰

仮に自由意思の問題を措くとしても、ロボットに刑罰を科す必要はあるであろうか。法人の場合、「企業活動として公害犯罪や経済犯罪が行われることがあるのであり、その処罰が個人に及ぶにすぎないとすると、企業の従業員にすぎない個人に刑事責任を押しつけることによって、企業が刑事責任を受けないで莫大な利益をあげていくのを抑制でき

(13) 増田豊『規範論による責任刑法の再構築——認識論的自由意志論と批判的責任論のプロジェクト——』(勁草書房、2009年)477頁。

(14) 増田・前掲注(13)478頁。

ないのであれば、刑事制度は公正かつ有効とはいえない⁽¹⁵⁾ことから、法人処罰の必要性は認められて然るべきである。これに対して、ロボットの不処罰により不都合な事態が生じるとは考えにくい。この点、ヒルゲンドルフは、ロボットに民事制裁を課す前提として、ロボットの使用者に万一の場合に備えて保険契約を結んだり、あるいは銀行口座に一定額を預金しておくよう法律で義務付けることを提案している。しかし、ロボットに制裁を課さずに、直接、ロボットの製造者・使用者に損害賠償を求めれば事足りるであろうし、また、刑事制裁が問題になる場合も製造者・使用者に罰金刑を科せば十分であると思われる。

確かに、損害の発生にロボットが介在する場合、ロボットを製造・使用する自然人の過失(予見可能性・注意義務違反)や因果関係の認定が困難になることは想像に難くない⁽¹⁶⁾。しかし、だからといってロボットに対する民事・刑事制裁を安易に肯定するわけにはいかず、それには慎重な検討を要する。というのも、ロボットに制裁を課すといっても、結局はその背後者(自然人)に何らかの不利益を負担させるものであって、それは、過失や因果関係の立証のない法的責任を——いわば「ロボットに対する制裁」という裏道を通じて——当該自然人に認めることを意味するからである。

あるいはロボット自体に財産を所有する権利を認め、その財産に対して制裁を課すということも考えられる。しかし、このように考えるには、その前提として、ロボットを権利の主体として認める必要がある。そして、それは、ロボットを——我々人間と同格の——「自律的な人格」とみなすことを意味する。しかし、ロボットに人格性を認めるのであれば、ロボットの所有権を認め、財産刑を科すことが可能になる一方、人格の不可逆的な変更を強いるロボトミー手術が人間に対する制裁としては許容されないのと同様、人格を認めうる程の自律的なロボットのプログラムを強制的に変更することも許されないし、また、死刑に相当する廃棄・破壊も(原則として)禁止されることになるだろう。

(15) 山中・前掲注(1)196頁。

(16) この点については、富川・前掲注(2)116、114頁以下参照。

3. ロボットの人格性

ロボットに対する刑罰の是非を考えるにあたって問題になるのは、刑罰の必要性だけではない。前述のように、ロボットに責任を問い、刑罰を科すということは、同時に当該ロボットを一人の人格として認めることになる。したがって、そもそもロボットに人格性を認める余地があるのかを問わなければならない。

この点、カントは、その定言命法において、「人間ばかりでなく、およそいかなる理性的存在者も、目的自体として存在する、すなわちあれこれの意志が任意に使用できるような単なる手段としてではなく、自分自身ならびに他の理性的存在者たちに対してなされる行為において、いついかなる場合にも同時に目的と見なされなければならない」〔傍点著者・訳者〕⁽¹⁷⁾と説いていた。カントの時代においては、人間以外に理性的存在者である人格を観念することはおよそ不可能であったといえよう⁽¹⁸⁾。それでは我々は現代に至り、ついに人間以外の理性的存在者「ロボット」を見出すことになったのだろうか。

この問いは、現在の社会状況を前提にする限り、否定されることになる。ヒルゲンドルフ自身が指摘しているように、「ヒューマンズミ的な観点からは、機械は、我々が我々の世界と生活をより生き生きと『人間らしく』形成するための手段である」⁽¹⁹⁾に過ぎず、それ自体を目的に

(17) イマヌエル・カント(篠田英雄訳)『道徳形而上学原論』(岩波書店、1960年)101頁。

(18) アルトゥール・カウフマンも、「動物は、調教され、折檻され、譴責され、そして無害なものとなされるが、犯した責任に贖罪することはできない。刑罰を受けることができるのは、自己の行為に対して責任を負うことができ、かつ刑罰の意味を有責な悪行に対する当然の罪悪苦(Übelleiden)として精神的に理解することのできる本質存在だけである」、つまり「責任と刑罰は、人間存在それ自体においてのみ基礎づけることができるのであり、それらは、人間一般をはじめ人間たらしめるものの中に自己の根拠を有しなければならず、したがって、人間の不変の本性、つまり人間に、そして人間だけに固有の存在態様に関係づけられる」として、人間のみが責任と刑罰を負いうる存在であるとしている(アルトゥール・カウフマン(甲斐克則訳)『責任原理』(九州大学出版会、2000年)175頁)。

存在するものではない。理性的存在者である人格を何らかの「手段」として使用することは許されないわけだが、我々人間の利便性を向上させるための手段としてロボットを使用している現状からすると、人工知能の開発が進み、ますます自律的に作動するようになったとしても、ロボットに人格性を認める素地は未だ整っていないと言わざるを得ない。ロボットに対する責任・刑罰を論じるには、ロボットが自己目的な存在者になるのを待たなければならず、それはもうしばらく先の時代の話であると思われる。

(19) *Hilgendorf* a. a. O. (Anm. 2), S. 35 f.