ロボット・ＡＩに対して「刑罰」を科すことは可能か

根津洸希

一　はじめに――四つのテーマの提示

悪事を働いたロボットやそのＡＩに「刑罰」を科す、というとＳＦ小説のテーマであるように聞こえるかもしれない。しかしながら、ロボットやＡＩの可罰性を巡る問題は、既にＳＦ小説のモチーフになるにとどまらず、法的議論においても顕在化しつつある[1]。それはたとえば、自動運転車両[2]が「自らの意思」、すなわち製造者・販売者・利用者

などのあらゆる人間側の故意・過失が無かったにもかかわらず、人を傷害した場合を想像すれば、この問題が遠い将来の問題ではないことに思い至るであろう。[3]

しかし他方、「自らの意思」で人を傷害した自動運転車両を法廷に召喚し、口頭弁論をし、判決を宣告し、宣告された刑期を刑務所の中で過ごさせることは、非常に滑稽なことのように思える。しかしそれがなぜ滑稽なことのように思えるのかを、きちんと理論的に言語化することは存外難しいことのように思われる。

というのも、ロボットの可罰性を巡る議論の現状をみてみると、各論者は必ずしも同じレベルで議論を戦わせているわけではないように思われるからである。一方で「ロボットに『責任』は観念しうるか」だとか、「『責任』の前提たる自由意思を持ちうるか」[4]といった、ロボットの責任を論じることによって可罰性を検討する犯罪論体系的なアプローチがあり、また一方で、「ロボットは規範の名宛人たりうるか」ないし「『人格』たりうるか」[5]を論じることによって可罰性を検討する前法律的アプローチ、あるいは「ロボットに『刑罰』を科することは可能かつ必要か」[6]という刑事政策的なアプローチが存在する。しかしながら、「ロボットには責任を観念しうる」という主張に対し、「しかしロボットは『人格』ではない」あるいは「ロボットに『刑罰』は必要ない」と反論しても、各主張が噛み合わず、あまり実りの多い議論にはならないであろう。

それゆえ本稿では、レベルの異なる議論を混淆し、しかも抽象的な問題設定により更なる混乱をもたらしてしまうことを避けるため、より具体的かつ限定的な問題を設定したい。すなわち、自動運転車両が「自らの意思」で人を傷害し、かつその自動運転車両に「責任」や「人格性」が仮に認められるものだとして、その自動運転車両を「刑罰」に処すことは可能か、という問題である。

したがって本稿では「ロボットは『責任』を有しうるか」といった問題には立ち入らない。そうではなく、「責任」も「人格性」も「必要性」も仮に全て満たされたとして、自動運転車両のようなロボットやそのＡＩに「刑罰」を科す際に支障となりうる現実的な諸問題を扱う。具体的には、①ロボットやＡＩに「責任」や「人格」を付与したとして、責任を負う「一人」の人格の範囲はどこまでなのか（「刑罰」を受けるのは誰かという問題：二章）、②「一人」の範囲を画定したとして、その刑罰に実効性はあるのか（「刑罰」はロボットやＡＩにとって苦痛となりうるかという問題：三章）、③実効性ある「刑罰」が考案できたとして、その「刑罰」は近代刑法の原則に沿うか（近代刑法における意味での「刑罰」と呼べるかという問題：四章）、④①～③の問題が全て解決したとして、ロボットやＡＩに「刑罰」を科すことは、人間にとっての「刑罰」の意味に影響をもたらさないか（ロボットやＡＩを「処罰」することによる「刑罰」という語の意味変容という問題：五章）、である。

なお、本稿ではロボットやＡＩという用語が多々用いられるが、本稿にいうロボットやＡＩないし自動運転車両という語で想定されているのは、今後数年～十数年で実装が予想される技術であって、一〇〇年先の技術までをも念頭に置くものではない。この種の議論をする際には、論者によって想定される「ロボット像」が異なることによって議論が噛み合わないことも多々あるように見受けられるため、付言する次第である。(7)

二　「刑罰」を受けるのは誰かという問題

「ロボットやＡＩは『人格』たりうるか」(8)という問いに対して、仮に「『人格』たりうる」との結論が導き出された

として、それによってロボットやＡＩに対する「刑罰」は実現に向けてひとつ歩を進めたことになるのであろうか。たとえば、自動運転車両やそのＡＩが法的人格としての地位を仮設された場合に、その自動運転車両を処罰することの障壁の一つが取り除かれたことになるのであろうか。

そこで考えねばならない問題として、その「人格」の一個性が挙げられよう。自動運転車両のＡＩに人格性が付与された場合に、その人格は車両一台ごとに宿っているのであろうか。それとも、同車種・同ＯＳのＡＩは同一人格、すなわち全体で一人の人格として扱われるのであろうか。(9)

仮に、車両一台ごとを一人の人格として取り扱うこととしよう。(10) その場合、「自らの意思」で人を傷害しようとした車両こそが単独正犯となろう。したがって同車両のみが「刑罰」を科される主体となるわけであるが、この「刑罰」は有意義なものであろうか。というのも、同車両は確かに「自らの意思」で人を傷害したわけであるが、同車種・同ＯＳのＡＩは基本的なスペックを共有しているのであるから、同様に「人を傷害する意思」を有する潜在的な可能性があるのではないか。そうであるとすれば、車両一台ごとに人格を付与することは、その車両に「刑罰」を科すことに一定の意味を持つこととなろうが、そもそもその「刑罰」自体に意味はあるのかという疑問は残る。少なくとも、同種事故の再発防止には資さないであろう。

なお以上の検討は、自動運転車両がスタンドアローン状態で運用された場合を前提としている。しかしながら、現在の自動運転技術はそれ単体で運用されるわけではなく、他の自動運転車両などとインターネットを経由して情報を交換しながら走行する。このような車両はコネクテッドカーと呼ばれるが、「コネクテッドカーとは、ＩＣＴ端末としての機能を有する自動車のことであり、車両の状態や周囲の道路状況などの様々なデータをセンサーにより取得し、

ネットワークを介して集積・分析する」[11]ことで運転の安全性を高めたり、事故時には警察や救急に連絡する機能を有している。

このような技術を前提として、本稿で設定された問題を再び考えてみよう。「自らの意思」で人を傷害した自動運転車両は、インターネットサーバーを通じて常に他の自動運転車両と通信し、絶えず相互に情報交換をしている。各車両の情報は常に最新の情報に更新され、走行体験から得られたあらゆる情報が各車両に同期されているものとしよう。その場合、実際に「自らの意思」で人を傷害した車両のＡＩと、その他多数の同型・同ＯＳのＡＩの間に差異はないことになる。なぜなら、一台の車両が「知っている」ことは、常に同期され、全ての車両が「知っている」ことになるからである。[12]

そうであるとすれば、最早一台の車両を「一人」の人格とみなすことは難しいのではないか。というのも、一台の車両は、無論物理的には一個体であることに間違いないが、各車両は独立した個人というよりは、ネットワークによって構成された情報の総体の一部、人間でいえば身体の一部のようなものと考えることもできるからである。一台の車両は個でありかつ全体なのである。[13]

そのように考えてみると、サーバーを介して接続されている全車両やそのインフラ全体を、総体として一人格とみなすと考えることも可能であろう。もしこの人格に「刑罰」を科すこととなった場合、先述の問題は部分的に解消される。すなわち、「自らの意思」で人を傷害した車両とは別車両ではあるが、その潜在的な危険性を有する車両をも「刑罰」に処することが可能となる。より厳密にいえば、個々の車両を各自処罰するというよりは、ネットワーク上に存在する情報の総体としての人格を処罰するがゆえに、その帰結として個々の車両が処罰されることになる。同種事故

の再発防止には効果が高いため、一般予防論的見地からは正当化されうるかもしれない。無論、その処罰によって自動運転機能という交通インフラ全てが麻痺しかねないことは自明である。なぜなら情報共有しているあらゆる機器やＡＩ、場合によっては情報インフラ全体がその「刑罰」の対象となるからである。

このように、自動運転車両に「刑罰」を科そうという場合、結局のところ「誰」を処罰すべきであるかを、一義的に決定することが難しい。というのも、自動運転車両のＡＩに人格を付与したところで、その一人格の範囲が確定できないからである。そして各車両単位で人格を付与するにせよ、サーバーを共有する全てのＡＩを総体として一個の人格を付与するにせよ、困難が生じるであろうことは既に述べたとおりである。

この点、法における人格は、法秩序が何を人格として扱うかであって、決して自然人を念頭に置いて構成された概念ではないという理解が存在する。

伝統的な見解は、法的主体という概念と人格という概念を同一視してきた。その定義によれば、人格とは権利・義務の主体としての人間である、とされる。しかし人間に限らず、人間以外の存在、すなわち一定の団体、たとえば組合、株式会社、自治体、あるいは国家も人格であるとされるのだから、人格という概念は権利と法的義務の「担い手」と定義される。人間に限らず、人間以外の存在も担い手としての機能を果たしうるのである。権利と法的義務の「担い手」という概念は、伝統的な法的人格理論において重要な意義を有している。権利と法的義務の担い手が人間である場合、自然人という語が用いられる。他方、権利と法的義務の担い手が人間以外の存在である場合、法人という語が用いられる。その際、自然人が「本来の」人格と考えられる一方、法人は「人工的

な」人格、すなわち法律学によって人為的に構成された、「リアル」ではない人格として対置される。たしかに、法人も「リアルな」人格であると論証しようとする試みも存するところではある。しかしその試みが徒労に終わることによってなお一層、いわゆる自然人も法律学による人工的な構成物であって、自然人もまた法人に過ぎないのだ、ということが明瞭になってくる。(14)

すなわち、法的人格のプロトタイプは自然人だというわけではなく、自然人も法人も、どちらも法律学による人工的な構成物であるという点に変わりはないというのである。このような理解を前提とすれば、次にロボットやＡＩをこの文脈でいうところの法人とみなしてよいかという問題が生じる。この問題について次のような見解がある。

社団法人と物のあいだにあるものとして財団（法人）が考えられるかもしれない。たとえば民事法においては、およそ財産でしかない物の集合がそれ自体として法人格を認められる場合が存在する。一般社団・財団法人法により設立される財団法人が典型であるが、民法九五一条は「相続人のあることが明らかでないときは、相続財産は、法人とする」と定める。この相続財産法人は、特に設立手続きを踏まずに物である財産に法人格が付与されるもので、財団法人として目的遂行のために主体として行動しうる。また、破産宣告がなされると破産者の財産は財団となるが、かつてはこの財団に法人格を認める説も主張されていた。すなわち「破産財団に法主体性を認める」説がそれであり、「権利義務の帰属点としての主体性を承認」し、明文の規定を欠くがいわば「暗星的法人」としての法人格を認めようとする兼子仁の学説が著名である。兼子は、破産法学の目的論を根拠にケルゼンを引

きつつ、かかる法人格肯定説がもっとも適切であると論じている。このように社会的有用性が高い場合に財産に法人格を与えること自体は珍しいことではない。(15)

このように結局のところ、法人格というものが有用性という観点から認められるのであるとすれば、その一個の人格の範囲に関しても同様の観点から画定する余地はあろう。つまり、どの範囲までが一個の人格であるか、ではなく、どの範囲までを一個の人格として扱うことが有用か、というアプローチで人格性の範囲を画定するのである。無論、その有用性の検討に際しては、上述したように「刑罰」として有意義であるかも考慮されねばならないであろう。

三　「刑罰」はロボットやＡＩにとって苦痛となりうるかという問題

上述のように、法的人格のプロトタイプは自然人ではなく法人であり、結局のところ法的人格の付与の可否は有用性によって決定されるがゆえに、その一個の人格の範囲も同様の考慮によって確定されるのであれば、「刑罰」を受けるのは誰かという問題は一応の解決をみるかもしれない。では「刑罰」を受ける主体が確定されたとして、その「刑罰」はロボットやＡＩに何らかの影響を与えうるであろうか。

「刑罰」は非難としての害悪であるとされ、「苦痛あるいは通常は不快であるとされるその他の結果を含むものでなければならない」(16)とされる。そして我が国の現行刑法が予定している刑種は、罰金刑、自由刑、死刑の三種である。これら刑罰が人に科された場合、死刑を除くこれら刑罰が有する「苦痛」を契機として、犯人が自らのなしたことの

重大性につき身をもって理解し、真摯に反省するということを、我々は期待する。しかしながら、そもそもロボットやＡＩにとって、現行法が予定する刑種から「苦痛」を感じることができるのであろうか。

罰金刑が人間にとって苦痛であると考えられるのは、金銭的負担によってたとえば欲しいものが買えなくなるだとか、衣食住に割かれるべきコストを節約し生活レベルを下げねばならないだとか、自らの欲求を我慢せねばならないことがその理由の一部となっている。しかしながら、そもそも財産権の主体となりうるかは別論として[17]、ロボットには（それがプログラムされない限りは）欲求というものが存在しない。また何らかのエネルギーの補給さえできれば、衣食住にコストはかからない[18]。それゆえ、罰金刑やそれに類似した何らかのリソース制限という形での刑罰は、ロボットやＡＩにとって苦痛とはならない。自動運転車両に罰金を科しても、ガソリンが買えなくなる程度であって、それによって苦痛を受けるのはむしろ、自動運転車両が使えなくなってしまう利用者たる人間の側であろう。

自由刑が人間にとって苦痛であるのは、移動の自由が制限されることによる不都合と、寿命という人間にとって有限のリソースを、（様々な自由が制限される）不自由な状態で消費せねばならないことに一因がある。しかしロボットやＡＩの場合、先述のように欲求がない。とりわけ自動運転車両においては、移動の自由を付与されたとしても、自ら目的地を設定してドライブを楽しむ欲求はない。また寿命もない以上は、自由刑によって消費されてしまうリソースも存在しない。したがって自動運転車両を刑務所（ないし物理的な移動を不可能にする駐車場など）に留めても、それは苦痛とはならない。

死刑に関しては、それが人間にとってのあらゆる自由の否定であることは言わずもがな、死に対する直観的な恐れも死刑という刑罰のもつ苦痛の一部であろう。繰り返しとなるが生への欲求がないロボットやＡＩには、死への恐れ

はない。また、先述したコネクテッドカーの場合、仮に「自らの意思」で人を傷害した車両をスクラップにして「死刑」に処したとしても、その「身体的」な「死」は意味をなさない。というのも、同車両のデータはインターネットを介してサーバーに保存されているのであり、罪を犯した車両の「悪しき意思」は社会から追放されていないことになるからである。それゆえ、人を傷害した自動運転車両をスクラップにして、それを仮に「死刑」と呼んだとしても、やはり「苦痛」にはなりえないであろう。

それゆえ現行刑法が予定している刑種によっては、ロボットやＡＩに対して「苦痛」を与えることは不可能であると言わざるを得ない。しかしながら、「刑罰」にとって、その刑罰を科される者が実際に苦痛を感じている必要はないとする見解も散見される。そこでは、刑罰を科される者が実際に苦痛を感じることは重要ではなく、①一般的・客観的にみて「害」となる取り扱いを受けているという事実が重要であるとする見解[19]や、②苦痛を感じているように「見える」ことが重要であるとする見解[20]がそれである。

前者の見解は、「刑罰は常に処罰される者にとって有害なものになるとは限らない」のであって、「例えば、鞭打ち刑はたいていの人にとって有害なものであろうが、打たれることにこの上ない快感を覚える者にとっては有害だとはいえない」、あるいはたとえば「ホームレスにとっては一般社会で当てのない生活を送るよりは刑務所で暮らした方がよほどましである[21]」とすれば、この刑罰は苦痛たりえないという主張に対して反論する形で生じてきた。同見解によれば、刑罰を受ける者が何に「苦痛」を感じるかという主観的事情には依存せず、客観的に有害であるということが刑罰の構成的要素であるという。「なぜならば、ある取り扱いを刑罰とみなすかどうかという問題と、その取り扱いを受けている者が実際に処罰されているかどうかという問題とは、切り離して考えることが出来るからである。仮

に、先の例のマゾヒストやホームレスは害を被っていないので実際には処罰されていないということを認めたとしても、鞭で打たれることや刑務所に閉じ込められることが一般に人に害を与える性質であることを認める限り、これらの措置を刑罰と呼ぶことは依然として可能である」[22]というのである。以下では便宜上、同説を「客観的害」説と呼ぶこととする。

刑罰を科される者が実際に苦痛を感じることは重要ではないとするもう一方の見解は、その刑罰が客観的に有害であるかではなく、その刑罰によって受罰者が苦痛を感じているように「見える」かが重要であるという（以下、この見解を「見える」説と呼ぶ）。すなわち、「ここで考えるべきは、人間であっても実際には刑罰が痛みとして効かない者がいるなかで、同じ人間であるからという理由で効いているようにみえることで、私たちは均質性を擬制している点である【傍点原文ママ】」[23]。つまり、実際には人間同士であってもある苦痛に対する反応は異なりうるから、他人が感じている苦痛がどれほどのものか、あるいはそもそも本当に苦痛を感じているのか自分にはわからないのであるが、刑罰を科される者が苦痛を感じているように「見える」ことによって、多かれ少なかれ自身が感じる苦痛と似たような苦痛を感じているのであろう、という推測が成り立つというのである。「自分が感じるのと同じような苦痛を感じているのであろう」という推測が、自己と他者との均質性を担保しており、この均質性こそが刑罰にとって重要であるという。

問題はロボットやＡＩが人間との均質性を有しうるか、である。

ここで重要な鍵となっているのが、制度全体を構成する我ら人間がすべて等しくかけがえのない生命を持って

おり、痛みや苦しみを感じうる主体であるという可傷性（vulnerability）への意識だと言うことは、おそらく許されるだろう。そこに存在するような痛み・苦しみを前提として、それが加害者（集団）にも等しく担われることが責任の実践なのだと考えるならば、そのような可傷性を持たず、また本質的に複製可能であってかけがえのなさを持たない（と我ら人間が想定する）ＡＩやロボットによって責任が果たされることはありえないということになるのかもしれない。我々がロボットやＡＩに対する刑事処罰という観念に納得し難い違和感を覚えるとすればそこにあるのは、そのようなかけがえのなさが我々とは共有されていない、均質性が存在していないという感覚なのだと思われる。【傍点原文ママ】[24]

しかしながら、「見える」説によれば、この「かけがえのなさ」に由来する均質性も、「ＡＩを意図的に個体化させ（られたように見せ）たり、他の方法で均質性を補ったりすることも考えられる」[25]という。

さて、上記の二説を見比べ、差異と共通点を抽出してみたい。差異としては、「客観的害」説は、一般人の目から見て通常有害であれば、刑罰が有すべき有害性という要素にとって十分であるとする一方、「見える」説からすれば、一般人の目から見て刑罰を科される者が苦痛を感じているように見えることが要求される。先述のマゾヒストの例を考えたとき、「客観的害」説からすればマゾヒストへの鞭打ちはなお有害性を有しうるが、「見える」説からすれば、打たれて悦に入っているようにしか見えないのであれば、有害であるとはいえないことになる。

このような差異は認められるものの、共通点もまた認められる。一般人の目から見て通常有害であるかだとか、苦痛を感じているように見えるかということを考えるときには、その外部的観察者である我々一般人が、実際に刑罰を

科される者と自らとの立場を交換したときに、苦痛を感じるであろうか、という推量が背後に控えている。立場の交換可能性が全くない場合、その対象にとって通常有害であろうと推測することはできないし、苦痛を感じているように見て取ることも難しいであろう。道端に転がっている石の立場になってものを考えることは難しいし、それゆえその石を蹴飛ばすことは石にとって通常有害であるのか、ないし苦痛を感じているように見えるのかを論じることは不可能であろう。

この立場の交換可能性はまた、観察者と立場の交換相手との一定の均質性への信頼を前提としている。すなわち、自身が感じることと、立場の交換相手が感じることが、ある程度均質であろうという信頼である。これを共感や同情[26]と言い換えることもできよう。

それゆえ、究極的には将来ロボットやＡＩに対して我々が共感を有しうるかが分水嶺となる。現状の技術の発展状況に鑑みれば、たとえば自動運転車両がたとえ非常に丁寧に運転をしてくれ、乗客を気遣ってくれたとしても、その車両に人間に対するのと同程度の共感や同情の念を感じることは難しいであろう。しかし、ロボットやＡＩとの交流を描いた映画やドラマなどを見て、感情移入した人々が涙を流すことは珍しいことではないから、ロボットやＡＩとの間に情緒的なつながりを持つことは原理的に排除されているわけではなさそうである。今後の技術の発展の次第によっては、ロボットやＡＩに対して科された「刑罰」を見て、応報が達成されたと感じることも可能となろう[27]。

ロボットやＡＩにおいても刑罰の「苦痛」を観念する余地が存在したとしても、他方で現行刑法上、ロボットやＡＩの処罰に適した刑種がないことは先に述べたとおりである。この点につき、今井教授は再プログラミングという措置こそロボットやＡＩに対する「刑罰」たりうるとして、次のように述べる。

刑罰の機能としての、過去の違法行為（違法な法益侵害を惹起した行為）を回顧的に非難する点（刑罰の応報的機能）との関係では、違法結果を惹起させたプログラム（その該当部分）の削除が、ＡＶ【自動運転車両の意：引用者注】に対する社会の応報感情に即した対応となろう（特に、ＡＶがＮＰ【自然人の意：引用者注】の倫理的判断を代行するようになれば、そうした判断を下した部分のプログラム除去が必要となろう）。これができない場合には、一定期間、プログラムを作動させないこと（temporarily inactivation）で対応すべきであろう。（自動運転システムを構成する）プログラムの全面消去は、ＡＶとの関係では、その死刑に相当するが、その執行は、比例原則（犯罪結果と刑罰との比例制維持の要請）を踏まえて判断されるべきである。

他方で、将来の違法行為を防止するという刑罰の（予防）目的からも、違法結果を惹起させたプログラム（その該当部分）の削除が要請されよう。これに加えて、同種結果の再発防止に向けて、プログラムの改良という措置も要求されるように思われる。[28]

たしかに、再プログラミングを科すのであれば、人間でいうところの再犯防止にも資する。しかしこれに対しては思想刑にあたるのではないかとの懸念も示されている。[29] ロボットやＡＩにも人格性が肯定されたうえで、再プログラミングを科すとすれば、そのロボットやＡＩの「人格改変」を許すことにはなるのではないか、という懸念である。もしこれを許した場合、問題はロボットやＡＩのみにとどまらず、我々人間にも及ぶ。というのも、人格性を有する存在に対して思想刑（洗脳）を許すというのは、人間に対しても同じことがいえるからである。この逆推論は、科学

的去勢の問題を改めて考えるきっかけとなろう。[30][31]

四　近代刑法における意味での「刑罰」と呼べるかという問題

刑罰の実効性の見地から、仮に再プログラミングをロボットやＡＩに対して科し、かつその措置が我々人間の目から見て（あるいはほかのロボットやＡＩの目から見ても）、苦痛を伴うものであろうと思える場合、それを「刑罰」と呼ぶ余地も出てくるように感じる。たとえば、見た目も仕草もほぼ人間と変わらず、普段コミュニケーションを楽しむことが出来るようなパートナーロボットが、再プログラミングの苦痛に悶えている姿を見て、多少なりとも「かわいそうだ」と思うのであれば、この措置を「刑罰」と呼ぶことの違和感は徐々に小さくなってくるのではないだろうか。

しかし、なお検討を要するのは、この「刑罰」が近代刑法の諸原則と矛盾することは本当にないのかという点である。これは**二章**にて検討した人格の一個性の問題にも若干関係している。

すなわち、**二章**にて述べたように、ロボットやＡＩ、あるいは本稿の関心でいえば自動運転車両に各自の人格性を付与した場合、スタンドアローン状態で運用されているのであれば、悪事を働いたロボットやＡＩだけが再プログラミングという「刑罰」を受ければ事足りるのであるが、その「悪しき意思」がインターネットを通じて他のロボットやＡＩに共有されてしまう場合、すなわちコネクテッドカーを想定するとき、その「刑罰」にはほぼ全くといってよいほど意味はない。その「刑罰」によってはその「悪しき意思」を撲滅することは一切できないからである。

この問題の対策として、一定の利用目的の場合には、ロボットやＡＩに相互通信機能を持たせないだとか、相互通

信の内容を制限するといった方法も考えられよう。たとえば、話し相手になってくれるパートナーロボットのような技術を念頭に置く場合、そのパートナーロボットは所有者である人間のプライバシーに関わる情報（名前や住所は無論のこと、場合によっては恋愛や病気、信条や職業に関わること等）を保有しうる。そのような情報を他人のパートナーロボットと共有することは重大なプライバシー侵害であるし、そもそもそのような情報をパートナーロボット間で共有する必要性も低い。このような用途のロボットやＡＩ等は、スタンドアローンでの運用の余地があり、この種のロボットやＡＩが悪事を働いた場合は、そのプログラムを削除するなり改善する再プログラミングにも意味は認められよう。

では自動運転車両の場合はどうか。結論から言えば、自動運転技術はスタンドアローン状態での運用は想定しにくい。というのも、自動運転技術においては、常に他車との情報交換をすることによって、道路の混雑状況だとか車間距離だとかについての情報を得て、その都度の判断を形成する。もし自動運転車両がスタンドアローン状態で、情報のアップデートがされないまま運用されたならば、道路工事によってあけられた大穴に飛び込んでしまう可能性が生じうる。

しかしながら、各車両を一人格として捉えた上で、人を傷害した一台の車両のみを「刑罰」に処することの無意味さは先に述べたとおりである。おそらく、再プログラミングによって取り除かれた「バグ」に関するアップデートデータは、同車種・同型ＡＩにも適用されねば、再発防止の観点からは全く意味をなさない。

しかしながら再プログラミングが同車種・同型ＡＩにも科される場合、近代刑法の原則との抵触が生じる。すなわち、個人責任の原則との抵触である。というのも、たとえば一台の自動運転車両が事故を起こしたことを理由に、全ての同車種・同型ＡＩに対して再プログラミングという「刑罰」が科されるのであれば、これは他人の犯罪を理由と

して、犯罪を犯していない者までもが処罰されるという連帯処罰と構造的には全く同じであることになる。しかしながら、「刑罰は、実際に犯罪を犯したもの、あるいは犯罪を犯したと思料される者に対して、その犯罪を理由として科されねばならない」[32]のであって、連帯責任は近代刑法の原則に反する。

無論、近代刑法と理論的に整合しないことがロボットやＡＩに対する措置を直ちに無為なものとするわけではない。ロボットやＡＩ、自動運転車両といった先進技術に関しては、既存のシステムの枠内で問題解決を図ることは重要なのではあるが、その枠に収まりきらない場面も当然に想定される。その場合、従来の刑法の原則との抵触があるのであれば、新たなパラダイムへの移行が求められるのであって、ロボットやＡＩに対する統制手段がおよそ挫折してしまうわけではない。

五　ロボットやＡＩを「処罰」することによる「刑罰」という語の意味変容という問題

仮にロボットやＡＩが責任主体として法における人格であると認められ、ロボットやＡＩの有する人格の範囲が確定され、ロボットやＡＩにとって再プログラミングという措置が苦痛を伴うように見えるようなかたちで執行され、そしてその措置が近代刑法の原則と抵触する部分については、ロボットやＡＩの領域に限ってその原則を修正する（あるいは二章にて述べたように、ネットワークを通じて情報共有をしているシステム全体を一人格とみなす）ことによって、ロボットやＡＩに「刑罰」を科す種々の障壁が解消されたとしよう。その場合、理論的にはロボットやＡＩに「刑罰」を科すことも可能となっているはずである。しかしながら、ロボットやＡＩに「刑罰」を科すことによる副作用はないの

であろうか。

この副作用を考えるにあたって、本稿が扱う問題とは毛色が異なるものの、問題意識としては共通しているように思われるため、少し長くなるが伊藤教授のリスク社会に関する論考を以下に要約する。(33)

伊藤教授によれば、リスク社会化と厳罰化という本来矛盾する二つの動向が同時に進行しているという。同論文は、リスク社会化と厳罰化がいかなる点で本来異なるものであり、そしてなぜこの両者が交差するのか、この両者が交差することによって犯罪統制はいかなる方向に変化しうるのかを問う。

リスク社会においては、主に統計的手法によって犯罪のリスクを評価・把握し、犯罪統制を最適化しようとする。その統計的手法がゆえ、犯罪者個人に着目するのではなく、コスト・ベネフィットの観点から総数としての犯罪の最小化が目指され、事前的・予防的な犯罪統制が志向される。またリスクファクターの集団として犯罪を把握するため、犯罪者個人の内的なものを問題としないという意味で「脱モラル」的でもある。他方、厳罰化は刑罰という事後的・懲罰的な犯罪統制を志向する。また、犯罪統制を最適化するというよりは、その過剰さにおいて感情的であり、犯罪者個人をモラル的に非難することを求めるという点で、リスク社会と厳罰化は対極に位置するという。

ではリスク社会化が進展する現在、なぜその対極に位置する厳罰化が志向されるのであろうか。伊藤教授によれば、現在の厳罰化の流れは公衆・メディア・政治の三者の相互作用により形作られているという。すなわち野心的な政治家が、メディアを通じて殊更に市民の不安を煽り、犯罪に対する強硬策を打ち立てることで支持を集めるという方法により、厳罰化が形成されるというのである。従来はプロフェッショナルである我々研究者がこのような感情的な厳罰化傾向に歯止めをかけてきたのであるが、現代において政治家はメディアを通じて直接的に市民へのアピールをす

るため、我々プロフェッショナルはバイパスされてしまっている。あるいは、我々プロフェッショナル自身が厳罰化を志向している、という可能性もあると指摘される。

無論、このような傾向は、安易に「ポピュリズム」として非理性的な感情の表出ととらえることも適切ではないという。論者によってはこれを「民主化」と呼ぶ向きもあり、そのようにとらえた場合、非合理な感情に対する理性的抑制としてのプロフェッショナリズム、という枠組みで語ることはできなくなる。それゆえプロフェッショナルはいかなる使命を負うのかが問われているのである。それを明らかとするためには、厳罰化とリスク社会化という現状がもたらす犯罪統制への影響を概観せねばならない。

犯罪統制は①司法制度による法的な統制、②ソーシャル・ワーク組織による規律的な統制、③犯罪予防機関による保険数理的な統制、という三者のバランスによりなされる。リスク社会論との関係で論じられるのは③保険数理的な統制である。というのも、保険数理的な統制は、統計的手法により犯罪を把握し、そのリスク評定に応じた予防手段を採用することにより、全体としての被害を最小化することに関心を寄せており、リスク社会論と共通の関心を有しているからである。他方、一九七〇年代以降の社会復帰思想の衰退を受け、これとの関係で②規律的な統制が後退する。また、同時期の福祉国家主義から新自由主義への転換もあって、この後退は決定的なものとなる。

これにより、リスクに対する事前的な予防であるところの保険数理的な統制と、規律的な統制を欠いた、すなわち社会復帰を目指さない事後的な制裁であるところの法的な統制の二極によって犯罪統制がなされることとなる。この再社会化を志向しない事前的予防・事後的制裁の接合は、「排除社会」をもたらしうる。

しかし先述のように、この法的な統制における厳罰化をいかに解するか、すなわち単なる感情的な処罰欲求ととら

えるべきか、それとも刑事司法の民主化ととらえるべきかによって、プロフェッショナルとしての我々の使命は変化しうる。この際に、ただプロフェッショナリズムの「理性」の高みから、ポピュリズムの「感情」の非合理を慨嘆するだけの立場からは、排除社会の到来に対する、いかなる有効な抵抗の手立ても生じないという。

以上が同論文の要約であるが、同論文をそのまま本稿の問題意識にあてはめることはできないため、今一度、伊藤教授の主張を整理してから、抽象化・構造化し、本稿の問題にパラフレーズすることを試みたいと思う。

伊藤教授によれば、構造的に全く異なるものであるはずのリスク社会化と厳罰化が同時に進行しており、これはメディアを媒介したプロフェッショナルのバイパス、あるいはプロフェッショナル自身の内的変化によってもたらされうるという。これを犯罪統制の観点から分析すると、前者のリスク社会論は保険数理的な統制、すなわち事前的予防、後者の厳罰化は法的な統制、すなわち事後的制裁に対応する。そして社会復帰思想の後退により、再社会化が犯罪統制の目的から抜け落ちたがために、事前的・事後的排除を志向する社会が到来しかねない、という。しかし注意すべきは、この厳罰化を単なるポピュリズムと見るのか、犯罪統制の民主化と見るのかによって、我々プロフェッショナルの使命は変わりうるのであって、ただ居丈高にポピュリズムの非合理を嘆くだけでは何の役にも立たないというのである。

これをやや抽象化すれば、以下のようになる。現在の厳罰化傾向が仮に、自らが犯罪被害を被ることへの不安感によって、予防的に用いられるのであれば、犯罪統制は危険回避・不安解消のための単なる排除手段に堕しかねない。他方、この厳罰化傾向が犯罪統制の民主化を意味するのであれば、我々プロフェッショナルに求められるのは、この厳罰化傾向を一方的に抑制することではなく、その理論的基礎付けを提示するアドバイザーとしての使命である。

これを本稿の問題意識にパラフレーズしてみたい。これまで述べてきたように、ロボットやＡＩに「刑罰」を科すことは、理論構成によっては全く不可能であるわけではない。技術的な条件さえそろえば（再プログラミング措置を受ける際には「苦しそう」な表情を浮かべるよう設定しておくなど）、ロボットやＡＩに「刑罰」を科し、それによって市民が満足を得ることもできよう。

しかしその刑罰への欲求が、「自動運転車両によって自分が被害を受けるかもしれない」という単なる不安や、「もし自身が被害を受けた場合に、誰も責任を取らずに泣き寝入りせねばならないなど、許せない」というストレスに起因するものなのであれば、感情的なポピュリズムという問題性は本稿における議論にもあてはまる。(34)

しかしまた他方で、ＡＩやロボットに対する「刑罰」をはじめとした法規制は、今後避けては通れない問題であろうし、単なる先進技術への不安感であるとして片づけることはできない。「責任の所在を明らかにする」「答責の間隙を埋める」ということは我々プロフェッショナルの使命だからである。

現在、責任の所在が不明瞭である領域に、新たな責任主体（電子的人格）を構成して、その主体に「刑罰」を科すための法的・技術的整備を進めることは、今後不可避となろう。しかしその動機が、単なる不安解消のためのスケープゴートをつくることの一環なのであれば、「刑罰」がストレス解消手段に堕してしまうこととなる。そうなのであれば、我々はその「刑罰」の意味変容に対してプロフェッショナルとして反対せねばならない。そうではなく、先進技術の利点を享受し、欠点を補う法規制により、より良き社会を構想する一環としてＡＩやロボットに「刑罰」を科すことが検討されるのであれば、刑罰は責任非難であるという意味の枠内において、我々プロフェッショナルは市民的討議に道筋を示すことが使命である。

伊藤教授は、厳罰化を例にとりながら、ただペシミスティックに現状を嘆くだけの「プロフェッショナル」は、何の役にも立たない（Nothing Works）と戒めておられたのではなかろうか。

六　おわりに

以上の検討を要約すれば、次のようになる。ロボットやＡＩに対して、たとえば事故を起こした自動運転車両に「刑罰」を科す際、どこまでが一人の人格なのかという問題が生じるものの、その範囲を有益性の観点から画定する余地もありうる（二章）。その一人の人格に対し、現行法は有効な刑種を予定していないが、再犯防止の観点からは再プログラミング措置が考えうるところであり、その措置を施す際にロボットやＡＩが苦痛を感じているように「見える」のであれば、この問題も克服しうる（三章）。再プログラミングという措置の技術的特性上、未だ犯罪を行っていないロボットやＡＩないし自動運転車両にも、刑罰内容と同様のアップデートを施す必要があるため、個人責任の原則との抵触が考えられるが、ロボット法という新たな制度構築や原則の修正によって、あるいは二章にて検討した人格の範囲を改めて検討することによって、この問題もまた克服される余地はある（四章）。以上のように、ロボットやＡＩに「刑罰」を科すことは、理論的には全く可能性が無いわけではないが、その「刑罰」は真に「犯罪に対する責任非難」という枠内で用いられているかはなお慎重に検討されるべきである。単なる被害への不安から、スケープゴートとしてロボットやＡＩに「刑罰」を科すのであれば、それは「刑罰」の意義を変容させてしまう（五章）。

なお本稿で得た結論であるところの、ロボットやＡＩに「刑罰」を科すことは、若干の無理を承知でいえば、理論

的には全く可能性が無いわけではない、というテーゼは、本稿で扱った四つのテーマを検討した限りでは、というにとどまる。当然、本稿で検討していない、想像だにしない問題点がなお存在しうるところではあるし、本稿で扱った論点に関しても検討が不十分であった点もあろう。それらは今後の課題となろうが、今後不可避となってくるロボットやＡＩへの法規制を巡る議論に、問題提起として一石を投じることになるのであれば幸いである。

＊　＊　＊　＊　＊

少しばかり伊藤先生の思い出についてここで語ることをお許しください。

私自身は刑事法専攻に身を置きながらも、伊藤先生から直接のご指導をいただいたことはありませんでした。しかし研究会などでご一緒させていただく機会があるたび、伊藤先生は気さくにお声をかけてくださりました。個人的な食事会を企画したこともあります。また懇親会などお酒の席ではその朗らかなお人柄で、私のような大学院生にも分け隔てなく、（ここでは書けないようなことも含め）いろいろな逸話・裏話を教えてくださいました。

伊藤先生のそのおおらかなお人柄を象徴するエピソードをひとつ、この場を借りて紹介させていただきたいと思います。

とある研究会が終わり、その後懇親会を催すため居酒屋に入りました。最初の乾杯用のビールを注文し、店員がジョッキを配る際、その店員が手を滑らせて机の上のジョッキを倒してしまいました。それによって不幸にもジョッキ一杯分のビールが、その机の真下にあった伊藤先生の鞄にかかってしまいました。革製の鞄はずぶ濡れになり、鞄

の中までビールが滲みているであろうことは明白でした。周りの参加者は騒然とし、「パソコンや締め切り前の原稿等が入っているのではないか…」と心配する声が上がり始めました。それを聞いた店員は、顔面蒼白となってしまい、呆然と何をすべきか分からなくなってしまったようでした。私も、これは店長を呼ぶなどして、損害を確認するなり苦情を言うなりしようとしかけたところでした。しかしそのとき、伊藤先生ご本人はいたって穏やかに、笑いながら店員に言いました。「その鞄、ビール好きなんだよ。なんたって俺の鞄なんだから。」と。ご自身にとって不利益となることを被ったにもかかわらず、一切店員を責めることもせず、これほどまでに粋で優しいことを、しかも咄嗟に言えるものであろうか、と非常に感動した覚えがあります。

いつも穏やかに笑っておられる、お優しい先生でした。早すぎるご逝去が残念でなりません。心よりご冥福をお祈りいたしますとともに、拙いながら本論文を伊藤先生に捧げます。

（1） *Susanne Beck*, Grundlegende Fragen zum rechtlichen Umgang mit der Robotik, JR 2009, S. 225-230. など。なお近時の議論状況を概説するものとして拙稿「ロボットの可罰性を巡る議論の現状について」比較法雑誌五一巻二号一四五頁以下。

（2） 本稿で「自動運転車両」という場合、いわゆるSAE規格にいうレベル4以降を指し、初期設定完了後は人間の介入が無くとも一定の裁量のもと、自ら道路状況を判断して走行する車両を指すものとする。つまり、目的地を設定して発車した以後は、その車両の乗客は一切運転に関与することが出来ない、という状況を想定されたい。

（3） 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、たとえば今井猛嘉「自動化運転を巡る法的諸問題」IATSS Review Vol. 40, No. 2. 五六頁以下。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―」愛媛法学会雑誌四三巻三・四合併号一頁以下、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. 三三頁以下、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. 二九頁以下。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. 四一頁以下など。

（4） *Eric Hilgendorf*, Können Roboter schuldhaft handeln? Zur Übertragbarkeit unseres normativen Grundvokabulars auf Maschinen, in : Susanne Beck (Hrsg), Jenseits von Mensch und Maschine. *Jan. C. Joerden*, Strafrechtliche Perspektiven der Robotik, in : Hilgendorf / Günther (Hrsg), Robotik und Gesetzgebung.

（5） *Susanne Beck*, Intelligente Agenten und Strafrecht. Fahrlässigkeit, Verantwortungsverteilung, elektronische Personalität, Studien zum deutschen und türkischen Strafrecht-Delikte gegen Persönlichkeitsrechts im türkischen-deutschen Rechtsvergleich (Band 4), Ankara 2015, S. 179–195. 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』」千葉大学法学論集三一巻三・四号頁以下。*Sabine Gleß = Thomas Weigend*, Intelligente Agenten und Strafrecht, ZStW2014 ; 126(3). *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.

（6） *Sacha Ziemann*, Wesen Wesen seid's gewesen?, in : Hilgendorf / Günther (Hrsg), Robotik und Gesetzgebung.

（7） ロボットはせいぜい電卓に毛が生えたようなもの、ということを前提としている論者（たとえば Joerden, a. a. O(Fn. 4)）と、ロボットと人間との差異は、金属でできているか蛋白質でできているかの僅かな違いでしかない、ということを前提とする論者（Hilgendorf, a. a. O (Fn. 4)）とでは、合意に達する望みは薄い。

（8） *Greß = Weigend*, a. a. O. (Fn. 5). *Asaro*, op. cit (note 5).

（9） この問題を暗示するものとして、川口浩一「ロボットの刑事責任 ― ロボット刑法序説」『市民的自由のための市民的熟議と刑事法：増田豊先生古稀祝賀論文集』一一五頁以下。

（10） ロボットやAIに所有者情報や識別符号を付与して、その識別符号ごとに人格を付与することを検討するものとして Alain Bensoussan, Le droit de la robotique : aux confins du droit des biens et du droit des personnes - «Une démarche éthique est indispensable dans la construction d'un droit de la robotique »

（11） 総務省「平成二七年版情報通信白書」一八三頁。

(12) このように同一プログラムによる「累積効果」について危惧するものとして *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. 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?』」千葉大学法学論集三一巻三・四号頁以下。

(13) 大屋雄裕「人格と責任―ヒトならざる人の問うもの」『AIがつなげる社会』三五九頁。

(14) Hans Kelsen, Reine Rechtslehre, 2. vollständig neu bearbeitete und erweiterte Auflage (1960), S. 176.

(15) 小林史明「権利主体性の根拠をAI・ロボットから問い直す」『市民的自由のための市民的熟議と刑事法：増田豊先生古稀祝賀論文集』一四六頁以下。

(16) *H. L. A Hart*, Prolegomenon to the principle of punishment, "Panischment and Responsibility"p. 4.

(17) 所有権などの民事法上の一部の権利をロボットにも認めようとする試みとしてAsaro, op. cit. (note 5.)。財産主体とみなすものの、自ら積極的に財産を処分することに係る民事法的な財産権ではなく、便宜的に賠償主体とするための保険法上の財産権を肯定することに解決の可能性を見出すものとして *Beck*, a. a. O. (Fn. 5)。

(18) 罰金刑によってエネルギーの補給が出来なくなるよう仕向けることも出来ようが、そもそもエネルギー不足による機能停止やリソース制限による「不自由さ」がロボットやAIにとって「苦痛」となるかは疑問である。

(19) 高橋直哉「刑罰の定義」駿河台法学二四巻一―二号一〇二頁。

(20) 小林・前掲注(15) 一五一頁。

(21) 高橋・前掲注(19) 一〇二頁。

(22) 高橋・前掲注(19) 一〇二―一〇三頁。

(23) 小林・前掲注(15) 一五一頁。

(24) 大屋・前掲注(13) 三五八頁以下。

(25) 小林・前掲注(15) 一五二頁。

(26) 小林・前掲注(15) 一五二頁は、これを「愛着」とも表現する。

(27) 小林・前掲注(15) 一五二頁は、結局のところロボットやＡＩに「愛着」を感じることが出来るかが重要であるとする。

(28) 今井猛嘉「自動車の自動運転と刑事実体法―その序論的考察」『西田典之先生献呈論文集』五二九頁。

(29) Sacha Ziemann, Wesen Wesen seid's gewesen?, in : Hilgendorf/Günther (Hrsg), Robotik und Gesetzgebung.

(30) 姜暻來「韓国における性犯罪者に対する化学的去勢―性暴力犯罪者の性衝動薬物治療に関する法律の概観」比較法雑誌四六巻二号七五頁以下。

(31) なお思想刑であるという点から直ちに悪しきことであるといえるかについて、小林・前掲注(15) 一五二頁の脚注において「もちろん危険除去や社会防衛を重要視するならば、洗脳教化刑を忌避する私たちのほうに問題があるのかもしれず、ＡＩやロボットの刑罰論を検討することによって私たちが啓蒙されることになるのかもしれない。」と述べられている。

(32) Hart, op. cit (note. 16) p. 5.

(33) 伊藤康一郎「理性と感情―リスク社会化と厳罰化の交差」犯罪社会学研究三二号七八頁以下。

(34) 事実、アメリカの自動車メーカー Uber の公道実験での事故を巡るニュースにおいては、事故当時乗員であった Uber 社員がスマートフォンを見ていた様子が繰り返し放映され、メディアは「嬉々として」自動運転車両や同社員の過誤を報じていた印象がある。伊藤教授も上記論文にて指摘しているように、メディアは種々のリスクを報道することにより、市民に直接的に不安を植え付ける機能を有している。

（本学大学院法学研究科博士課程後期課程在籍）