

Gunkel, *The Machine Question*, chapter 1 要旨

久木田水生

2012 年 12 月 28 日

1 Moral Agency

1.1 Introduction

キューブリックの『2001 年宇宙の旅』では、宇宙船を制御する人工知能 HAL が不具合を起こして、HAL の機能を停止しようとする乗組員を殺害する。このような場合、HAL はこの行動に対して法的・道徳的責任があるのだろうか？ すなわち HAL は道徳的行為者になることが可能なのだろうか？

フィクションだけでなく現実においても、自動ロボットによって人間が殺された事件が起こる一方で、ますます人間に近い能力を持つロボットが開発されている。現状においてはこういったロボットの行動の liability は設計者やユーザーに帰される。しかし liability の何らかの側面が機械それ自体に帰されることが可能だとしたら、moral responsibility もまた機械に移されるだろう。

問題は、どの時点で、そしてどのような根拠で、機械がその行動に対する責任を持つと考えることが形而上学的に可能、あるいは道徳的に responsible になるのだろうか、ということである。

1.2 Agency

この問題に取り組むためには「道徳的行為者性」という言葉の意味を定義する（あるは少なくとも特徴づける）必要がある。

道徳的行為者は行為者の部分集合である。そこでまず行為者について考えよう。Himma (2009) の「標準的見解」を参照すると、行為者性は、act あるいは action として見なしうる何らかのことは行うことができることという考えと結びついている。さらに行為者性にはある種の「意図」が行為の背後にある必要がある。さらに道徳的行為者であるためには、何らかの道徳的な criteria あるいは stipulation に従っている必要がある。Himma (2009) によれば道徳的行為者とは、自由な選択をする能力、何をすべきかを熟考する能力、そして道徳的規則を理解し典型的な事例においてそれを正しく適用する能力、そして意図的に行動を行うことができるものである。

この定義にはいくつかの欠陥がある。第一にこの定義は「自由意志」、「熟考」、「意図」等の複雑で難しい概念を含む。第二に、この複雑さのために、この定義は普遍的なものではなく、他の多くのありうる定義の一つに過ぎない。そして道徳的行為者性についての一つの定義を選択するということは、それ自体が道徳的帰結を持つ行為、道徳的に中立ではない行為である。

1.3 The Mechanism of Exclusion

Heidegger の「道具的定義」によればテクノロジーの役割あるいは機能とは、使用者である人間が特定の目的のために採用する手段であるということである。McLuhan の *Understanding Media: The Extensions of Man* によればすべてのテクノロジーは人間の能力の拡張であり、メディアである。

このように考えるとロボットや AI、その他のコンピューターシステムも responsible agent とは見なされない。人間だけが道徳的行為者である。道具を責めることは存在論的に間違ったことであるばかりか、倫理的にも怪しい。それは人間が自らの行為の責任逃れをするために使う方法の一つだからである。

道具的定義の欠点の一つは、それが人間中心的な理論であるということである。しかし人間の境界は常に確定していたわけではない。また人間中心主義は排他的であり暴力的である。

またテクノロジーのすべてが単なる「道具」とみなせるわけでもない。自律的な「機械」は道具と区別されるべきだ。Langdon Winner は *Autonomous Technology* において、自律的な機械とは人間の意志と独立に行うするものと定義している。だとすれば自律的な機械は道具主義的な定義には当てはまらないテクノロジーである。だとすれば自律的機械の到来は道徳的行為者性の理解に重要な変化を惹き起こすだろう。

AI やロボットが人間の複雑さは精妙さを備え、感情的知性を備えるようになったならば、そのような人工物を単なる道具として扱うことはひょっとしたら非道徳的なことになるかもしれない。もし機械が人間のような合理性を持つことになれば、Kant によればそれは「目的の王国」の一員でなければならない。

そのようなテクノロジーはまだ実現しそうにないため、道具主義者たちはその可能性を考慮していない。しかし現在でも人間とは独立に行動する機械のプロトタイプは作られており、これは行為者性の帰属の問題を複雑にしている。例えば自動的に学習するシステムなどはすでに作られている。Matthias (2004) が言うには、そのようなシステムの引き起こしたことの結果を、従来のように人間に帰すことは誤りである。Mowbray (2002) によれば MMO (Massive Multiplayer Online) では、自動的に活動するボットが登場人物として社会の中に存在しており、従来の行為者＝道徳的行為者という対応が失われ、理論的かつ実践的な困難を引き起こしている。このようなボットは誰の責任でもなく被害を引き起こしうる。というのもプログラムと、それと相互作用する人間の行動や心的感情状態、他のボットの行動、環境とそのコミュニティの社会経済などの組み合わせを含む、様々な状況が組み合わさってそのような被害が起こるからである。

道具主義的定義によればこのような状況は、人間が機械に対する制御を失った状態として理解される。ここにはテクノロジーによる人工物は人間によって制御されている、かつそうであるべきだ、という前提がある。しかしこの前提は疑わしい。

1.4 The Mechanism of Inclusion

Singer は種の境界を形成する生物学的事実には道徳的重要性 significance がないという。このために道徳的行為者性の問題は「パーソン性 personhood」に言及し、それに依存してきた。

パーソンの概念は普遍的ではないし、常に安定していたわけでもない。またパーソンは人間と一対一には対応しない。胎児や脳死患者はパーソンを持たないし、法人、知性を持つように思える異星人、高等哺乳類はパーソンを持つ。

パーソンの範疇と人間の範疇を区別することは、様々な抑圧と排除を促進し正当化したが、しかしそれは皮肉にも人間以外の動物や人工物のような他者を、正当な道徳的主体として考慮することを可能にしてきた。

オーストリアではチンパンジーのマッシューに法的後見人が認められことを求めた裁判が行われたことがある。そこではマッシューがオーストリアの法に従って一パーソンとして考えられるべきだということが主張された。この裁判では結局、マッシューがパーソンであるか否かという問題は棚上げされた。

機械のパーソンが請願されたことはないが、しかしその可能性については多くの議論がなされている。David J. Calverley は機械が人間の心的状態に比較しうる特徴を示すように設計されるようになれば、法がそのような存在者をどのように扱うかということは、設計者および社会全体にとって非常に重要な問題になるだろう、と述べている。また Peter Asaro は、法人間以外の存在者にパーソンのものである法的権利を認めている事例を指摘している。すなわち法人の例である。アメリカの法律によれば、「議会の定めるいかなる法の意味を決定する際にも、文脈がそうでないことを明示しているものでない限りは、「パーソン」あるいは「誰であれ」という言葉は個人のほかに、法人、株式会社、協会、企業、パートナーシップ、団体、合資会社を含む」。しかしここでもパーソンという言葉が定義されているわけではない。

1.4.1 Personal Properties

「パーソン」という言葉は重要であるが、しかし厳密な定義を与えられていない。一方で何かがパーソンを持つか否かと言うことは、パーソンを形作る重要な性質を持っているかいないかによって決定できると主張する人々もいる。そのようなプロセスはまず、パーソンを構成する性質がどのようなものであるかを明確化しなければならない。そしてさらにある対象がその規準を見なすか否かを証明しなければならない。

西洋哲学は伝統的に、パーソンに意識（自己意識）と知性（合理性）を結び付けてきた。そして機械や動物はそれらを持たないという理由でパーソンを持たないものとされてきた。しかしこのような考えは様々な理由から反論を受けている。Tom Regan は動物に意識があるということを5つの根拠からなる「累積的論証」によって主張している。その根拠とは私たちの常識、言語実践、人間と動物の厳格な二分法に基づく排他主義に対する批判、動物の行動、そして進化論である。

Regan や Searl は人間だけが意識を持つものではなく、高等な哺乳類は意識を持つと主張する。しかしこれは Descartes などによって引かれた排他主義の線引きを少し広げただけに過ぎない。

一方で機械は動物よりも意識を持つ見込みが高いと思われる。人間と同等の知性と意識的行動を機械において実現することは AI 研究の当初からの目標であったし、多くの批判にも関わらず AI 研究者たちはその目標が達成可能なものと考えている。Cf. Owen Holland, “Editorial introduction” in Holland ed. *Machine Consciousness*.

ここでの主要な問題は、動物や機械が意識を持つか否かということよりも、「意識」という言葉が曖昧ではっきりとした意味を持たないということである。

1.4.2 Turing Tests and Other Demonstrations

意識のようなパーソンを構成する性質を定義することの難しさは問題の一面でしかない。そういった性質が特定の対象に存在しているかを認識することの難しさもあるのである。

Turing は言語を使ったコミュニケーションにおいて人間と区別できない機械は知性を持つとみなすべきだと主張した。David Levy は意識の問題も Turing と同じようにアプローチされるべきだと主張する。

このアプローチは実際には Descartes によっても試みられたものである。Descartes によれば動物や機械が人間ではないと認識できる確実な規準があると言う。それは言語と理性（これらはともに古代ギリシャ語ではロゴスと呼ばれた）の欠如である。人間であればどれほど愚かなものであろうと話しかけられた状況に応じて異なる適切な語の配列を発することで答えることができる。しかし動物と機械にはこれができない。動物と機

械はロゴスを欠いているという点で人間と区別される。そしてまたこの点において動物と機械は同等である。つまり動物の行動を模倣する機械と動物を厳密に区別することはできない。

しかしこのようなアプローチは現在ではもはや有効ではない。1967年、Joseph Weizenbaum のおしゃべりロボット ELIZA は、質問されたことに対して人間のユーザーと異なる配列の語を発することで会話をした。仮想世界やインターネットに配置されたより洗練されたボットとのコミュニケーションの経験によって、機械と人間の間の境界はますます曖昧になっている。そして同じような発見は動物に関してなされている。霊長類を使った様々な実験が、動物にも洗練された言語能力があることを確証している。

理性と合理性に関しては状況はますます複雑である。*Ratio* という語はそもそも計算を意味するものだった。そして Leibniz がやろうとしていたのは「すべての理性の真実がある種の計算に還元される一般的な方法を作る」ことだった。現在のコンピュータは数学的演算や複雑な定理の証明において人間に勝るのみならず、人間の言語を翻訳し、チェスのグランドチャンピオンを負かし、ジャズの即興演奏をする。Brooks が言うように、理性はもはや人間を定義するバリエーションではない。「今日、ほとんどの人は、適切なソフトウェアと適切な問題領域が与えられたならば、コンピュータは推論し、意思決定し、目標を持つと言うだろう」(Brooks 2002, 170)。

さらには人間ではなく機械のみが合理的行為者になりうると論じる理論家もいる。Nedau は、人間の行為が理性によって引き起こされていることはごくまれ、という。一方で機械は完全で間違いのない論理的処理によって判断を行うようプログラム可能である。従って Nedau は機械のみが完全に合理的でありうるという。そして合理性が道徳的意思決定の基本的な要件であるとすれば、機械だけが正当な道徳的行為者だと考えられるのである。

しかしこの議論の本当の問題は、動物や機械がパーソンを構成する性質を持つか持たないかを、何らかの議論、証明、テストによって示すことではない。本当の問題は Dennet や Derrida が指摘するように、ある外的に観察可能な現象から内的な働きに関する想定へと飛躍するという、根拠のない推論 unfounded inference なのである。これは議論のどちらの側も行っている。Kant が指摘するように、見えているものからものそれ自体について推論をすることはできない。Dennett が結論付けるように、内的生活を持っているように思える何かが本当にそれを持っているということを証明することはできない。哲学者、心理学者、神経科学者がどれだけ議論を重ねようと、「他者の心」問題は、経験科学のアプローチによって解決できるものではない。

1.5 Personal Problems and Alternatives

確かなことは“person”あるいは“person-making qualities”といったものが、曖昧なものだということである。しかしながら DeGrazia のように、それらが曖昧であるからこそ、人間以外の者にそれらの概念が開かれているのだと考える論者もいる。一方で、そのことは何が道徳的な道徳的ステータスを認められるのかという重要な決定を恣意的で、不整合で、総体的なものにするという可能性もある。

さらに問題を複雑にしているのは、そのような決定が利益の絡んだ人々 (interested party) によって行われているということだ。Johnson and Miller (2008) によれば、コンピューターシステムが「道徳的行為者」になれるかどうかについての判断は、その決定から何らかの利益や損を得る人々が行っている。また Lucas Introna (2003) がいうには、倫理的重要性を決めるのは人間であり、人間は自分たちの能力や経験に基づいて、その決定の基準を導き出している。結果として、「パーソン」の概念は人間が例外であるという立場を異なる名前で導入するに過ぎない。

1.5.1 Rethinking Moral Agency

Deborah Johnson (2006) によれば、道徳的行為者性をめぐる議論の枠組みは、「コンピューターが道徳的行為者であるか、それともそれが道徳的ではないかのどちらか」という選択を迫るようになっている。Johnson は視点を変えて、議論の用語を再構成することを提案する。「コンピューターが倫理的行為者であることを否定することはコンピューターが道徳的な重要性を持たないということと同じではない。またコンピューターが道徳的であると主張することはそれが道徳的行為者だと主張することと同じではない。

Johnson によれば、道徳的行為者性をコンピューターには与えず、一方でそれが倫理的な重要性を持つことを認識することは可能である。Johnson は、コンピューターが道徳的行為者になることはできないが、しかしそれは志向的でありうる、と言う。それはその設計者の志向的行為によってコンピューターに組み込まれたものである。人間が人工物を用いて行為を行う時、その行為は人間自体の志向性と効果 (efficacy) および、それが用いる人工物の志向性と効果によって構成されている。そして人工物の志向性と効果はさらにその設計者の志向性と効果から構成されている。従ってユーザーと設計者と人工物のすべてが道徳的評価の焦点になる。

しかしこの志向性の三つ組は、通常の道具主義よりも複雑ではあるが、しかし人間例外主義の枠を超えるものではない。Johnson のパラダイムにおいてもやはり人間だけが正当な道徳的行為者である。

別の選択肢は道徳的行為者の基準を下げることである。J. P. Sullins (2006) によれば、道徳的行為者性がパーソン性と結び付けられる限り、ロボットが道徳的行為者と認められる可能性は低い。しかし Sullins は道徳的行為者が必ずしもパーソンである必要はない、と主張する。彼の議論はロボットの道徳的行為者性についての 4 つの哲学的見解を概観することから始まる。それは以下のとおりである。

- 現在のロボットは道徳的行為者ではない。しかし将来そうなる可能性はある。(Dennett 1998)
- コンピューターとロボットは決して道徳的行為者にはならない。(Selmer Bringsjord 2008)
- 人間は道徳的行為者ではないが、ロボットは道徳的行為者になりうる。というのも道徳的行為は十分に考えられた理由に基づかなければならないが、人間は完全な合理的存在者ではないから。(Joseph Emile Nedau)
- 「精神なき道徳性」を認めればロボットが道徳的行為者になりうることを認められる。(Sullins 2006)

Sullins は「精神なき道徳性」は、自由意志や志向性などの心の哲学における困難を回避することで、道徳理論における様々なパラドクスから免れることができる、と考える。Sullins は道徳的行為者の基準として次の三つだけを挙げる。

- 工学的な意味における自律性。すなわちロボットが他の行為者あるいはユーザーの直接的に制御されていないこと。
- 弱い意味での志向性。これは Dennett が論文「志向的システム」において発展させた概念で、ある対象に信念や欲求を帰属させることによって、その対象の行動が説明・予測できるならば、それは弱い志向性を持つとされる。
- 責任。これも他者の心の問題を避けて、単なる見かけだけで十分とみなす。

Sullins の議論は Floridi and Sanders (2004) に基づいている。Floridi and Sanders (2004) は道徳哲学の主要な問題は、その分野が人間中心的な行為者性の概念によって不当な制約を受けていることだ、と述べる。このために道徳哲学は近年の「分散された道徳性」のような革新に付いていけない。道徳的行為者性概念の完全な改訂が必要である。

その改訂は Floridi and Sanders (2004) が「抽象の方法」と呼ぶものによって進められる。これは何らかのシステムとその文脈を記述・分析・議論する際に、異なるレベルの規準を定式化するものである。Floridi たちは道徳的行為者としての抽象化のレベルは、相互作用性 interactivity、自律性 autonomy、適応可能性 adaptability の三つを含むものであるとする。この抽象化のレベルでは、人間（子供を含む）だけでなく、ソフトウェア・エージェント、組織や企業、様々な動物も行為者として認められる。しかしこれだけではまだ道徳的行為者として認められるのに十分ではない。ある行為が道徳的行為として認められるのは、それが道徳的善または悪を引き起こしうる時である。ある行為者が道徳的行為者として認められるのは、それが道徳的行為をなしうる時である。この特徴づけにおいて重要なのは、それが現象学的なものであり、心理学的なものではないということである。これはある種の「精神なき道徳」である。

Floridi たちの議論は人間以外のものに道徳的行為者性を認める可能性を持つものであるが、しかしそれには少なくとも 3 つの問題点がある。

1. 適切な抽象化のレベルが何かということに関して議論の余地がある。
2. 誰が適切な抽象化のレベルを決めるのかということに関しては政治倫理的な次元がある。
3. このアプローチは道徳的行為者性の定義を与えるものではなく、操作的な限界を与えるのみである。これは根本的な哲学的問題に対して「工学的な解決」を与える試みである。

1.5.2 Functional Morality

道徳的行為者性にまつわる相当の形而上学的・認識論的・倫理的困難に対処する一つの方法は、大きな問題をすっかり回避することである。Wendell Wallach (2008) が「機能的道徳性」と呼ぶ戦略においては、機械の行為者性は解決できない問題かもしれないとされるが、しかし現実世界においてますます自律性を高める機械による意思決定の帰結については考慮する必要がある、とされる。機械の道徳性についての哲学的な議論はさておき、機械が破滅的な帰結をもたらすかもしれないということを認識し、私たちは私たちが制御できることに取り組まなければならない。(Cf. Susan and Michael Anderson (2007a))

機能主義のアプローチには、Kant の批判哲学の試みと不思議な類似性がある。カントは実際的に重要な責任の問題に取り組むために、大きな形而上学的・認識論的・メタ倫理的な問題を解決することを必要としなかった。ただし機能主義は機械の行為者性・パーソン性・意識などを否定するわけでも受け入れるわけでもない。これらの問題そのものは私たちの能力の限界を超えている可能性を認識する、というカントの批判哲学のスタンスを取るだけである。

アシモフのロボット工学の三原則もまた機能主義的アプローチの例である。それらはロボットが意識やパーソン性を持つことを要求してはいない。

機能主義の現実的な応用としては Anderson, Anderson and Armen (2004) によって提唱された機械倫理 (machine ethics; MA) がある。コンピューター倫理が主にコンピューターを道具として用いる人間の行動の帰結に関心を持つのに対して、機械倫理は機械の行動の帰結に関心を持つ。このようにして、機械倫理は人間中心の伝統に異議を唱え、倫理の主体を広げることを主張する。そしてこの目的のために機械倫理は完全に機能主義的なアプローチを取る。Susan Leigh Anderson (2008) によれば、ME は機械自体の倫理的ステータスについて私たちが判断を下すことを要求しない。ME の関心は

- 倫理的意思決定が計算可能であるのか
- 行動のための適切な倫理的基準を機械にプログラムすることができるのか

という点にある。

倫理的意思決定の計算可能性は、行為功利主義や義務論などの伝統的な道德哲学が追求してきたことに合致する。これらは倫理的意思決定の合理的機械化を目指してきたのである。Henry Sidgwick (1981) によれば「倫理学の目的は、振る舞いの正しさないし合理性に関してほとんどの人間が持っている見かけ上の認識を**体系化し**、そこから誤りを取り除くことである (the aim of Ethics is to *systemize* and free from error the apparent cognition that most men have of the rightness or reasonableness of conduct)」。

従って西洋において道德は通常、行動についての体系的な規則にその本質があるものと考えられてきた。そしてその規則は多くの状況において様々な行為者によってアルゴリズムのようにコード化され、実装される。Cf. 十戒やカントの定言命法など義務倫理。ベンサム「道徳算術」のような功利主義。

道德が計算可能かという問題に際して、Susan and Michael Anderson (2007a) は「プログラムすることが論証である」というハッカーの格言に従って、いくつかの作業的プロトタイプを作成する。そのような試みが最終的に「倫理的な知的行為者」を生み出すかどうかはまだ分からない。しかし彼らは機械に明示的に倫理的な要素を組み込むことが可能であることを、「限定された領域に概念を応用するという証明」によって明らかにした。さらに Anderson たちはこのことがより良い道徳理論の発展に結び付くだろうと主張している。

同様の機能主義的アプローチは Wallach and Allen (2009) にも見られる。彼らは道德的行為者性にまつわる困難で深遠な哲学的問題の重要性を認識しつつ、そのことは機能的な道德的能力を備えた「人工道德的行為者 (artificial moral agent; AMA)」の開発を妨げるものではない、と言う。彼らは機能的 AMAs の設計における三つのアプローチの費用対効果分析を行う。そのアプローチとは以下のものである。

- トップダウン：あらかじめ特定された倫理理論を取り上げ、その計算的要件を分析して、その理論を実装することのできるアルゴリズムとサブシステムの設計の指針とする。アシモフのロボット工学三原則に基づくロボットや Selmer Bringsjord の仕事などはトップダウンアプローチの例である。しかし Wallach と Allen によれば、このアプローチは極めて限定された状況でしか使えず、AMA にそれが従う曖昧でない規則を与えることは実現できないだろう。
- ボトムアップ：試行錯誤のプロセスによって道德的行為を導出する。Peter Danielson の「バーチャルゲームにおけるバーチャルロボット」、およびジレンマ反応規範進化プロジェクト (Norms Evolution in Response to Dilemmas; NERD)。このアプローチは倫理の社会的本性に注目する」という点で独自のメリットがある。しかし現実世界に应用するには困難がある。
- ハイブリッド：トップダウンとボトムアップの両方のアプローチを利用する。Anderson と Anderson の MedEthEx は。あらかじめ定義された *prima facie* 義務と学習アルゴリズムの両方を採用している。

機能主義のアプローチは有望な結果を生み出しているが、しかし少なくとも三つの重大な困難がある。

- 機能主義は道德性を試験する方法を必要とする。Allen, Varner and Zinser (2000) はある種の道德的チューリングテストを提案している。そこでは被験者は AMA に道德に関する質問をしてそれが人間か機械かを判定することを求められる。しかしこのテストは推論能力 (discursive abilities) に不当な重点を置いている、と批判される。道德性のテストには、状況の理解も必要だと Stahl (2004) は批判する。
- 一般に機能主義のアプローチは、機械による意思決定がもたらしうる危機を管理したいという欲求によって動機付けられている。このような動機は明確に人間中心的である。この人間中心主義は Wallach と Allen のような機能主義者に関しては特に問題ではないかもしれない。しかし Anderson たちは、人

間中心的伝統一般に対する明確な異議申し立てとして、特にコンピュータ倫理の構造的限界に対するオルタナティブとして、MEを導入し推進している。従って彼らにとってはこの人間中心主義は、その目的に反するものである。その一方で機能主義は奴隷の倫理とでも呼ばれうるものを創始している。Bryson (2010) は明確にロボットは奴隷であるべきことを主張する。しかし Robdney Brooks (2002) のように、機械が知性や意識や志向性の兆しを見せ始めたならば、そのとき機械の well-being に対して私たちは責任を負うことになる、と考える人々もいる。

- Terry Winograd (1990) は、解釈を経た判断 (interpretative judgments) なしに規則が従われる「心の官僚主義」に対する警戒を呼び掛ける。「人が自分の仕事を規則の正しい適用であるとみなす時、個人的な責任とコミットメントが失われる」と Winograd は言う。また Mark Coeckelbergh (2010) は、規則に従うだけで、価値や道徳を理解しない「サイコパスロボット」を生み出す危険性を指摘する。ただし Winograd と Coeckelbergh は、機能主義的なロボットに精神がないと断じている時点で機能主義の重要な前提の一つを無視している。機能主義はロボットに心があるかどうかに関しては私たちは知りえないという前提を採っているからである。しかしそのことは問題を解消するわけではない。機能的に設計されたシステムが本当に実質的な道徳的行為者なのか、プログラムに従うだけの人工的官僚なのか、それとも見かけだけ道徳的で、じつは潜在的な危険性をはらむサイコパスなのかどうかを知らないままに取り残されるならば、問題はなお悪い。

2 Summary

機械の問題は道徳的行為者性について尋ねることから始まった。この選択は道徳哲学の歴史によってあらかじめ定められたものである。ギリシャの古代から現代まで、道徳哲学はほとんどもっぱら行為者指向的に行われてきた。

行為者の視点から考察されるとき、道徳哲学は必然的に、道徳的行為者のコミュニティに含められるべきは誰で、考慮から除外してもいいのは何か、という排他的な決定を下す。しかしこの「誰」という語のもとにおかれるべきものが実際に誰なのかという問いに完全に決着がついていたことはかつて一度もない。道徳哲学の歴史的発展は、かつては除外されていたものが、道徳的行為者のコミュニティに入ることを認められていく、前進的展開 (a progressive unfolding) として解釈できる。

この前進にも関わらず、機械はいまだにこのコミュニティに入っていないし、その候補者として考慮されたこともない。機械は人間の特定の目的のための単なる道具として見なされ続けている。このテクノロジーの道具主義的理解は驚くべき水準で受容され標準化されている。機械の道徳的ステータスに関するこの決定は、自律的システムとロボットに適用した時、特に興味深い形を取る。Bryson (2010) などによって、「ロボットは奴隷であるべきだ」と主張されるのである。このとき道具主義的理論は、人間例外主義と人間の権威を創始し確保するために、とりわけ有用であることが分かる。

しかしながら道具主義と人間中心的な伝統の下部構造とその道徳的帰結に関しては、その見直しを迫るかなりのプレッシャーがある。これは少なくともハイデggerの批判的な介入に始まり、そして動物の権利運動と近年の AI とロボット工学の進歩を通じて持続してきたものである。しかし以前は除外されてきたこれらの他者へと考慮を拡張することは、道徳的な「パーソン性」の概念に取り組み直すことを要求する。そしてこれは容易なことではない。とりわけ、あるものをパーソンたらしめるのが何かについてはほとんど、あるいはまったく同意がないのである。

一つの「パーソンたらしめ」性質として、私たちは意識について考察をした。しかしこれもまた存在論的、認識論的な困難に突き当たった。何が「意識」であるのかについて私たちはまったく理解していないように思われるのである。これを回避する試みは、機能主義などを含め、どれもうまくいっていない。

かくして機械の道徳的行為者性についての問題に答えることは、まったく簡単でもなければ決定的でもない。これはそもそも「道徳的行為者」という語が曖昧で、不確定で、そして雑音の多い概念だという事実によるものである。機械の道徳的行為者性についての問題を吟味することによって明らかになったのは、道徳的行為者性の概念が、まったくもって混乱したものであり、私たちが実際に道徳的行為者であるかどうか不明でない、ということだ。他者の道徳的行為者性は、それらについての道徳的考慮に先だてて見出されるものではない。それはそれらとの相互作用のプロセスにおいて、それらとの関係性において、それらに与えられるものである。しかしこの問題はもはや行為者性の問題ではなく、**受容者性** *patency* の問題である。