

研 究

ロボットの処罰可能性を巡る議論の現状について

Zum heutigen Diskussionsstand um die Strafbarkeit der Roboter

根 津 洸 希*

目 次

1. はじめに
2. ロボットの処罰可能性を巡る見解の対立
 - (1) 積極説—— Hilgendorf の見解
 - (2) 消 極 説
 - (3) 中 間 説
3. 若干の検討
 - (1) 積極説について
 - (2) 消極説について
 - (3) 中間説について
 - (4) 共通の認識？
4. 私見の方向性
5. おわりに

1. はじめに

我が国では2020年に完全自動運転自動車の部分的導入、2025年に一般市場販売が検討されており¹⁾、この完全自動運転自動車についての適切な法

* 中央大学大学院法学研究科博士課程後期課程在学中

1) 内閣府「戦略的イノベーション創造プログラム（SIP）自動走行システム研究開発計画」http://www8.cao.go.jp/cstp/gaiyo/sip/keikaku/6_jidousoukou.pdf。
本稿が想定する「完全自動運転自動車」とは、いわゆるレベル4（加速・操舵・制動を全てドライバー以外が行い、ドライバーが全く関与しない状態）を

規制が喫緊の課題となっている。ロボットや人工知能 (AI) の法規制というのは決して未来の話ではなく、せいぜい数年後生ずるであろうほどに現実味を帯びてきている。現に、アメリカの自動車メーカーであるテスラ・モーターズの自動運転自動車が、テスト中に死傷事故を生じたというニュースは記憶に新しい²⁾。

このように、ロボットは我々の社会に進出してきており、当然人間との交流の機会も今後増えていくこととなろう。そうなれば、ロボットが人間を傷付けてしまいうるということも想像に難くない。その際、ロボットの刑法的問題を論じるにあたって、以下のような設例を挙げてみたい。完全自動運転自動車が、誤作動により歩行者に衝突して傷害を負わせた、という事例を考える。この際、当該誤作動の原因として考えられうるのは、①製造者が製造時に誤ったプログラミングをした、②販売時にディーラーなどでの納車整備の際に瑕疵があった、③購入した利用者が不適切な情報を入力した、といった点である。よりこの問題を困難たらしめるのは、当該完全自動運転自動車に自律学習機能を有する AI が搭載されている場合である。AI が走行体験から情報を獲得し、場合によっては自らのプログラムを改変・最適化することも考えられるのである。この自律学習機能の複雑さによって、当該誤作動の原因を立証することはほぼ不可能となろう³⁾。

指す。本稿の問題意識は、もはや人間が操縦できないような乗物やロボットが関与した犯罪についての責任の所在にある。

2) New York Times “Inside the Self-Driving Tesla Fatal Accident”, http://www.nytimes.com/interactive/2016/07/01/business/inside-tesla-accident.html?_r=0.

3) *Susanne Beck*, Grundlegende Fragen zum rechtlichen Umgang mit der Robotik, JR 2009, S. 225 ff. なおこの問題を避けるために自律学習・自動更新機能を搭載させないという選択肢もありうるが、たとえば道路工事の情報を更新せずに、道路の穴に落ちてしまう危険性があるのであれば、自動運転機能としては実用に耐えない。トレーサビリティの観点からシステム更新・改変の際には随時システムログを残し、誤作動の原因を突き止められるようにするという方法もあるが、そもそもその膨大なシステムログを解析し、原因を突き止めることのできる人間がそもそも存在するか、という問題もある。

また、立証の問題をクリアしたとしても、さらなる問題に直面することとなる。たとえば完全自動運転自動車が、そのAIの「自らの意思で」歩行者に衝突し、傷害を負わせた場合を考える。この事例では、少なくとも製造・販売時の最先端の知見によれば、プログラミングに欠陥はなく、販売時のチェックにも一切の瑕疵が認められなかったということを前提とする。すなわち、人間側に一切の落ち度がなかったとして、当該自動運転自動車が自ら人間への害意を生じ、傷害を負わせた場合、どのような解決の可能性がありうるであろうか。

可能性の一つとして、ロボットやAIの文脈においては過失概念を修正するという方法がありえよう⁴⁾。これによれば、当該自動運転自動車を製造したメーカーは、「いずれ事故は生じるであろう」ということは認識しているのであるから、この抽象的な予見で足りることとし、この予見可能性を根拠として過失責任を問う、ということになる。

しかしながら、これはメーカーにとって酷である。製造・販売時に欠陥・瑕疵はないにもかかわらず、「何かしらの事故」という非常に抽象的な予見だけで過失責任が肯定されてしまえば、そもそもそのような技術を研究・開発すること自体に抑制的な作用がはたらいってしまうこととなる。刑法が技術発展の妨げとなってしまう、これは我々にとっても好ましいことではない。

他方、上記の事例でメーカーの責任を一律して否定した場合、人が傷害を負うという重大な結果が生じているにもかかわらず、多くの事例において誰も責任を負わないこととなってしまう（いわゆる答責の間隙）。この

4) *Susanne Beck*, *Intelligente Agenten und Strafrecht. Fahrlässigkeit, Verantwortungsverteilung, elektronische Personalität, Studien zum deutschen und türkischen Strafrecht - Delikte gegen Persönlichkeitsrechte im türkischen-deutschen Rechtsvergleich* (Band 4), Ankara 2015, S. 179 ff. なお本論文は既に千葉大学共同研究チーム「ロボットと刑法」にて紹介した。拙稿「スザンネ・ベック『インテリジェント・エージェントと刑法—過失 答責分配 電子的人格』」千葉大学法学論集31巻3・4号頁以下。

ジレンマを解決するためには、いかなる当事者のいかなる責任分配がなされるべきであろうか。

上述のような問題関心から本稿は、ロボットやAIが何らかの損失をもたらした場合、責任の分配はいかになすべきかという問いに対し、諸外国においける「ロボットは法的責任を負うるか」という議論を参照する。我が国における来るべき時代の責任分配のありかた⁵⁾を探る第一歩として、以下ではその議論の状況を概観し、若干の検討をくわえるものとする。

なお本稿における用語法として、「ロボット」とは「人工知能により自律的に学習、行動する有体の機械」、「インテリジェント・エージェント」とは「人工知能により自律的に学習、行動する無体のプログラム」を指すこととし、両者を主に有体か無体かで区別する。しかしながら、インテリジェント・エージェントにより制御される機械は最早無体のプログラムにすぎないとはいえず、両者の境界は必ずしも一義的に決定できないと思われる。また本稿の問題関心からすれば、「人工物であっても自律的である」という場合、責任を肯定できるか」という点に重点があるのであって、有体か無体かの差異は本質的な点ではないため、原語の都合上訳し分けることもあるが、本稿では両者をほぼ同義のものとして扱う⁶⁾。また、

-
- 5) 自動運転自動車が交通事故を起こした場合の、運転者と自動車メーカーの間の責任分配に触れる我が国の議論として、今井猛嘉「自動化運転を巡る法的諸問題」IATSS Review Vol. 40, No. 2, 56頁以下。同様に自動運転自動車の刑法上の問題を扱うものとして、岡部雅人「自動運転車による事故と刑事責任—日本の刑法学の視点から—」愛媛法学会雑誌43巻3・4合併号1頁以下、池田良彦「自動運転走行システムと刑事法の関係」自動車技術 Vol. 69 No. 12, 33頁以下、中山幸二「自動運転をめぐる法的課題」自動車技術 Vol. 69 No. 12, 39頁以下。民事責任との関連で論じるのは、中川由賀「自動運転導入後の交通事故の法的責任の変容～刑事責任と民事責任のあり方の違い～」中京ロイヤー Vol. 25, 41頁以下。
- 6) 以下に検討対象として挙げる諸文献をみる限り、本稿と問題関心を共有していると思われるため、そのように解することに特段の支障はないと思われる。実際、同じ文献中でも両者が混在している箇所がある。また特段の理由がない限り「機械」も同様である。

そこでいう「自律」という語も、特段の理由がない限りは「人間が直接指示をせずとも、ある程度の自由裁量のもと行動する」という最広義の意味で用いることとし、必ずしも「自己決定権を有している」だとか「他者からの干渉を排除しうる」といった意味までも含意するものではないとする。

また、本稿の問題意識はあくまで、ロボット自体の処罰可能性の検討にあるため、関与した背後者たる人間の責任、とりわけメーカーの製造物責任については立ち入らないこととする。

2. ロボットの処罰可能性を巡る見解の対立

ロボットは法的責任主体たりうるか、という問いに対し、見解は大きく分けて3極に分かれている。すなわちこれを肯定する積極説、それに対してロボットはおよそ責任を負えないとする消極説、刑事上の責任は問えないが民事上の責任は観念可能であるとしたり、新たな種類の法的主体の導入により責任を観念したりする、バリエーション様々な中間説が存在する。以下、代表的な論者の見解を詳しくみていく。

(1) 積極説—— Hilgendorf の見解

積極説を代表する論者は Eric Hilgendorf である。Hilgendorf は意味論的規約主義 (semantischer Konventionalismus) を前提として、言語の転用可能性という観点からロボットに責任を肯定しようとする。

ある概念が意味を獲得する方法は二つある。すなわち用語法と定義である。その際、用語法は非常に大きな意義を有する。重要なのは、一般的な用語法の文脈において、記号の意味はその記号の使用者によって一方的に画定されうるものではなく、記号の使用者の解釈を主眼に決定されるわけでもない。潜在的な名宛人の理解にこそ重点があるのである。つまり、“記号の意味とは、その言語共同体において正し

いものと受容されている理解なのである。話者がその表現をもって話す内容は、あらゆる潜在的な聴者の理解によって話者がいかに理解されうるか、にかかっているのである⁷⁾。

すなわちある概念やある語の意味内容は、その言語を用いる言語共同体の共通認識によって画定されるのであり、話者がその語に込める意味ではなく、聴者が受け取る意味にこそ重点があるというのである。

しかしながら、ある語がその言語共同体において常に同じ意味で用いられているわけではなく、場合によっては矛盾することさえあるという。それゆえ学問においては、概念や語の意味内容を明確化する必要がある。その際、概念と語の意味を明確化するための方法こそが、注釈（*Explikation*）と定義（*Definition*）であるという。注釈が既存の概念を構成要素へと分解し、各要素を検討することによって意味を画定する試み⁸⁾である一方、定義は新たな概念に意味を付与する試みであるという。

定義とは言語的意味の画定である。定義とはしたがって、意味の記述ではなく、新たな言語的意味を規約（*Konvention*）に基づき規定することである。それゆえ定義は決して正しい、正しくないといった二者択一ではなく、せいぜいより適切（合目的）であるか否かの相対的な差があるに過ぎない⁹⁾。

7) *Eric Hilgendorf*, Können Roboter schuldhaft handeln? Zur Übertragbarkeit unseres normativen Grundvokabulars auf Maschinen, in: Susanne Beck (Hrsg.), *Jenseits von Mensch und Maschine*. S. 122. なお本論文は既に千葉大学共同研究チーム「ロボットと刑法」によって紹介されている。伊藤嘉亮「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—』」千葉大学法学論集31巻2号31頁以下。

8) *Hilgendorf*, a.a.O. (Fn. 7), S. 123. なお、同論文には「注釈」についてこれ以上の説明がない点を見るに、着目すべきは後述の「定義」の方であるように思われる。

9) *Hilgendorf*, a.a.O. (Fn. 7), S. 123 f.

定義による概念の意味画定は、学問といった一定の目的のためになされるものであるから、付与された意味内容が絶対的に正しいだとか誤っているだとかを論じることはできず、その目的によりよく即しているのかといった観点から評価しうるに過ぎないのだという。それゆえ、言語の意味内容は常に一定の意味に固定されているのではなく、歴史的過程の中で変化するとされる。

上述の検討から Hilgendorf は、ロボットに「答責性 (Verantwortung)」や「責任 (Schuld)」という語を転用することが、原理的に排除されるわけではないとする。意味論的規約主義によれば、その語を用いる言語共同体の共通理解こそが語の意味内容を決定するため、言語共同体がこれらの語をロボットにも転用可能な意味内容であるとの合意をしているのであれば、それで十分であるからである。現にドイツにおいては、企業は民法上の答責性を負うるのであって、人間以外の存在への転用も実際になされてきたというのである。

しかしながら、刑法上の責任は「人的非難可能性」であるとされ、その意味自体に争いがあるものの、そこでは意思自由の存在が大きな意味を有している。ロボットが自由意思を有しうるかが、刑法上の責任のロボットへの転用の可否を決する。

これについて Hilgendorf は、「意思自由という観念は単なる一つの要請であって、それ自体証明できるものではないし、科学的に分析できるものでもない。21世紀の初頭、意思自由というドグマは脳研究によって根幹から揺るがされ、法律学は、意思自由は（不可欠な）フィクションであるという Kohlrausch の見解へと立ち戻った¹⁰⁾とする。そして「意思自由が単なるフィクションに過ぎないのであれば、そのフィクションがロボットにも拡張されてはならないという理由はないのではないか。フィクションとは合目的性から生じる設定であり、それは『現実の』状況を反映するものではなく、むしろ時として現実と相容れない場合もある¹¹⁾という。それ

10) Hilgendorf, a.a.O. (Fn. 7), S. 129.

11) Hilgendorf, a.a.O. (Fn. 7), S. 129.

ゆえ「ロボットの『意思自由』をフィクションとして導入することも、相応の実務の必要性が生じた場合には考えられよう。しかし現状その必要はないように思われる」¹²⁾とする。というのも「ロボットの態度を修正するためには、プログラム変更をするのが最も簡単な手段であり、「様々な含意を伴う『刑罰』という概念を用いる必要はない」¹³⁾からであるとする。

以上より、ロボットも(フィクションではあるが)意思自由を有しうる余地はあり、したがってロボットの刑法上の責任について論じることも意味論的規約主義によれば原理的に不可能というわけではないという。

(2) 消 極 説

上述の積極説に対し、ロボットは原理的に責任を負えないとする消極説も存する。むしろ議論の現状を概観するに、積極説は少数説にとどまっている¹⁴⁾。消極説の論拠は様々で、以下では代表的な論者の見解をみていくこととする。

i) Joerden の見解

Jan C Joerden は前述の Hilgendorf と同様に、ロボットに意思自由を肯定することは可能であるとしつつも、人間が有する自由の構造は完全にはロボットに転用できないという。刑法の文脈において、自由という語を用いる場合、「従来から自由の二つの側面が区別されている。すなわち、意思の自由と評価の自由(Bewertungsfreiheit)である」¹⁵⁾。この両者はそれ

12) Hilgendorf, a.a.O. (Fn. 7), S. 130.

13) Hilgendorf, a.a.O. (Fn. 7), S. 131.

14) Andreas Matthias は積極説に立つが、刑罰論において特別予防論を採用しており、これは現在の通説的理解とは異なる。以下の Spiegel Online のインタビュー記事を参照。http://www.spiegel.de/wissenschaft/mensch/strafrecht-fuer-roboter-maschinen-sollen-fuer-eigene-fehler-haften-a-569004.html。我が国における積極説として松村格「システム論と自由意思—ヨナルト論文『刑法と自由意思』への釈明—」『法の理論Ⅱ』。

15) Jan. C. Joerden, Strafrechtliche Perspektiven der Robotik, in: Hilgendorf / Günther (Hrsg), Robotik und Gesetzgebung. S. 201. なお本論文は既に千葉大学共同

ぞれ行為帰属の段階と責任帰属の段階とに振り分けられるという。

ある行為の構成要件該当性を検討する際に問われるのは専ら、刑法的に重要な結果——たとえば殺人罪という人の死——が惹起者としての他人——たとえば殺人罪の正犯——と結びつくか否かである。換言すれば、殺人行為の帰属が問題となるのであって、この行為がどのように評価されるべきかという問いとは全く別問題である。しかしある状況が後に可罰的な行為であったと評価されるためには、その人の態度が自由な決断によるものであったといえなければならない。そうでなければその人は行為をしたのではなく、単にそのような事象が生起したとしかいえないからである。

このことが意味するのは、その人が別の行為をすることもできたであろう、という可能性が存在しなければならないということである——たとえば、その人がそうしようと思えば、他人を殺すことを差し控えることもできたであろう、ということである。無論これに対し、絶対的強制下にある者の態度は、行為とはいえない。なぜなら彼の態度は意思的なものではなく、絶対的な力によって強制されており、他の行為を選択することができないからである [……]。

行為帰属という事実の帰属の段階における自由——これは意思自由と呼ばれる——と、刑法的非難を可能たらしめるに不可欠な自由とは区別されねばならない [……]。

[……] 責任帰属のために前提となる自由は、単なる意思自由ではなく、それ以上の、法律に従うあるいは違反する決断を可能にする自由である。ここでは端的に評価の自由と呼ぶこととする。

たとえば未成年であることや病気であることを理由に、適法/不法を区別できない者は、評価の自由を有していない。この者にその殺

研究チーム「ロボットと刑法」によって紹介されている。今井康介「ヤン・C・イェルデン『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31頁以下。

人行為が帰属されるという意味では、この者はたしかに行為をしており、その行為は禁じられ、すなわち違法であるといえる。しかしその態度により彼に責任が帰属されることはない。なぜなら未成年であることや(精神の)病気であることを理由に法を知らない者は、法に従うあるいは違反する決断をなしえないからである[……。](傍点は原文イタリック)¹⁶⁾

すなわち人間の自由には二つの側面があり、一方が行為論に影響を与える意思自由、もう一方が責任論に影響を与える評価の自由であるという。意思自由はある行為が有意的になされたか否かを問題とするに過ぎない一方で、評価の自由は適法性/違法性に対する理解を前提として自らの行為を決定する自由であるとの理解を示している。

このような理解を前提とすれば、「遅かれ早かれロボットは、意思自由を有しているかのような印象を我々に与えることも可能となろう」¹⁷⁾。その際、実際に意思自由を有しているか否かは問題とはならないという。なぜなら、我々人間同士であっても真に意思自由を有しているのかは互いに知りえないのであるから、これはロボットであっても同様であるという。つまり、外部的な観察により、意思自由を有しているようにみえることこそが重要なのである。

同様のことが評価の自由にもあてはまる。つまり、ロボットが評価の自由を実際に有しているかではなく、評価の自由を有しているように見えるか否かが重要であるというのである。「しかし、いつの日かロボットが自由な意思で法に違反する決断をした、すなわち有責的に行為したのだと、チューリングテストによって我々が納得する日が来るとは思えない」¹⁸⁾という。ロボットが、チェスのような一定のルールの中で人間を凌駕することはありうるところだが、コミュニケーションに関しては歴史の話や冗

16) Joerden, a.a.O. (Fn. 15), S. 201 f.

17) Joerden, a.a.O. (Fn. 15), S. 203.

18) Joerden, a.a.O. (Fn. 15), S. 204.

談、価値判断も理解する必要があるため、この能力を得ることは原理的にきわめて難しいというのである。

それゆえロボットに刑法上の責任を認めることができるか、という問いに対しては、「かなり明快にノーという返答がなされねばならないように思われる」¹⁹⁾という。

ii) Gleß = Weigend の見解

Joerden と類似の帰結に至るのは Sabine Gleß = Thomas Weigend である。Gleß = Weigend はロボットの刑法上の問題として、「ロボットは行為しうるか」と「ロボットは有責たりうるか」という問いを挙げる。

前者の行為能力の問題について、「インテリジェント・エージェントは『任意に』行為しうるかという問いを生じるが、因果的行為論によれば態度が意思により制御されていることに対して高いハードルを課してはいない。任意の行為でないとされるのは、人の身体が（制御できない反射的反応のように）『勝手に』動いたり、あるいは他人に物理的な対象として利用される場合（たとえばAがBをCに向かって突き飛ばすなど）に限られる」²⁰⁾とする。他方、目的的行為論による場合には「人間の行為は、その行為者の目的志向的な意思の表出であるといえる場合にのみ、かつまさにその理由から、刑法上重要なのであり、「インテリジェント・エージェントは自らの『意思で』目標を設定し、その目標にしたがって行為を規定できる」²¹⁾わけではないから、行為能力が否定されるという。

結局のところ、インテリジェント・エージェントが（刑法上の）行

19) Joerden, a.a.O. (Fn. 15), S. 203.

20) Sabine Gleß = Thomas Weigend, *Intelligente Agenten und Strafrecht*, ZStW 2014; 126 (3), S. 571. なお本論文は既に千葉大学共同研究チーム「ロボットと刑法」によって紹介されている。伊藤嘉亮「ザビーネ・グレース=イエシュック・ヴァイゲント『インテリジェント・エージェントと刑法』」千葉大学法学論集31巻3・4号31頁以下。

21) Gleß = Weigend, a.a.O. (Fn. 20), S. 572.

為能力を有しうるかという問いは、定義の問題であらう。つまりあらゆる『任意の身体動作』を行為であると定義する、『因果的な』、単なる外部的観察によれば、インテリジェント・エージェントを行為者とみることもできる。行為概念により多くの内容を盛り込み、自己の意識に基づく目標設定という内容をより多く盛り込むほど、インテリジェント・エージェントが行為能力の要件を満たすことが難しくなるのである²²⁾。

また責任能力に関する問題に関して、1952年の連邦通常裁判所（BGH）の判例²³⁾を挙げ、「インテリジェント・エージェントは、現在のところ道徳的自己決定をなしうる人格ではなく、せいぜい与えられたタスクをプログラムに従って処理するコンピューター・テクノロジーとみられるに過ぎないから」、「[……] どんなに高知能な機械であっても、刑法上の責任を想定することは絶対に不可能である」²⁴⁾という。

しかしその際、意思自由はインテリジェント・エージェントの責任を否定する論拠にはならないという。

[……] 機能主義的な観点からいえば、犯罪において生じる社会的コンフリクトの原因として答責的であるとみなされる者が、責任ある者と記述される。したがって「自由意思」の観念もどうとでも解釈できる。つまり「行為を自ら決定する自由」というのは自然現象ではなく、一定の社会的目的に資すべき記述の、すなわち自由を相互承認す

22) *Gleß = Weigend*, a.a.O. (Fn. 20), S. 572 f.

23) BGHSt 2, 194. 本件では禁止の錯誤の犯罪体系論上の位置付けが問題となったのであるが、S. 200にて「責任非難の内在的根拠は、道徳的に成熟した人である以上、その者は自由答責のかつ道徳的に自己決定をなしえ、したがって法に従い不法を排し、自らの態度を法的当為たる規範に適合させ、法的禁止を回避する能力を有するという点にある」とした。

24) *Gleß = Weigend*, a.a.O. (Fn. 20), S. 574.

るという前提のもとでの社会制度の産物なのである²⁵⁾。

以上の点は、意思自由の位置付けが若干変わることを除けば、その帰結においてほぼ Joerden の見解と軌を一にする。しかし Gleß = Weigend はさらに、Joerden が指摘していない問題についても指摘している。すなわち「インテリジェント・エージェントは法における人格であるといえるか」と「インテリジェント・エージェントを『処罰』することは可能か」という問いである。

前者の人格性に関しては、「自己反省の能力こそが、人間に自己決定を可能たらしめているのである。あるいは Kant が明らかにしたように、人間は自らを意識するからこそ、自らを『理性によって彼が自らに与える法則にしか規定されないもの [……]』とみなすのであ」って、「[……] しかしインテリジェント・エージェントは自らの自由を意識できない」がゆえに「[……] 人格性 (Person-Seins) の本質的な要素が欠ける」²⁶⁾という。

後者の処罰可能性については、インテリジェント・エージェントの背後にいる人間を責任から解放する意味では法的にも意味がないではないが、インテリジェント・エージェントそれ自体の処罰にはあまり意味がないとする。

[……] たとえばインテリジェント・エージェントを物理的に破壊したり、完全に一から再プログラミングしてしまえば、それはたしかに一見して、一種の死刑のような、「人格に対する」処罰であるようにもみえる。しかし、——インテリジェント・エージェントが人間が有するような生き（永らえ）たいという意味を獲得しない限りどのみち——その処罰はインテリジェント・エージェントにとって、人間と同じほどのインパクトを与えるものではない。このような理由から、

25) Gleß = Weigend, a.a.O. (Fn. 20), S. 574 f.

26) Gleß = Weigend, a.a.O. (Fn. 20), S. 569 f.

インテリジェント・エージェントは刑罰をそれ自体として感受できないため、現在の開発状況においてはインテリジェント・エージェントの刑事処罰は無駄に終わる。人間はロボットの破壊をその過誤行為に対する非難的制裁ととらえることができるが、ロボットは非難の意味を理解できないのである²⁷⁾。

すなわち、ロボットないしインテリジェント・エージェントを「処罰」する場合には、その処罰が「非難」を意味するということが、処罰する側の人間だけではなく、処罰される側のロボットないしインテリジェント・エージェントにも共有されていなければならないというのである。現状、ロボットないしインテリジェント・エージェントは彼等自身の自由や生命といった観念を有しておらず、それゆえ処罰に意味を見出すことができないために、処罰をする意味に乏しいという。

iii) Ziemann の見解

Gleß = Weigend と同様に処罰の不可能性という論拠から消極説に立つのは、Sacha Ziemann である。Ziemann は三つの観点から機械の可罰性を否定する。すなわち、①刑法の人間中心主義、②機械を処罰する刑種の問題、③刑事政策的必要性への疑義、である。

①の論拠について Ziemann は、「刑法は人間である犯罪者にのみ向けられているのである。この原則は刑法の人間中心主義と呼ばれる。」²⁸⁾ という。たしかに「過去を振り返れば、啓蒙期以前は刑法秩序に数多くの非人間的存在が包摂されていた」し、「動物の処罰や動物の刑事手続きに関する記録も存する」²⁹⁾。しかし「非人間中心主義的刑法観は啓蒙期に終焉を

27) Gleß = Weigend, a.a.O. (Fn. 20), S. 578.

28) Sacha Ziemann, Wesen Wesen seid's gewesen?, in: Hilgendorf / Günther (Hrsg), Robotik und Gesetzgebung, S. 185. なお本論文は既に千葉大学共同研究チーム「ロボットと刑法」によって紹介されている。田村翔「サシャ・ツィーマン『機械の本性とは何であったか? 機械刑法をめぐる議論について』千葉大学法学論集31巻3・4号31頁以下。

29) Ziemann, a.a.O. (Fn. 28), S. 186.

迎える。刑法においても、中心点がもっぱら人間へとシフトしたのである。すなわち、一方で自由な行為をなしうる、理性を備えた個人としての、他方で法律に義務を負っている国民ないし臣民としての人間である。この基準に照らせば、もはや動物はここには含まれない。なぜなら動物を自由と理性を有するものと認めることはできず、また動物は国民でもないからである³⁰⁾。それゆえ、機械のような人間以外の存在に刑法の適用を認めることは、中世以前への退行であるという。

また、②について、機械を実際に処罰する際に、いかなる処罰が考えられるかという点を指摘する。その際に考えられる刑としては、「(機械の解体による)機械の『身体』への措置、(スイッチを切り、再起動することによる)機械の『生命』への措置、(電磁記録の抹消、再プログラミングによる)機械の『人格性』への措置³¹⁾であるという。

しかし「これら3つの措置は刑法典に取り入れられていないにとどまらず、啓蒙期以降の現代刑法にその措置を認める余地もなく、刑罰的措置と認めないとするのが妥当である³²⁾という。というのもこれらの措置はつまり『体罰』ないし『身体刑』、『死刑』、『洗脳』であって、ドイツでは「このような刑罰的措置は、犯罪者の身体や人格性、生命に対する侵害であって、憲法上禁止されている³³⁾からである。

③の刑事政策的必要性への疑義、というのは、いわゆる「答責の間隙」論に対する疑問である。機械が自律的に学習し、自ら判断をなし、行動するという性質上、それを開発した人間にも、利用する人間にも、その機械が具体的にどのような行動をするかについて厳密に予想することはできない。それゆえそのような機械は最早人間のコントロールが及ぶ範疇にはなく、誰も責任を負うことができない、というのが答責の間隙論である。

これについて Ziemann は「人間に答責できない場合に、ロボットに答

30) Ziemann, a.a.O. (Fn. 28), S. 187.

31) Ziemann, a.a.O. (Fn. 28), S. 189.

32) Ziemann, a.a.O. (Fn. 28), S. 189.

33) Ziemann, a.a.O. (Fn. 28), S. 189.

責することによって埋め合わせてしまおうという発想は説得的ではない』という。

まずもって、そのような発想は刑法理論的に合理的ではない。というのも、機械によって惹起された法益侵害に関してまでをも刑法的帰属のプロセスに載せることは、広きに失する。たとえば刑法には、危険源に対する監督義務や保安義務があり、それに違反すれば刑罰が科される。この危険源には、監督が必要な危険な物や危険な動物だけが含まれるわけではない。一定の条件下では、たとえば過度に熱心な警備員や出世欲が強すぎるセールスマンといった、自己答責的に行為する人間もその監督を義務付けられるような危険源となりうる。刑法的帰属のプロセスのこういった『発想の豊かさ』があれば、潜在的な答責性の間隙の領域は、非常に小さいと思われる³⁴⁾。

つまり、刑法を機械にまで適用するのではなく、それを利用する人間に注意義務や保障人義務を認めることができるのであるから、「答責の間隙」論が強調するほどの間隙はそもそも生じないという。Ziemannによれば、注意義務や保障人義務でなお捕捉できないような間隙は、最早「間隙」ではなく社会全体での「引き受けるべきもの」であるというのである³⁵⁾。

(3) 中間説

以上、積極説と消極説を概観してきたが、これらにくわえて第三の見解が存在する。ロボットは責任を負いうる／負えないといった、二者択一的な結論に至るのではなく、ロボットの特徴に鑑みて負いうる責任とそうでないものを区別するアプローチである。以下ではこれら中間的な見解を概観していく。

34) Ziemann, a.a.O. (Fn. 28), S. 190.

35) Ziemann, a.a.O. (Fn. 28), S. 191.

i) Asaro の見解

Peter Asaro は、法秩序は原則的に人間を前提として形成されているが、例外的に非人間的な存在も部分的に法における人格性を付与されるのであるから、ロボットはこれに含まれると主張する（ロボット準人格 quasi-person 説）。

未成年者は準人格の最たる例である。未成年者は、成人が享受する権利のすべてを享受するわけではない。とくに未成年者は契約を結んだり、様々な法的合意の当事者とはなりえない。なぜなら未成年者はそのような権利を有していないからである。未成年者がこのような法的合意の当事者となるのは両親や法的後見人の行為を通じてのみである。この意味で、未成年者は完全な法的身分を有していない。無論、子供を殺害することは、成人を殺害するのと同じく謀殺であり、この意味では子供でも法的な人格なのである。もっといえば、実際、成人よりも手厚い保護が与えられている。したがって子供は一種の準人格あるいは法的準行為能力者としてみなされている。慢性的な精神疾患を患っている人はかなり子供に似たような扱いを受ける。完全に一人前の法的な人格でさえ一時的な判断力の減退を主張することがあり、それによりたとえば一時的な精神病や、不意に意識を失ったりといった一定の状況下では、その行為に対する責任が減じられるのである。重要なのは、法的主体性のいくつかの側面は、完全に一人前の人格や完全責任が欠けた存在にも適用するということであり、法の下でいくつかのロボットはこの種の準主体として認められると考えることも合理的であるように思われる³⁶⁾。

ロボットが法における（準）人格であると認めたとして、それは法的評価においていかなる帰結をもたらすのであろうか。とりわけ、本稿が関心

36) Peter Asaro, A Body to Kick, but Still No Soul to Damn: Legal Perspectives on Robotics, Robot Ethics: The Ethical and Social Implications of Robotics, p. 180.

を寄せる責任という観点においては、どのような結論が導かれるのか。

この点につき Asaro は民事法と刑事法の差異に着目し、とりわけ刑罰の定義と関連してロボットの責任について以下のように論じている。

刑法は違反行為者を処罰することに関わる法域であるが、民法は主に違反者に強制的に損害賠償をさせることに関する法域である。この差異の基礎には重要な原理がある。すなわち、犯罪というのは、罰せられるべきなのである。その際、その犯罪による直接の損害に対する賠償の有無は関係がない。別の言い方をすれば、犯罪によって害された者というのは、社会全体なのである。だからこそ、刑事事件は国家や“人民”によって訴追されるのであり、違反者が負う責任というのは社会に対して負うものなのである。刑罰には様々な現象形態が存在する一方で、伝統的に刑罰にとって重要であると考えられてきたのはいくつかの意味で誤りを正すという点である。すなわち、違反者は社会に対する責任を果たすべきである（応報）、違反者は今後違反を繰り返さないように、改善されるべきである（改善）、社会における別の者が、似たような違反を犯さないように抑止されるべきである（抑止）、といった点である。

ロボットに刑法を適用するには二つの根本的な問題がある。(1)犯罪行為というのは、道徳的主体がこれをなす必要があり、(2)ロボットを処罰することなど可能か否かが定かではない。道徳的主体性は民法にとって構成的な要素ではない一方で、刑法にとっては構成的な要素なのである。そして道徳的主体性は、我々の刑罰に対する観念（応報、改善、抑止）に深く関わるのである。道徳的主体性は様々な形で定義されるが、刑法の文脈でいえば、究極的には、非難すべき性質を有し、処罰されることができるという自律的主体としての立場に値するものでなければならないのである。道徳的主体性がなければ、損害とはいえても（つまり民事責任はあっても）、有罪であるとはいえない。したがって、道徳的主体が存在しないのであれば、社会に対して負う

べき責任もまた存在しない。単に事故や自然のいたずらではあっても、犯罪ではないのである。同様に、道徳的主体のみが、道徳的性格の発展・矯正を含め、改善をされうるのである。そうでなければ、それは単に問題の修正に過ぎない。そして終局的には、ある道徳的主体が、誤った選択や行為を理由として処罰された他の道徳的主体の選択や行為と、自らの潜在的選択や潜在的行為との類似性を認識することができてはじめて、抑止は意味をなすのである。道徳的主体によるこの選択の対称性がなければ、そして道徳的主体の類似性に対する認識がなければ、刑罰が抑止効果を生ずることはない³⁷⁾。

それゆえ、ロボットは民事責任を負うことはできても、道徳的主体性が欠けるために刑事責任を負うことは難しいという。無論、Asaro はロボットの刑事責任を一律に否定しているのではなく、あくまで非常にハードルが高いといっているにすぎない。とりわけ、一部の事例においては、法人のような非人間的存在にも刑罰は観念されるという。なぜなら「企業は自らや株主のために利益を出すために存在しており、したがって金銭的な制裁を加えうる。そして独占禁止法違反などの一定の場合には、企業の解散もありうる。刑事事件によって収監されることはないものの、企業の責任者はその個人の行為について訴追されうる」³⁸⁾からである。

しかしながらその際に、法人とロボットの差異に注意を払わねばならないという。第一に、法人すなわち企業は利益追求を目的としているが、ロボットは必ずしもそのような目的を有しているとはいえないという点である。そうであれば金銭的制裁は意味をなさない。またロボットに何らかの目的が観念されたとしても、それは各々のロボットごとに異なりうるものであって、これを裁判所が公平に裁くことは難しいという³⁹⁾。

第二に、ロボットは法人とは異なり、具体的な身体を有しているという

37) Asaro, op. cit., p. 180-181.

38) Asaro, op. cit., p. 182.

39) Asaro, op. cit., p. 182.

点である。それゆえ、法人とは異なり取監という措置も可能ではある。しかしながら、それによって刑罰目的が達成されるかには疑問があるという。というのも、企業処罰であっても、人間に特有の願望や恐怖などを前提としているからである。すなわち、ロボットには苦痛や寿命がないことから、拷問や死刑を執り行ったとしても、応報、改善、抑止には役立たない。「人間を害する動物の場合と同じく、害をなしたロボットはすべて破壊するという政策もとりえようが、それは結局のところ、本来の意味での刑罰というよりはむしろ個人による将来的な害を避けるための予防的手段に過ぎない」⁴⁰⁾という。

ii) Beck の見解

中間説に立ちながらも、Asaroとは異なるアプローチをとるのはSusanne Beckである。Asaroがロボットを未成年者や法人といった、法における準人格とのアナロジーからロボットにも準人格性を付与しようとしたのに対し、Beckはそのようなアナロジーには限界があり、むしろ全く新しい法的存在の構想を導入するべきであるという⁴¹⁾。

また、上述のZiemannが「答責の間隙」は憂慮するほどのものではないというのに対して、Beckは立証の困難性から「答責の間隙」は避けられないという。すなわち、ロボットが人を傷付けてしまった場合、製造者や販売者、利用者の過失が問題とされるが、ロボットが自律学習により自らのプログラムを絶えず改変しているような場合、どの時点で誰に過失があったのかを立証することはほぼ不可能だというのである⁴²⁾。

40) Asaro, op. cit., p. 182.

41) Susanne Beck, Sinn und Unsinn von Statusfragen, in: Hilgendorf, Eric (Hrsg.): Robotik und Gesetzgebung, S. 256

42) Susanne Beck, Google-Cars, Software-Agents, Autonome Waffensysteme – neue Herausforderungen für das Strafrecht? in: Susanne Beck, Bernd-Dieter Meier, Carsten Momsen (Hrsg.), Cybercrime und Cyberinvestigations, S. 15. なお本論文は既に千葉大学共同研究チーム「ロボットと刑法」にて紹介した。拙稿「スザンネ・ベック『グーグル・カー、ソフトウェアエージェント、自律的武器システム—刑法にとっての新たな挑戦?』」千葉大学法学論集31巻3・4号頁以下。

無論、そのような場合には一律して利用者などに負責してしまう、というかたちでの答責分配もありうると指摘している。しかしながら、それは二つの理由から否定されるべきであるという。

【一律に利用者には負責してしまうと】明らかににロボットの行為が構造的欠陥に基づく場合を除き、利用者はロボットに委託したあらゆる行為に対して責任を負うこととなろう。しかしこれはやはり、委託の趣旨を没却し、したがって社会的にも受け入れがたいであろう。たとえば、自動運転自動車というのは、人間の運転手が非常に疲れているけれどもあまり休憩を取りたくないような場合に用いられる。もし運転を委託するという措置が、たとえばロボットの方が疲弊した人間よりもミスをしなれないといったような理由から、社会的に受容されている場合、もっぱら運転手が自動運転を使ったという理由によって運転手が自動運転自動車のすべての判断に刑法的な責任を負わねばならないとするのは矛盾している。というのも、運転手は自動運転モードのときでも自らが運転しているのと同じくらいの集中力を保たねばならないということになろうからである。長時間自動モードにする場合、これは酷であろう。もしそうだとすれば結局、自動運転の利用は意味がないことになってしまう⁴³⁾。

有り体にいえば、自動運転自動車などの自律的ロボットは、人間の負担を減らすために人間の判断を一部委託されているのであるから、その委託された行為について全て利用者たる人間が責任を負うとすれば、結局人間の負担は減らないこととなり、本末転倒だというのである。

もう一つの論拠として、利用者あるいは製造者に一律して負責してしまうと、そのようなロボットを購入しようという動機や、開発に注力しようという動機を萎えさせてしまうことになるというのである。とりわけ刑法

43) Beck, a.a.O. (Fn. 4), S. 189.

は「有責な、つまり明らかに非難すべき行為を処罰する」のであって、その非難が「一定のコミュニケーション手段としての一定の作用」を有する限りで用いられねばならないという⁴⁴⁾。そのような刑法の回顧的性質に鑑みれば、刑法は新たなテクノロジーが有する危険性をコントロールすることには不向きなのであるから、刑法が積極的に用いられることによって、テクノロジーの発展が阻害されてはならないというのである⁴⁵⁾。

それゆえ Beck はこの「答責の間隙」を埋める一つの方策として、法における新たな人格概念、すなわち「電子的人格」を導入すべきであるという⁴⁶⁾。Beck の見解が Asaro の見解と異なるのは、Asaro がロボットを既存の法秩序に組み込むためにアナロジーを論拠に準人格として扱った一方、Beck は「ロボットがいくつかの点で人間にどれだけ『似ている』かを問ういずれの試みも、結局は推測にとどまらざるをえない」し、「ロボットという存在は、一切の生物学的な特性を有さず、人間にプログラムさ

44) Beck, a.a.O. (Fn. 42), S. 17.

45) Beck, a.a.O. (Fn. 42), S. 33 f.

46) Beck, a.a.O. (Fn. 4), S. 189 ff. なお Beck は「答責の間隙」を埋める方策として、「電子的人格」の導入が唯一の解決策であるとは述べていない。現行法の枠内で解決しようとするのであれば、過失概念を修正し、予見可能性を不要と解することによって、製造者やプログラマーに答責する方法もあるとしている。この方法によればたしかに、「人が害されたにもかかわらず、責任を負う者がいない」という状況は避けられる。無論、Beck はこの方法によることに疑問がないではないとしている。注意義務を広く認めることで過失を広く認めたととしても、結局誰の過失で当該事故が起きたのかを厳密に立証することは難しいという点、技術発展を阻害してしまいうる点は既に述べた通りである。また、注意義務の認定に関しても、ロボットのような先進テクノロジーの場合、注意義務の基礎となる明確な行為準則やガイドラインが未確立である点も障害となるという。製造の際の様々な製品規格を基準に注意義務を画定し、製造責任を問うということもできようが、そもそもそのような製品規格は民間の協会等が策定するものであって、民主主義的な基礎を有していない上に、刑法が保護しようとする利益と民間の製品規格が追及する利益とは必ずしも一致しないのであるから、刑法上の注意義務を民間規格に依拠させてしまうのは適切ではないと指摘している。

れたものであるという点で、全く新たな種である」として、既存の法的存在に関連付けて法秩序へと組み込むことを拒絶している⁴⁷⁾。

Beckによれば、アナロジーを論拠として法秩序に組み込む方法によると、ロボットと他の法的人格との類似性を検討し、その類似性を存在論的な基礎付けとして、法的人格が有する権利や義務の一部をロボットにも認めるという論理をたどることになるという。たとえば、ロボットには学習により情報を収集し、それに基づく判断形成をなすうという点に鑑み、これは部分的に人間と一致するのであるから、意思の自由やそれに基づく責任というものをロボットにも観念できる、という論理である。

しかしそうした場合、様々な問題が生じる、とBeckは指摘する。まず、そのような存在論的な根拠を人間自身が有しているのかが厳密には証明できないという点である。それはたとえば意思の自由が科学的には実証できないことと同様である。また現実的な問題として、そのような「人間に準じた存在」を創造し、場合によっては廃棄、破壊してしまうことは、許されないのではないか、という疑問がある。そして刑法に関わる点でいえば、類似性や存在論的な基礎は複数あろうが、刑罰を科すために必須の要素はどういったものなのかが一義的には答えられないし、その答えを得るには刑罰論にまで遡らねばならないという⁴⁸⁾。

したがって、アナロジーによって既存の法的存在の地位を一部転用することで法秩序に組み込む方法は取らず、完全に新たなカテゴリーとして「電子的人格」を導入するのが、少なくとも現状に鑑みれば無難であるという。具体的には、商業登記簿への登録のように、そのロボットの性能や利用用途などを登録し、その登録番号に基づきそのデータを照会した人間が当該ロボットの危険性を判断できるようにしたり、部分的に財産権を認め、問題を起こしたロボットに関わった人間の過失が立証できないような場合には、そのロボットの有する財産から賠償させるようにするような制

47) Beck, a.a.O. (Fn. 41), S. 253.

48) Beck, a.a.O. (Fn. 4), S. 190 f.

度設計が望ましいという⁴⁹⁾。

3. 若干の検討

以上、ロボットの処罰可能性を巡る議論の現状を概観してきた。以下では各説に対し若干の検討を加えたい。

(1) 積極説について

積極説に立つ Hilgendorf は、意味論的規約主義の観点から、ロボットに責任を観念する余地を残した。より厳密には、ロボットにも責任を認める可能性については否定していない。ただし、現状ではロボットに責任を観念する必要はなく、またロボットのバグを取り除く措置のことを刑罰と呼ぶ必要はないとする。

たしかに刑罰という語は様々な意味を含意しており、それをそのままロボットに転用しても刑罰という語が含んでいる意味の全てを適切に反映できないかもしれない。たとえば、刑罰は責任非難であるとすれば、刑罰にはコミュニケーション的契機が含まれる。その際、相互理解なくしてコミュニケーションは成立しない。非難が非難としての意味を伝達できないのであれば、それはコミュニケーションとはいえないということである。ロボットが非難を理解しないのであれば、刑罰によるコミュニケーションは成立しないために、ロボットに対する措置に刑罰という語を転用することはできないであろう。

しかしながら、Hilgendorf の立つ意味論的規約主義を前提とすれば、重要なのはその言語共同体における共通理解なのであって、必ずしも刑罰という語の転用が排除されるものではない。これを前提にしてたとえば、刑罰の目的や本質が法共同体の構成員の応報感情の慰撫にあると仮定した場合、刑罰にとって法共同体の構成員が満足することこそ重要なのであるか

49) Beck, a.a.O. (Fn. 41), S. 256 f.

ら、違反行為をしたロボットの破壊（死刑）によって構成員の満足が得られるのであれば、それは刑罰と呼びうることとなろう。実際にそのような傾向は我が国やドイツでもみられるのであるから⁵⁰⁾、責任や刑罰という語を転用する必要がないと言い切れるかはなお検討の余地がある。

(2) 消極説について

Joerden は、上述の Hilgendorf と同様に、ロボットに意思自由を仮定する。Hilgendorf と異なるのは、意思自由を行為論の問題であるとする点である。有意的行為論的な観点から、外部的観察として意思の自由があるように見えることが重要であるとする。

しかし、Joerden によれば、ロボットには「評価の自由」がないことを理由に、責任が欠けるといふ。「評価の自由」に関しても、その自由を有しているように見えることが重要であるという点は、意思の自由と同様であるが、ロボットが評価の自由を獲得することは難しいという。

問題は、評価の自由が意味する内容である。Joerden は評価の自由が欠ける場合の例として、未成年者や精神病患者を挙げている。しかし、これらの者には評価の自由が欠けるといわれるのではなく、通常は責任能力が欠けるとされる。評価の自由とは、是非弁別能力（あるいはさらにその基礎となる能力）のようなものであろうか。

この問いに答えたのが Gleß = Weigend である。Gleß = Weigend は基本的には Joerden と同様の見解に立っており、ロボットの行為性に関しては抛って立つ行為論次第であるとしており、ロボットの責任能力については BGH の1952年判例を引用しつつ、これを否定している。すなわち、BGH が挙げた道徳的自己決定という基準を満たさないというのである。

Gleß = Weigend はさらに人格性と処罰可能性について触れており、前者は Kant を引用して道徳的自己反省能力の欠如を理由に、後者は生への

50) 刑罰が市民の不安を払拭するために用いられることの問題性について、詳しくは拙稿「刑罰における応報と予防の関係性について」大学院研究年報第44号 法学研究科篇379頁以下。

欲求といった能力の欠如を理由に、いずれも否定している。

しかし、法における人格にとって、道徳的自己反省能力は構成的な要素であろうか。というのも、ドイツにおいて企業のような法人も法的人格とみなされており、法における行為主体性や、保護客体性を有しているものの、(企業としての)法人自身が道徳的自己反省能力を有しているかは甚だ疑問である。道徳的自己反省能力がない場合でも、必ずしも人格性が否定されるわけではないのではないか⁵¹⁾。

後者の処罰可能性についてより詳細に説明し、これを可罰性否定の論拠にするのは Ziemann である。Ziemann は具体的な刑種にまで踏み込んで、ロボットに対する諸措置を刑罰と呼ぶことはできないとした。そして、ロボットへの刑罰を観念する以前に、そもそもロボットへの答責移譲によっても「答責の間隙」は危惧するほど生じないという。解釈的な取り組みによって、ロボットに対する注意義務や保障人義務を認定することは可能であり、したがって「答責の間隙」は憂慮すべき問題ではないという。

たしかに解釈学的な努力によって、注意義務や保障人義務を広く認めることは可能であろう。しかしながら、むしろ問題は本稿の冒頭でも述べたように、立証の困難性にあるのである。製造・利用に大勢の人間が関与するロボットの性質上、注意義務が肯定されたとしても、結果に現実化した注意義務違反がどれであるのかを特定し、立証することが困難なのである。

(3) 中間説について

Asaro は上述の消極説とは異なり、ロボットに部分的に人格性を付与しようとする。従来より非人間的的存在にも法における人格性は付与されてきたのであるから、それ自体は何ら驚くべきことではない。また積極説・消極説の論者もロボットを論じる際に法人を引き合いに出している点からみ

51) 無論、ドイツにおいても法人処罰はなされていないため、刑法上の意味で人格性がないということであれば、Gieß = Weigend の指摘するとおりである。

でも、積極説や消極説の理解と全く前提が異なるというものでもない。

消極説が主張するのと同様に道徳的主体性の欠如を理由として、Asaro もロボットに刑法上の人格性を付与するまでには至らない。Hilgendorf と同様に、刑罰には様々な目的が観念されるがゆえに、道徳的主体でなければ刑罰の目的を達成できないというのである。したがって Asaro は、人格性を肯定する意味で中間説に分類できようが、処罰可能性を否定する意味では消極説に近い。

ところで、道徳的主体を対象とするものでなければ、刑罰の目的は達成されないとしたが、はたして本当にそういえるのであろうか。というのも、Asaro は刑罰目的に改善を含めているが、ロボットの改善はむしろ人間よりも容易であるように考えられるからである⁵²⁾。また、改善を刑罰目的とするいわゆる特別予防論からすれば、刑罰は行為者の危険な性格を矯正する処分であって、そこでは行為者は半ば病人のような扱いを受け、単なる処分客体にすぎないとされる。そこでは道徳的主体性は何ら問題とはなっておらず、むしろ道徳的主体性を強調すると、そのような主体を強制的な処分の客体にすぎないとする事の妨げとなろう。

また、一部の道徳的主体性が欠けることを理由に処罰が不可能であるというが、具体的には何が欠けるがゆえに刑罰を観念できないというのであろうか。刑罰にとって不可欠の道徳性とはいかなるものかについて説明がないまま、処罰の不可能性を導くのであれば、これは結論先取りではなかろうか。

こういった基礎付けの問題を回避するのが Beck である。Beck は Ziemann とは異なり、「答責の間隙」は決して小さいものではなく、それを埋める一方策として「電子的人格概念」の導入を提案する。その際、Asaro が陥った基礎付けの問題を回避するために、既存の法的人格とのアナロジーでは考えない。民事上の概念として便宜的に主体とみなすのであ

52) まさにこの観点から Matthias, a.a.O. (Fn. 14). はロボットを処罰可能であるとする。

る。

具体的な制度設計として、ロボットに一部の財産権を肯定し、賠償主体とするという。たとえば、ロボットにはその労働に応じて、給料が支払われねばならず、過誤があればその給料から賠償金を支払う、という運用がなされるのだという⁵³⁾。

しかし、ロボットは給料を欲することはない。指摘するまでもなく当然ながら、人間が金銭を欲するのは、前提として衣食住にコストがかかり、生命を維持するにはそのコストをまかなうことが不可欠だからである。ロボットは衣食に窮することなく、またそのロボットの所有者のもとで労働をしている限りで住居（と呼べるかは別にして）も自らコストを支払って所有する必要はない。これを前提とした上で、なお一定の金銭をロボット自身に所有させることからすれば、おそらく Beck がイメージしているのは、給料として財産権を付与するというよりは、我が国でいう自賠責法上の「運転者」として賠償主体とするといった構成であろう。

(4) 共通の認識？

興味深いのは、各説の論者は意思自由を、ロボットの責任を否定する理由にはしないという点である。というのも、一見すればロボットはプログラミング・データの集合体であり、データの処理に関しても一定のアルゴリズムに従って行動しているように思われるからである。それにもかかわらず、ロボットにも意思自由を肯定しようというのは、意思自由が法的フィクションであるという Roxin の見解が広く受け入れられているからであろう。むしろ Joerden は同論文で、人間的な必然性（食事を採らねば生きていけないだとか、寿命があることだとか）が無い分、むしろ人間よりも自由である可能性すらあるという。したがって、意思自由は人間とロボットを区別する基準にはならない、ということが共通の認識となっているようである。

53) Beck, a.a.O. (Fn. 41), S. 256.

また他方、意思自由は肯定できたとしても、その他の能力や性質が無いことを理由に処罰を否定するのも、各説に共通する点である。Hilgendorfは刑罰の多義性を理由とし、JoerdenやGleß = Weigend, Asaroは道徳的な能力ないし刑法上の人格性の欠如を理由として、処罰を否定する。しかしながら、これらは責任論上のどの要件に還元されるのかについて言及するものは少ない。この点については今後の議論を俟つほかない。

4. 私見の方向性

以上の検討と共通認識を踏まえて、若干私見の方向性を示すこととする。

ロボットに責任を肯定できるかという点に入る前に、まず前提として、ロボットの利用に関する答責分配を人間側の負担を増やすことによって解決することには賛成できないということを先に述べておく。つまり、たとえば自動運転自動車が他人を傷付けてしまった場合に、(たとえば過失概念や作為義務を拡張することなどによって)常に製造者あるいは利用者のうちのただ一人に全責任を負わせるかたちで責任分配することには反対である。これには二つの理由が挙げられる。

第一に、Beckも指摘するように、人間側の負担増による解決を採用すれば、技術発展を阻害することにつながる。すなわち、たとえば製造時点や利用時点で予見することが不可能であるような事故により他人を傷付けた場合に、予見可能性を不要と解するよう過失概念を修正⁵⁴⁾することで製造者あるいは利用者⁵⁴⁾に全責任を負わせるとするなどの方法によって解決しようとするれば、製造者や利用者は未知の危険にすら責任を負うこととなる。製造者は、このようないつ処罰がなされるかわからない、不安定な地位に置かれてまでこの技術を開発しようとは思わないであろうし、それは

54) 実際に、予見可能性を注意義務内容の確定のための一基準に過ぎないものとする理解は我が国でも見られる。樋口亮介「注意義務の内容画定基準—比例原則に基づく義務内容の画定」『山口厚先生献呈論文集』(2014) 227頁以下参照。

利用者においても似たようなことがあてはまるであろう。刑法の威嚇力が技術発展の妨げとなってしまうのである。

第二に、やはり Beck が指摘しているが、人間側の負担増による解決は、ロボットの利用の趣旨を没却してしまうように思われるのである。ロボットの利用は、人間の日常生活において、人間の負担を軽減する点にその趣旨がある。たとえば自動運転技術は、自動車の運転をロボットや AI が代行することによって、人間をその労務から解放しようとする点に目的があるのである。過失を広範に肯定することにより、自分で運転する以上に神経を尖らせて自動車の挙動を注視しなければならないのであれば、自動運転技術の目的はまったく達成されないことになる。また、自動運転技術は免許を有していない者や、高齢者・障害者の移動を容易にしようとするものであるが、処罰のリスクから利用に躊躇せざるをえないということにもなりかねない。モビリティを抑制してしまう可能性に鑑みても、人間側に答責するという方法は妥当ではない。

無論、「利益を得ようとする者は、不利益をも甘受せねばならない」という民事法的な発想もありえよう。ロボットを利用して利益を得たならば、ロボットを利用することで生じる不利益にも責任を負わねばならないというこの発想自体に異論はない。しかしこれはあくまで使用者責任といった民事上の賠償主体を決する論拠であって、刑法上の答責主体を決する理論として用いることはできないように思われる。刑法上の答責主体となるのはあくまでその危険源につき管理・監督の責任がある者であって、利益を得ているか否かはそこでは構成的な要素ではない。

したがって「答責の間隙」を人間側の負担によって埋めるのは妥当ではない。しかし、他方でこの新たな技術によって損害を被った者は一切保護されないとするのも不当である。また、害意を持った人間がロボットを用いて犯罪を犯した場合⁵⁵⁾に、その者を一切答責できないというのもまた不

55) たとえば害意を持った者が自動運転自動車の走行プログラムを一部改変し、無差別に他人に衝突するよう仕向け、学習機能などの累積効果によりロボットがさらに別の方法で他人を傷付けたような場合、プログラムを改変した者は

都合である。問題は、そのバランスである。

Hilgendorfのようにロボットを処罰対象と認め、刑事上の責任を肯定する方法も一つの解決策である。無論、Hilgendorfも刑事責任を認める可能性を指摘したのみで、積極的にそのような運用がなされるべきであると述べているわけではないことには注意が必要である。

しかし私見によれば、ロボットに刑罰を科すことはできないように思われる。というのも、ロボットの処罰には難点があるからである。

当然ながら、ロボットには寿命が存在しないために、自由刑のような一定期間自由を制限するといった、時間による制裁はほぼ意味をなさないのである。人間が自由刑を不快なものと感じるのは、移動の自由が制限されることに加え、それが長期に渡れば年老いていくからでもあろう。寿命がないロボットには、時間の概念がなく、自由刑は苦痛としての意味をなさない。さらに、死刑も同様に意義を有しない。たとえばそのロボットがインターネット接続し、自己のOSや蓄積データの総体（人間でいう「脳」）をサーバー上に保存し、常に同期していた場合、そのボディ（人間でいう「身体」）を完全に破壊したとしても、サーバー上に残しておいたデータを同型のロボットにダウンロードしてしまえば、完全な復旧が可能である。それゆえロボットには死刑を科しても有意義とはいえないのである。

唯一、効果があるように思われるのは、再プログラミングや「バグ」の修正である。特別予防論的な観点から、この措置を刑罰として正当化することもできるとする見解もある⁵⁶⁾。しかしながら、私見によれば、これにも難点がある。

まずその措置自体を刑罰と呼ぶかという問題がある。つまり、再プ

そのような因果経過を予測していないのであるから、故意が否定される余地が生じる。しかしこれを不当として予測・予見を抽象化することの問題性は既に述べた。認識が非常に抽象的なもので足りるとすれば、およそ自動運転自動車に乗る者は、罪に問われることとなりうる。ここにジレンマが存在するのである。

56) *Matthias*, a.a.O. (Fn. 14).

プログラミングというのは、人格の書き換えなのであるから、それは許容されるか、という問題である。悪事をはたらいた人間を、薬物の作用や外科的な治療によって善人につくりかえてしまうことを刑罰と呼ぶ者はいないであろう⁵⁷⁾。

その点をおいておくにしても、再プログラミングという措置の技術的な側面に刑罰との不親和性がある。というのも、再プログラミングや「バグ」の修正というのは、過誤行為を犯したロボットだけに適用すればよいものではない。同型・同プログラムのロボットは全て同じ過誤を犯す可能性を潜在的に有しているのであるから、同型・同プログラムのロボット全てにこのアップデートがなされる必要がある。しかしながら、Hartによれば、刑罰は「犯罪を犯した者に対して」、「犯罪を犯したことを理由に」なされるものであるから⁵⁸⁾、未だなお過誤を犯していないロボットにまで、特段の理由なく適用されるこの措置を刑罰と呼ぶことはできないであろう。

それゆえ、本論の結論としては、現状に鑑みればロボットに刑事責任を問う、あるいは刑罰を科すことに意味はないということになる⁵⁹⁾。それゆえ、刑法上の責任が問題となるのは、従来どおりあくまで過失が立証された人間にのみ、ということとなる。無論、再三指摘されているように、立証の困難性について回ることとなろうが、ロボットの学習システムにシステムログを具備させることを製造者に義務付けることなどによってアクセシビリティを確保することは可能である⁶⁰⁾。

57) Anthony Burgess, "A CLOCKWORK ORANGE".

58) H. L. A. Hart, *Punishment and Responsibility*, 1968.

59) ただし、積極説の検討にて述べたように、刑罰の目的を「応報感情の慰撫」に求めれば、意味はあることとなろう。それが効果や影響力という点で有意義か否かは別として、ではあるが。

60) ロボットの判断がいかになされたかを事後的に検証するための説明責任、プログラムの透明性の重要性を指摘するものとして、平野晋『「ロボット法」と自動運転の『派生型トロッコ問題』—主要論点の整理と、AIネットワークシステム『研究開発8原則』』NBL No. 1083, 29頁以下。なおシステムログを具

また、前章の検討において抽出された共通認識によれば、意思の自由が責任を肯定するための必要十分条件であるかについては、積極説と消極説の間で相違はあるものの、ロボットに意思の自由を認め、行為性を認めうることに異論はないようである。そうであれば、ロボットに行為主体性を認めることに特段の支障は無いようである⁶¹⁾。

このことは刑法理論上、大きな意味を有するようと思われる。ロボット自身が行為主体であると認められる場合、ロボットが彼自身の過失により他人を傷付けた場合、その結果は行為主体としてのロボットに帰属されるのであって、製造者や利用者にまで遡及しない。この遡及禁止効果は、人間側の負担増による解決に反対する私見の方針に沿うものである。

さらに害意を持った人間が、ロボットのプログラミングを一部変更することによって、他人を傷付けるように仕向けたような場合も、3章4節にて述べたように、ロボットは意思自由を有しうることからロボットの行為に行為性が認められるという共通認識が存在するゆえに、その共通認識によればロボットの行為への行為支配を想定でき、間接正犯構成により捕捉することが可能となる。

このような主体性に加え、電子的人格概念のような、法における新たな人格性までも導入すべきかは、一旦保留しておきたい。導入するにして

備すれば問題がすべて解決するものではないという点を付言しておく。というのも、①膨大なシステムログを解読するのは結局人間であるが、そのようなことが実際に可能か、②システムログがAIのハードディスク上に蓄積されると、メモリの関係上処理速度が徐々に低下し、それは自動運転のような0.1秒の処理スピードが事故の明暗を分けることからすれば好ましくないのではないか、③ハードディスクではなく、クラウドやサーバー上に保存すれば処理速度の問題は生じないが、ウイルス感染・ハッキングのおそれが生じるのではないか、といった疑問が生じるからである。

- 61) ただし目的的行為論によれば、行為は目的志向的であるがゆえに、行為性を肯定することは難しい。ロボットは自ら目的を設定して、そのための手段を選択し、行為するわけではなく、常に人間の手段として、目的を与えられて行為しているからである。

も、現状の技術に鑑みれば、刑事法上の処罰を可能にするほどの人格性ではなく、中間説が指摘するように、一部の賠償責任の主体足りうるような、たとえば我が国の自賠責法上の賠償主体として構成し、保険料は利用者が支払うといった運用が現実的であろう。

5. おわりに

以上見てきたように、積極説においても消極説においても、ロボットに意思の自由は肯定しようという点に争いはない。しかしその上で両者の結論を分けているのは、意思の自由が責任を肯定するための必要十分条件になっているかという点である。消極説や中間説は、人間にはある（と思われる）が、ロボットにはないものを理由に責任を否定している。たとえば Joerden の言葉を借りれば「評価の自由」が、Asaro の言葉を借りれば「道徳的主体性」が欠けるというのである。これらの要素が責任のどの要件に関連するのは定かではないし、これは何を非難するのかという問い、すなわち刑罰論に直結する問いである。刑罰による非難にとって不可欠な要素は何であるのか、そしてそれは犯罪論上の何に還元される要素なのか、という点に関しては、刑罰論の文脈における今後の課題となろう。

また表裏ではあるが、もう一つ積極説と消極説の結論を分けている点を挙げるとすれば、人間とロボットの間に質的な連続性があるのか、あるいは将来連続性が生じうるのかという点についての見解の相違がある。消極説は、そのような将来はありえず、人間とロボットが質的に断絶していると考えている節がある。しかし、たとえばロボットに「寿命」を設定するなど、条件面で人間と差異がないような状況にまで至った場合、ハードウェアが蛋白質で構成されているか金属で構成されているかの差異しかなければ、この差異は質的な差異と呼べるのであろうか⁶²⁾。無論、これ

62) フィリップ・ディック「アンドロイドは電気羊の夢を見るか？」なお付言すると Hilgendorf, a.a.O. (Fn. 7) は同論文を以下のような書き出しで始めている。「はるか昔、カシオペア座の中にあつて、中くらいの恒星の中くらいの惑星に

はまだSF小説が扱うテーマである。ロボットと人間の質的な断絶が存在するのか否かは、今後の技術の発展を俟って見極める必要がある。

おいて、金属でできた生き物の文化が芽生えた。彼らは自らを『ロボット』と呼んだ。彼らは次第にその惑星を支配し、文学や哲学、科学や法をつくりだした。偉大な哲学者たちが『善』や『悪』を論じ、『責任』とは何か、それはどのような条件で生じるのかについて問うてきた。ロボットたちの文明は、やがて宇宙船を建造するまでに発展し、ついには地球に降り立った。ロボットたちは地球上で、有機物でできた下等生物と出会った。彼らは自らを『人間』と呼ぶらしい。人間もどうやら行為をするようで、その行為を『善』とか『悪』とかといった基準で評価しているようである。また、人間は『責任』についても多少知っている風だった。しかしロボットの哲学者たちは、人間が本当の意味で行為できるはずはなく、『善』や『悪』を理解することも、責任を負うこともできないはずだと考えた。なぜなら、意思の自由はロボットにのみあるのだから。『行為』や『答責』ないし『責任』というのはロボットにだけあてはまるのであって、人間にはそんなものはない、と。』