

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」
Susanne Beck, Intelligente Agenten und
Strafrecht - Fahrlässigkeit,
Verantwortungsverteilung, elektronische
Personalität

根津 洗 希

- I. 導入
- II. 文献紹介
 - 1. 過失—従来のアプローチ方法
 - 2. ロボット工学における答責分配
 - 3. 過失概念の修正
 - 4. 電子的人格概念の導入
 - 5. 法的構造が有する社会への影響
- III. コメント
 - 1. 過失の修正について
 - 2. 電子的人格概念について

I. 導入

本稿にて紹介する論文は、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」である⁽¹⁾。同論文は2014年2月7日にスイスのバーゼル大学のシンポジウムにて、同教授が英語で報告した内容をドイツ語に翻訳したものであるが、その後、*Studien zum deutschen und türkischen Strafrecht - Delikte gegen Persönlichkeitsrechte im türkischen-deutschen Rechtsvergleich* (Band

4)に所収されるに至っている。

同論文の問題意識は、前号にて紹介した論文⁽²⁾と同様、先進テクノロジーに対して刑法はどのようにかわるべきか、という点にある。文献紹介に入る前に、本論文にて検討の前提となっている技術や用語法について補足する。本論文が想定している「自律的機械」あるいは「自律的ロボット」は、学習能力があり、その学習に応じて状況変化に順応可能な機械ないしロボットを意味する。つまり、出荷前のプログラミング時には必ずしもあらゆる状況においてロボットがとるべき行動について細かな指示がインプットされているのではなく、利用者に最大限のメリットをもたらすという基本的な大原則のようなもののみがプログラムされ、利用者自身がある程度このロボットを学習させ、訓練していくことを前提としているのである。ここでいう「自律」という語は、「一定の裁量を有している」という最広義の意味で用いるにとどまり、必ずしもそのロボットが自己決定権を有している、ということまでをも含意するものではない。

こういった自律的ロボットを用いる際には、自ら学習し、状況変化に合わせて自らを変更していくという性質上、このロボットがある状況でいったいどのような行動をとるのか、というのは厳密に予測することができない。また常に学習・最適化によりプログラムを自己改変しているがゆえに、いったいなぜその時にそのような行動をしたのか、ということを実後的に解明することも不可能である。本論文ではこのようなロ

(1) 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. 179-195.

(2) 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*, 2015, S. 9 ff. 文献紹介として、拙稿「文献紹介 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187頁以下。

ロボットが誤って他人に損害をもたらした場合、どのような解決が考えられうるかが検討されている。このような問題意識のもと、以下では本論文の「はじめに」と「おわりに」を除いた各章を要約して紹介していく。

II. 文献紹介

1. 過失—従来のアプローチ方法

従来からの過失犯論に自律的ロボットの問題をあてはめるとどのような困難が生じるであろうか。過失犯が成立するためには、注意義務違反と危険の予見可能性が必要とされるところ、以下それぞれ検討していく。

現在、ロボット工学において注意義務画定の参考となる基準はほぼ存在しない。注意義務を画定する目安として、工業規格であるISO規格やDIN規格は存在しているが、この規格を用いる際にはいくつか注意せねばならない点がある⁽³⁾。

第一に、これら規格は非国家的組織が策定した民間規格であって、必ずしも刑法上の注意義務の基準とは一致しないという点である。刑法は部分的に社会的な道徳をも保護するものであるから、民間規格が保護しようとする利益と完全に一致するものではなく、したがって民間規格が遵守されていた場合であってもなお過失犯の成立する余地はありうる。刑法と民間規格の保護目的が異なることを示す例として、DIN規格における身体的侵害の受忍基準と、刑法にいう「傷害」概念の相違が挙げられよう。DIN規格において、ロボットといっしょに作業する人間が怪我をした場合、その怪我が出血を伴わない場合には、身体的侵害の危険は受忍されうべきものであり、出血に至った場合には受忍されえず、すなわち規格違反であるとされる。しかし刑法における傷害概念は出血の有無ではなく、重大な身体傷害とそうでない場合を区別しているのである。

第二に、民間規格というのは非国家的組織が策定するものであるか

(3) 民間規格が注意義務に与える影響を論じるものとして、スザンネ・ベック「過失解釈における民間規格の意義」谷井悟司(訳)比較法雑誌50巻2号117頁以下。

ら、そこで策定された基準は民主主義的な正統性を欠いている。というのも、ある規格を策定する際に、その策定に携わる委員の選任方法や、その規格の起草、最終的な採決方法などが明らかではないばかりか、その規格に関連する業界の専門家が一人もいないということすらある。

しかしながら、以上のような難点はあるものの、実務的には民間規格に頼るところが大きい。このような民間規格に伴うオフィシャルなイメージから、多くの研究者や製造者はこれら民間規格を遵守していれば法的な問題はないと思っているからである。もし開発者がプログラミングしたロボットが他人を傷付け、その開発者が「民間規格に従ってさえいればよい」と確信していたような場合、その確信は注意義務違反の認定には影響を与えないが、違法性の意識を否定する余地はあろう。

過失犯の成立のためには第二の要件として、危険の予見可能性が要求される。自律的ロボットを利用する際に、このロボットが何かしらの損害をもたらすということは容易に想像できる。自動運転自動車や、軍事ロボットなどは格好の例であろう。しかし、具体的な結果の予見や因果経過の基本的部分などの予見は、ほぼ不可能である。というのも、学習・順応機能により自律的ロボットの具体的な行動は予測不可能になるためである。刑法上要求される予見可能性を、当該具体的事情や因果関係などに及ぶ具体的予見可能性とした場合、これが肯定される余地はほぼないが、他方で「およそ人を危殆化しうる」という抽象的予見可能性と捉えた場合に、これは広範に過ぎるものとなる。

2. ロボット工学における答責分配

従来の過失犯論による解決の難しさは上述の通りであるが、自律的ロボットが人を傷付けた場合の問題の複雑性はこれらの点にとどまらない。というのも、そのロボットには多くの人間がかかわっているからである。つまり、そのロボットの過誤行為の原因は、学習・順応プログラムを設定したプログラマーの行為なのか、製造・販売をした製造者の行為なのか、訓練・調整をした利用者の行為なのかをつきとめるのは困難であり、誰にどれだけの過失があるかを認定することはほぼ不可能であるといえる。

このような場合の答責分配の在り方として、以下の4つの方法が考えられうる。①関与者のうちの一人、たとえば利用者などを一律に答責的であるとみなす方法。②関与者のうち、過誤行為が立証できた者だけ答責的であるとみなす方法。③ロボットに関与した全ての人間を新たな法的人格として捕捉し、その法的人格が答責的であるとみなす方法。④そのロボットが社会的にみて重要なものであると捉え、そのロボットの危険性やそれまつわる答責性は社会全体で分担するという方法。これらの解決方法の背景には、ロボットの利用によって利益を得る者は、ロボットの過誤に対し金銭的に、場合によっては刑罰的にも制裁を科されるべきであり、その機械の態度はその受益者に帰属されるべきである、といった考えがある。たとえその機械の態度が、およそ人間の想像を超えていたとしても、である。

しかしながら、ロボットの文脈においては、過誤行為や損害賠償に対する答責のみが問題となるわけではない。望ましくない副次的作用に対する答責というものも忘れてはならない。また、刑法は回顧的に犯罪を非難するという機能を有するにすぎないから、こういった現代的なリスクを包括的にコントロールすることには向かないのである。

3. 過失概念の修正

以上示してきたこの複雑な問題に対して考えられる解決策として、ひとつに過失概念の修正がありうる。以下ではその修正過失犯論の骨子を示していく。

この修正過失犯論においては、最早危険の抽象的予見可能性は刑罰的答責の指標にはなりえない。というのも、自動車の利用や原子力発電など、統計学上危険性は存在するものの、社会的に受容されているものは多々あるからである。そのため注意義務の基準にウエイトが置かれることとなるが、この注意義務の基準に民間規格や刑法外の規定が用いられるべきでもない。なぜなら先に述べた通り、民間規格や刑法外の規定は、ある行為のメリット・デメリットを比較衡量するものであって、このような規範は必ずしも刑罰規範の保護目的とは一致しないからである。したがって、このような利益衡量は直接に注意義務の基準画定に用

いられるのではなく、ロボットの有用性と、ロボットが有する危険性を衡量した結果、なお社会的に受容されるものといえる場合に、「許された危険」の枠内で行われるべきである。また、その過失の答責は人間のみならず、ロボットも負いうるようなかたちで構想されねばならない。

この過失の答責分配の際に重要な要素が「信頼の原則」である。「信頼の原則」によれば、あらゆる者は原則的に他者の適法行為を信頼して良いことになる。ロボットの文脈においてこれが意味をもつのは、たとえば自律的ロボットが人を侵害した場合でも、プログラマーや販売者、利用者はそれぞれ互いに適法行為を期待しうる限りでは、誰も責任を負うことはなく、期待された適法行為を行わなかった者だけが侵害結果に責任を負うことになる。このように考えると、答責分配の結果、誰にも答責が残らないあるいは誰の答責も立証できない場合などには、誰も処罰されないということも考えられうる。

これに対して、以上のような答責の間隙を埋めるために、利用者に一律に責任を負わせるというアプローチもありうる。その場合、プログラミングや製造時の明らかな構造的欠陥がある場合を除けば、ロボットに代行させた行為のすべてに利用者が責任を負うこととなる。しかしながらこの帰結は受け入れられない。というのも、自律的ロボットに判断を移譲し代行させるのは、人間がその判断から解放されるためであって、つまり人間の負担を減らすためだからである。もし判断を移譲したとしても、そのロボットの行動にすべて利用者が責任を負わねばならないとすれば、結局人間の負担は実質的には減っていないのではなからうか。たとえば、長距離の運転でロボットによる自動運転モードを利用した場合、たしかに進路や速度調整はそのロボットが代行してくれるが、判断ミスによる事故などがあった場合にその利用者がその事故の責任を負うとすれば、利用者は休まずそのロボットの運転を注視せねばならず、結局自分で運転している場合と同程度の注意を向けねばならないことになってしまう。このような帰結は自律型ロボットを用いる意義からして本末転倒であるから、やはり答責の間隙は避けられないものと捉えた上で、むしろそのような間隙を社会が受容するようになってから規制がなされるべきである。

4. 電子的人格概念⁽⁴⁾の導入

過失犯論を修正するほかに、自律的ロボットという複雑な問題を解決する方法として考えられるのは、「電子的人格概念」の導入である。これは、企業などが法人として法における人格性を付与され、裁判などで企業が直接賠償命令の名宛人となるのと同様に、ロボットにも似たような人格性を付与する、というアプローチである。この人格性はあらゆる人工知能に観念可能である。

ここで注意せねばならないのは、ここでいう人格性はあくまで付与され、人格として扱われるということの意味するのには過ぎないのであって、ロボットを人間とみなしたり、人間に準ずる存在であると考えer必要はないということである。知能を有しているだとか、学習能力があるといった存在論的な理由からその人格性が付与されるわけではない。

逆に人間との類似性を理由として人格性を付与するということになると、様々な不都合が生じる。人間との類似性や、「人格を人格たらしめるもの」を有しているから人格性を付与するという場合、そもそも人間自体がその「人格を人格たらしめるもの」を有しているのか、そしてそれは何なのかを突き止めねばならない。また、人間との類似性から人格性を基礎付ける場合、「人間に類似した人格」を生み出したり、場合によっては破壊する権利が人間に認められるのか、という疑問も生じる。そして、ドイツにおいて法人処罰の可否について議論がなされた際に、刑罰の正当化根拠についての見解の差異や、法人という人格の法における位置付けの理解の相違が結論を分けていたように、このアプローチを採用するには刑罰論や法哲学の問題にまで遡らねばならなくなる。したがって現状ではこの電子的人格性の根拠として、人間との類似性のような存在論的基礎付けを用いない方が無難であろうし、存在論的な基礎付けを不要と解することは、ロボットの処罰に反対する論拠となろう。

(4) 同概念についてはスザンネ・ベック「ロボット工学と法：その問題、現在の議論、第一の解決の糸口」富川雅満(訳)比較法雑誌50巻2号95頁以下。

5. 法的構造が有する社会への影響

以上、近時のロボット工学の発展に伴う法的対応の様々なバリエーションを見てきたが、法と社会は相互に影響しあう関係にあるため、これら法的対応は社会にも影響をあたえうる。すなわち、法における人格概念に変更を加えた場合、我々の人格観も影響を受け、これは我々のアイデンティティにも影響を与えうるものである。それゆえ、社会情勢の変化に応じて様々な法的対応をとることが必要にはなるが、波及効果なども考慮しつつこれをなすことが重要である。

Ⅲ. コメント

以上が本論文の概要である。本論文も前号で紹介した同筆者の論文と同様、刑法による問題解決能力には懐疑的なスタンスを取っている。つまり刑法はあくまで回顧的に犯罪を非難するのであって、展望的な視点を有しておらず、そのため新たなテクノロジーの発展やそれに伴うリスクを制御することには不向きである旨、指摘するのである。それゆえ、ロボットに人格性を付与し、答責対象たりうる者としてカウントしたとしても、刑法は謙抑的に用いられるべきであるとしている。

その上で本論文は「ロボットが人間に危害を加えた」という場合に、二つの問題解決アプローチを提示している。すなわち、予見可能性を不要と解することによって、過失概念を拡張する方法と、ロボット自体に人格性を付与することによって、ロボット自身に答責性を観念する方法である。以下ではその両アプローチについて若干検討してみたい。

1. 過失の修正について

ベック教授によれば、自律的ロボットの行為は最早予測が不可能であるから、具体的予見可能性は認められえないという。したがって過失の負責の際には、注意義務違反だけが要件とされる。当該ロボットについて注意義務を負う者が答責的であるとされ、その注意義務違反を立証された者のみが過失の罪責を負うこととなる。誰の注意義務違反も立証ができなかった場合には答責の間隙が生ずることとなろうが、刑法の謙抑

の性質に鑑みれば、むしろその間隙は社会全体で負担すべきものであるという。

これについて我が国の判例は、過失責任の要件として予見可能性を要求してはいるものの、具体的予見可能性までをも要求してはおらず、「因果関係の基本的部分」についての予見があれば足りるとしている⁽⁵⁾。したがって問題は、我が国の判例が何を「因果関係の基本的部分」として、予見可能性の対象としているのかである。

我が国の判例は、予見可能性の対象を相当程度広く捉えていると考えられる。たとえば、荷台に人が乗っていることを認識せずにトラックを運転し、速度超過により操縦不能となって信号柱に衝突し、荷台に乗っていた者を死亡させた事案において最高裁は、「無謀ともいうべき自動車運転をすれば人の死傷を伴ういかなる事故を惹起するかもしれないことは、当然認識しえたものというべきである」と述べて業務上過失致死罪の成立を肯定した⁽⁶⁾。また近時の判例では、以前から陥没が多発していた人工砂浜で、新たな別箇所で陥没が発生し、被害者がその陥没孔に落ち込んで死亡した事案において、市職員たる被告人の安全配慮措置を怠った過失が問題となった。これにつき判例は、過失責任を肯定するには予見可能性が必要であり、その予見の対象は「因果関係の基本的部分」であると述べた上で、当該人工砂浜の他の箇所で大規模な陥没が以前からあったことを認識していれば人の生命に対する危険は予見可能であり、結果回避措置を動機付けるには十分である旨述べて、業務上過失致死の成立を肯定した⁽⁷⁾。

上述のような判例の立場を前提とすると、予見可能性の存否は、いわ

(5) 札幌高裁昭和51年3月18日判決(昭49(う)219号、いわゆる北大電気メス事件)など。なお最高裁第一小法廷平成28年5月25日決定(平26(あ)1105号)の補足意見において、「『因果関係の基本的部分』は何かを検討する手法によるのがふさわしい類型とはいえない」事例も存在しうることが示唆されている。

(6) 最高裁第二小法廷平成元年3月14日決定(昭61(あ)193号)。

(7) 神戸地裁平成23年3月10日判決(平22(わ)25号、明石砂浜陥没事故差戻後第一審)。

ば「当該注意義務を怠れば、当該結果に至りうるのは必定」といえる場合に肯定されるのではないか。そして、注意義務は、具体的結果の予見ではなく、その義務履行を動機付けるに十分な危険性を予見していた場合に肯定される。そうだとすれば、予見可能性は結果回避義務と独立の過失犯成立要件なのではなく、結果回避措置や注意義務の内容を画定するために用いられているようにも考えられる⁽⁸⁾。

我が国の判例がもし予見可能性を注意義務内容の画定のための下位基準と考えているのであれば、ベック教授が述べるように予見可能性を不要と解することによって過失概念を修正するという必要性は、少なくとも我が国においては乏しいといえる。

2. 電子的人格概念について

本論文の特筆すべきもう一点のポイントは、「電子的人格性」の導入可能性を正面から扱ったという点であろう。またその基礎付けに際しても、存在論的な基礎付けは不要であるとした点も興味深い点である。

たしかに、人間との類似性を論拠として人格性を基礎付けてしまうと、結局刑罰論や法哲学にまで遡らねばならなくなり、終局的には「人間とは何か」といった終わりのない議論に陥ってしまうおそれもあるため、法理論的にはこれを避けるのが無難であることは否めないであろう。

しかしながら、この電子的人格性はあくまで法実務的な実用性という要請によってのみ基礎付けられるのであろうか。存在論的な基礎付けがないということは、同時に電子的人格性を付与する条件や限界が画定されていないということをも意味するのではなかろうか。本論文では自律的学習ロボットが想定されていたため、このロボットに人格性を付与することにあまり違和感はない。では一定の裁量を有するという意味で自律的ではあるが、学習・順応能力のないロボットの場合にも、人格性は付与されるのであろうか。

(8) 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。なお過失概念修正の可能性については本稿と同時に掲載予定の伊藤嘉亮による紹介論文も参照されたい。

また、電子的人格の基礎付けとの関連で本論文では、刑罰の正当化根拠の理解によっても、この電子的人格が刑法的答責の対象となりうるかの帰結に影響を与える旨の説明があった。しかし詳しい説明がなかったため、以下では試論となるが、刑罰の正当化根拠が電子的人格の処罰の可否にいかなる影響を及ぼすかを検討してみたい。以下の検討では、帰結の異同をクリアなものにするために、相対的応報刑論や予防的統合論などのような複合体としての理論ではなく、一般予防論、特別予防論、応報刑論という単一理論を前提としたい。

消極的一般予防論の場合、そこで前提とされる人格は快苦の枠組で意思決定をする合理的経済人に尽きる⁽⁹⁾から、「人格性」という語へ要求されるハードルは低く、電子的人格であってもこの要請は満たしうるように考えられる。また、ロボットに禁則事項として刑事法の規定をインプットし、あらかじめ犯罪的な行動が取れないようにしておけば、ある意味で完璧な威嚇にはなるはずである。ただし、ロボットに対する刑罰は、それを予告することによる威嚇に尽きる。というのも、刑罰規定を禁則事項としてロボットにインプットしている時点で威嚇は成功しており、実際に刑を執行することによる威嚇力はほぼ皆無に等しいからである⁽¹⁰⁾。ロボット場合は、刑の執行よりも、そのような過誤を起こすバグを取り除くアップデートをするほうが犯罪予防に資するであろう。したがって消極的一般予防論によれば、電子的人格も刑罰の対象となること自体は排除されないが、刑の執行に意味はないということになる。

これに対し、積極的一般予防論によれば、刑罰は単なる威嚇ではなく、制裁による規範確証のためになされる⁽¹¹⁾。しかし先述の通り、ロボットの場合、刑の執行による規範確証によるよりも、事前のインプットによる方が効率的であるから、積極的一般予防論によって電子的人格

(9) Michael Pawlik, Kritik der präventionstheoretischen Strafbegründungen, Festschrift für Hans Joachim Rudolphi zum 70. Geburtstag, S. 213ff. は「人格」とはこのような合理的経済人に汲み尽くされるものではないとする。中村悠人「刑罰目的論と刑罰の正当化根拠論」現代法学28号178頁以下も同旨。

(10) ロボットには「寿命」がないため、自由刑のような制裁を苦痛であるとは認識しない。それゆえ実際の刑の執行自体は動機付けの意味をなさない。

を刑法上の答責主体と見ることは困難であるように思われる。そもそも、ロボットが規範意識を持ちうるか、そしてそれを強化しうるかも疑問であろう。

特別予防論からすれば、比較的容易に電子の人格に答責性を見出しやすいように思われる。というのも、特別予防論は刑罰を治療、犯罪者を病人のように扱うためである⁽¹²⁾。ロボットが過誤行為を犯した場合には、たしかに当該ロボットはバグや不都合を抱えており、これはいわば病人のようなものであるから、電子の人格も刑罰の対象となりえよう。また、一般予防論は刑の執行の説明に難があったが、特別予防論からすれば、学習能力のあるロボットを訓練することは再社会化教育のようなものであるため、刑の執行も効果があるといえるであろう。無論、そのような訓練を刑罰として執行することが効果的であるかは別途検討を要する。

これに対して応報刑論は、ロボット処罰の正当化に困難を伴う。カントの同害報復応報刑論の場合、人格は理性的主体であることが要求される。そこでは自由意思が前提とされ、道徳的価値判断が可能であることが必要となる⁽¹³⁾。人工物たるロボットに自由意思が存在するか、単なるアルゴリズム以上の道徳律を学習することが可能かという問題が大きな障壁となろう⁽¹⁴⁾。また、たとえ人格性の問題をクリアしたとして、刑の量定をする際、同害をどのように算定するのかという問題も生じる。すなわち、ロボットには寿命や死の概念が存在しないため、自由刑や死刑すらも彼らにとっては害悪ではないからである。それゆえロボットの過

(11) Wolfgang Frisch, Über das Verhältnis von Straftatsystem und Strafzumessung — Unrecht und Schuld in der Verbrechenslehre und in der Strafzumessung, GA 2014, S. 489 ff. 高橋則夫「刑法総論(第二版)」13頁、林幹人「刑法総論(第二版)」15頁など。

(12) Franz von Liszt, Der Zweckgedanke im Strafrecht, S. 29 ff.

(13) イマヌエル・カント「人倫の形而上学」『カント全集11』178頁以下。

(14) ロボットも自由意思を有しており、主体性を有しうるとするのは、松村格「システム論と自由意思—ヨンバルト論文『刑法と自由意思』への釈明—」『法の理論2』219頁以下。

誤行為がもたらした害悪に相当する害悪を付与することは困難であろう。ただし他の国民の応報感情を慰撫するという面を強調すれば、他の国民が納得する量刑が「同害」であるとみなすことはできよう。

ヘーゲルの応報刑論の場合、やはりカントの応報刑論と同じく人格性の問題を生じる。また規範違反に対する応報の場合、そもそもロボットの行為が規範に違反したといえるかが問題となろう。ロボットは規範を理解して内面化させているのではなく、外部的行為の準則として規範をインプットされているにすぎないため、規範に違反しているとは言い難いのではないか。さらに、カントの応報刑論と同様、規範違反に比例する量刑も問題となろう。

以上のように、予防論によれば電子的人格に答責や刑罰を観念することは可能であるように思われる。逆に応報刑論の場合には、非常に大きな困難を伴う。この差異は、一方で予防論が人格性や意思自由などに対して要求する内容が少なく、他方で応報刑論が人格性を内容豊かに考えているという点に生じる。したがって、応報刑論による場合は、電子的人格が高度に人間と類似しているということを前提とした場合にのみ、処罰が可能となろう。「存在論的基礎付けを不要と解することは、ロボットの処罰に反対する論拠となろう」というベック教授の説明は、このことを意味したものであると思われる。