

ロボット倫理はどこに向かうのか

根 村 直 美

[Abstract]

This paper defines a robot as an artifact that can communicate with human beings using language, gestures, or computer graphics. Such machines are already in use in numerous domains, and robots are poised to permeate aspects of everyday lives in human communities.

This paper also explores the argument on contemporary philosophy and ethics including critical posthumanism. This examination reveals that the concepts of *moral agent* and *moral patient*, human societies have assumed, are not grounded in the existence of a certain substance and that they are constructed in the concrete situation through the interaction of all entities containing both non-organisms and organisms.

Furthermore, two questions are asked herein: “Is a robot a moral agent?” and “Is a robot a moral patient?” It is concluded that the answers to both inquiries would depend on how a robot relates to its environment, which includes humans and their societies.

Thus, this paper demonstrates that a robot might be considered a moral agent or patient because it is a machine that is becoming increasingly adept at communicating with human beings.

I. はじめに

応用倫理学の文脈では、2000年代後半にすでに、ロボット倫理に関する議論の必要性が意識されはじめていた。とはいうものの、そうした議論はいささかSF的な思考実験と捉えられてきた。

近年、限定的ではあるが自律性をもつロボットが社会の様々な領域で活躍するようになっている。加えて、シンギュラリティ（技術的特異点）という概念が現実味をもって受けとめられるようになってきている。本稿は、こうした状況を背景として進展しつつあるロボット倫理の議論の現状を明らかにするとともに、今後どのように考えていけばよいのかを哲学・倫理学の立場から探ることを目的とする。

II. ロボットの定義

『ロボット白書 2014—社会を変えようとするとき、そこにロボット技術がある！—』は、ロボットについて完全に一般性をもった定義というのは実は存在しないとしつつ、ロボットの学術的な定義として、日本ロボット学会が編纂した『ロボット工学ハンドブック』（1990, 2005）の定義をいくつか紹介している（平井・水川・林原, 2014: 1）。1967年に開催された第1回ロボットシンポジウムでは、森政弘と合田周平により「移動性, 個性性, 知能性, 汎用性, 半機械半人間性, 自動性, 奴隷性の七つの特性をもつ柔らかい機械」という定義が提唱された。また、人間型ロボット研究開発の草分けであった加藤一郎は「①脳と手と足の3要素をもつ個体, ②遠隔受容, 接触受容器をもつ, ③平衡覚, 固有覚をもつ, これらの3条件を備える機械」としている。

スタンフォード大学の B. Roth の定義も紹介されている。Roth は「ロボットとは人間や他の動物あるいは機械と連携して仕事をする機械であって、自動型と半自動型がある。ロボットとは他の自動機械との区別はあまりはっきりしていなくて、かなり気まぐれで商業主義的なところがあり、また時間的にも意味は移り変わっている」と述べているという（平井・水川・林原, 2014: 1）。

一方で、実用的な「産業用ロボット」については、日本工業規格（JIS）の「産業用マニピュレーションロボット - 用語」（JIS B 0134-1998）の用語の 1100 番に「自動制御によるマニピュレーション機能または移動機能をもち、各種の作業をプログラムによって実行できる、産業に使用される機械」と規定されている（平井・水川・林原, 2014: 1）。

『ロボット白書 2014』によれば、ロボットの定義は「時と場合による」のであり、その定義が流動的な理由は、「そもそも時代とともに科学技術はもとより、産業構造, 社会制度, 文化なども変化し、ロボットの役割や受け取られ方も変遷してきている」（平井・水川・林原, 2014: 1-2）からである。

ところで、2018年の『情報処理』第59巻8号で、近年のAIブームの中で誕生している製品が人間とのコミュニケーションを重視し、家庭や社会など暮らしの中で活躍する製品であることを背景としつつ、「コミュニケーションをとる機械」としてのロボットに焦点を当てた特集が組まれた（真部・細野・今井, 2018: 684）。『ロボット白書 2014』が言うようにロボットの定義は多種多様であるが、本研究がグローバル研究センターのプロジェクト研究「ロボット倫理はどこへ向かうのか」の一環であり、そのプロジェクト研究が倫理学と情報学の立場からロボットを考えるものであることに鑑み、本稿では、『情報処理』の特集と同様のアプローチを取りたいと思う。すなわち、本稿では、ロボットを、必ずしも「機械の体」をもつものと捉えず、コンピュータグラフィックなども含め、「自律的に判断して動き、言葉や時にはジェスチャーを用いて人とコミュニケーションできる人工物全般」（今井, 2018: 692）を指すこととしたい。

Ⅲ. ロボット倫理の基礎理論—哲学・倫理学の立場から—

哲学者・倫理学者がロボット倫理を考える際に射程となる問いには、大きく言って2つの方向がある。1つには、ロボットが人間生活に組み込まれるようになりロボットの人間生活への影響が大きくなるにつれて、ロボットの振る舞いに一定の倫理的制約を課す必要が出てくるであろう。かくして、ロボットに課されるべき倫理的制約を考える、つまり、「ロボットが守るべき倫理」（久木田，2009）とは何かが問いとして設定されることになる。もう1つ、ロボットと生活の中で深く関わるようになるにつれ、人間はロボットに対してどのような配慮をすべきなのかを問うことができるであろう。すなわち、「ロボットに対する倫理」（久木田，2009）とはどのようなものなのかを問題とすることができるのである。

「ロボットの守るべき倫理」を問題にするためには、ロボットが「道徳的行為者」になりうることを前提とする必要がある。そこで、いわゆる AMA（Artificial Moral Agent）の概念が議論となってくるのである。その一方で、「ロボットに対する倫理」が問題とされるためには、ロボットが道徳的な配慮を受けなければならない対象である「道徳的受容者」、「道徳的被行為者」となりうるものが前提となってくる。

1. 道徳的行為者をめぐって

1-1 機械倫理の試み

機械倫理の主導者の1人 Susan L. Anderson がそうであるように（Anderson，2012），人工的な道徳的行為者をつくる試みは、これまでのところ「人間が従っている倫理的規則を明らかにし、それを機械にプログラムして実行させる」というパラダイムに従って進められている（久木田，2017：12）。

Anderson らが開発しているシステムは、何らかの規則ベースの推論システムを前提としているが、あらかじめすべての規則が与えられているわけではない。様々な事例において専門家が下す判断から、機械がその背後にある規則を学習するようになっており、そうした学習をして得た規則に基づき新しい状況においても適切に判断できるようになることを目指している（久木田，2017：12）。

このような発想の根底にあるのは、道徳とは理性に基づくものであり、行為者が正しい一般的基準と自分の置かれた状況を適切に理解しており、かつ、十分な思考能力を有しているならば、その行為者は適切に道徳的判断を下すことができる、という「合理主義」（久木田，2017：10）と呼ばれる思考である。この思考形式は西洋の倫理学においては1つの主流をなしてきており、機械倫理は西洋の倫理学の伝統を受け継ぐ試みと言えよう。

1-2 人工的道徳行為者は本当に「道徳的」と言えるのか

一方で、哲学・倫理学者の間では、ロボットが道徳的規則に従った判断を下し、その規則に従った振る舞いをしたとして、果たして、そのロボットを「道徳的」と評価しうるのかということが問われてき

ている。

この問題は、人工知能をめぐる「思考をしているかのように振る舞う人工知能は本当に思考をしていると言えるのか？」(久木田, 2017: 17), 「人工知能は心をもつのか？」といった問題と関連させて考えることができるものである。

Alan Turing は、人間の思考のプロセスは観察できないのであるから、外部から観察できる入力と出力の対応だけを思考能力のテストとしようと考えた。そのテストとは次のようなものである(人工知能学会, 2017)。2台のディスプレイの前にテストをする人がいる。1台のディスプレイには隠れている別の人が、もう1台は人間をまねるようにつくられたコンピュータが受け答えした結果がそれぞれ出てくる。テストをする人はどんな質問をしてもよいとする。詩を作らせたり、音楽の感想を尋ねたりする。また、コンピュータも人間をまねる努力をする。例えば、わざと計算に時間をかけたり、間違えたりする。こうして、テストをする人が、どちらが人間でどちらがコンピュータか分からなければ、このコンピュータには思考能力、あるいは、知性や心があるとするのがチューリングテストである。

これに反論したのが哲学者 John Searle である。Searle は「中国語の部屋」という次のような思考実験を考え出した(人工知能学会, 2017)。英語しか分からない人が部屋にいる。その部屋には、中国語が分からなくても、中国語の文字を書いてある通りに置き換えると、中国語の受け答えができてしまう完璧な説明書がある。つまり、この部屋の人、英語しか分からないが、中国語の質問に中国語で答えることができる。ということは、中国語の受け答えができるだけでは、中国語が分かるとは限らないことになる。同様に、まるで思考能力、あるいは、知性や心があるような受け答えができていてというだけでは、本当に思考能力、あるいは、知性や心があるかは分からないという反論である。

道徳的な行為についても同様の議論が可能であろう。道徳的問題に対して対応表を参照しながら答えを出している人がいるとする。外部から見ると、その人は道徳的判断をしているように見える。しかし、その人は道徳的な判断能力をもっているとは限らないのではないだろうか。

確かに、哲学・倫理学では、ある人の判断が道徳的かどうかを問うとき、その人がどう行動するのかを問うばかりではなく、どのような意図をもってその行動がなされたのかを問題にしている。例えば、Immanuel Kant は実践理性の無条件に善い行為をせよと命じる命法に従うことを道徳的な振る舞いとし、義務に従おうとする善意を重視した(Kant, 1994 [1785])。

こうした、意図や動機など道徳的行為者の〈心的な態度〉といったものを問題にする立場からすれば、機械倫理の試みに対して、「人工的道徳行為者は、本当の意味では、“道徳的”とは言えない」という批判が向けられることになるのである。

2. Floridi の道徳的行為者

情報倫理学分野では、人工的道徳行為者の意図や動機が問題にならないように道徳的行為者の概念を再構成しようとする動きがある。その1人が、イタリアの哲学者であるが、オックスフォード大学のスタッフとして主として英米圏で活躍する Luciano Floridi である。

Floridi は、「情報倫理」への統合的アプローチを提唱し、マクロ倫理としての情報倫理の中に道徳的行為者を位置付ける（Floridi, 2010 : 52-55）。

Floridi は、情報存在物がつくる情報環境を「情報圏（infosphere）」と呼ぶ。Floridi の言う「情報存在物（informants）」とは、情報的に理解されるすべての存在物であり、「一群のデータつまり情報オブジェクト」として記述することができるものである。

Floridi は、人工的行為者を道徳的行為者として認めることの問題を認識している。道徳性はふつう、責任（responsibility）を含意する。それゆえ、「しばしば、人工的行為者は自らの行為に責任を取ることができないため道徳的行為者となりえない」と議論されることになる。しかしながら、Floridi は、規範言説の中には、責任の所在を問う問題とは独立のものがいくらかでもあるとして、搜索救助犬の例を挙げる。搜索救助犬は行方不明者の後を探索するよう訓練されている。搜索救助犬はしばしば人命救助に貢献し、飼い主や救助した人から褒められ、報酬を受ける。だがこれは肝心な点ではない。人間は動物に感謝の念をもちうるが、犬にとってはゲームなのであり、その行為について道徳的な責任があると考えすることはできない。大切なのは、犬が道徳と関係のあるゲームに主要なプレイヤーとして参加したことであり、それゆえに、それらが起こした事柄の善悪について結果責任を負いうる（accountable）⁽¹⁾ 道徳的行為者として認められるのは正当であるという点なのである。

Floridi によれば、道徳的責任を問えなくとも、道徳的行為者は存在しうる。すなわち、道徳に関連する行為について結果責任（accountability）と能力のみがあり道徳的責任を問えない場合にも、規範的行為をするように促すことができるのである。Floridi は、道徳的責任には意図や自覚、その他種々の心的態度が必要であるが、道徳的責任と道徳に関連する「結果責任」とは区別できると考える。そして、「心をもたない道徳」というアプローチを探究するのである。

こうした Floridi の人工的道德行為者をめぐる議論は、「自由意志に基づいて行動する理性ある人間」（西垣、2007）を、道徳的責任を負う道徳的行為者とみなす西洋の倫理学の伝統を問い直すものとなっている。そして、そうした倫理学の伝統を問い直す動きは、現代哲学のうちに表れた動向とも呼応するものである。例えば、「心の哲学」の領域においては、「行動主義」といった立場が出現しており（ウォーバートン、2010 : 216-217）、Turing のように、観察できる入力と出力のみに注目しようとする考え方がうまれてきているのである。

3. クリティカルなポストヒューマニズムの立場からの考察

上述のような、「心をもたない道徳」、あるいは、外部から観察される道徳的な振る舞いのみを問題にしようとするアプローチとまではいかなくとも、人間の自律や主体性という概念の見直しは、現代哲学・倫理学の重要な課題の1つとなっていることは確かである。

<自律的な主体>の存在を前提とする西洋の<啓蒙主義的ヒューマニズム>の見直しの端緒を開いた1人が、Michel Foucault である（Foucault, 1966）。Foucault のそうした考えは、人間の認識過程を知の産出の中心に置くことを拒否した。すなわち、Foucault が注意を喚起しているのは、人間主体が意味を

発展させる言説的な構造である。主体は、言説に先んじて存在するのではない。言説によって確立された様式の中のみ存在する。すべての主体は、言説の中で形成されるのである。

啓蒙主義以来、「自分の理性に従って考え判断し行動することができる自律的な主体」としての自己という観念こそが、ヒューマニズムを支えてきた。〈啓蒙主義的ヒューマニズム〉は、自己が自由な意志をもち自律的な主体であることに道徳性の根源を見出し、その自由な意志を尊重することこそがその人の尊重となると考えてきたのである。

しかしながら、Foucault は、個人を、自律し自己を制御し自身の意志をもつ主体ではなく、言説の効果と捉えることで、〈啓蒙主義的ヒューマニズム〉の基盤を掘り崩した。そして、近年、クリティカルなポストヒューマニズムを展開する論者たちは、このような Foucault のヒューマニズム批判を受けた〈自律的主体としての人間〉概念の再構成を試みている。

例えば、Pramod Nayar のクリティカルなポストヒューマニズムは、自律的な主体に代わり、統一的でない主体像を提示しようとする (Nayar, 2014 : 8)。すなわち、そのポストヒューマニズムは、機械と有機的な身体、人間と他の生の形式が、多かれ少なかれ、境界が分からないようにつなぎ合わされており、相互に依存し、ともに進化する仕方に注意を払う。哲学的アプローチとしての Nayar のクリティカルなポストヒューマニズムは、人間の主体性を、共進化する機械と人間の寄せ集めと見るため、主体という観念の捉え直しを迫るのである。

Nayar はまた次のようにも言う (Nayar, 2014 : 10)。クリティカルなポストヒューマニズムは、人間の自律を、環境へ反応するシステムの自己制御として見ている。つまり、人間の内的な複雑さを、自己充足的な自律の印ではなく、むしろ、自身を制御し、組み込まれた環境から自分自身とその作用を切り離そうとするシステムの試みと見る。そして、そのポストヒューマニズムは、人間の主体を環境の中で生きられた経験によってうみだされるものと見るとともに、生きられた経験を相互的な関係性を通じ主体によって形成されるものと見るのである。

こうしたクリティカルなポストヒューマニズムは、実体的な道徳的行為者たる「自律的な自己」の存在を否定していると見ることができる。その一方で、〈自己〉やその〈心的な態度〉が言説システムを通じてうみだされることを否定してはいない。クリティカルなポストヒューマニズムは、道徳的行為者を社会的な相互作用、とくに〈言語〉を媒介とする相互作用の結果としてうみだされるものと見るのである。そうした立場に拠って立つならば、「自律的に判断して動き、言葉や時にはジェスチャーを用いて人とコミュニケーションできる人工物全般」(今井, 2018 : 692) としてのロボットに対し、道徳的行為者ではないと断言してしまうことはできないであろう。

IV. 道徳的受容者をめぐって

さて一方、「ロボットに対する倫理」が問題とされるためには、先述したように、そもそもロボットが道徳的な配慮を受けなければならない道徳的受容者、道徳的被行為者でありうるものが前提となる。

続いて、この問題について考察してみたい。

1. Floridi の道徳的受容者

まず前節で取り上げた Floridi が道徳的受容者についてどのように論じているか、その議論を見てみよう。Floridi は、情報的存在物はすべて固有の道徳的価値をもち、何らかの尊重に値すると論じる (Floridi, 2010 : 56-57)。

そして、「最小限の道徳的尊重」とは、「利害によらず認められ配慮される」ということである (Floridi, 2010 : 56)⁽²⁾。Floridi によれば、非生物的存在である情報的存在も道徳的配慮を受けるべき存在であり、道徳的受容者にほかならない。Floridi の情報倫理は、「存在物中心的、受容者指向的で、エコロジカルなマクロ倫理」なのである (Floridi, 2010 : 47)。

Floridi はまた、情報的存在、あるいは、情報圏一般にとって何が善なのかという、情報倫理によってなされる基本的な道徳的問いについて、次のように論じる (Floridi, 2010 : 59)。いかなる情報的存在物も何らかの基本的な倫理的要求の中心と認められる。そういった要求は承認に値するものであり、それと関連するあらゆる情報のプロセスの実行を規制するのに資するべきものである。したがって、道徳的行為者 A の決定や行動を受け入れるか否かは、A の行動が情報圏の善き在り方にかなる影響をおよぼすかに基づくべきだということになる。換言すると、情報圏を豊かにするか貧しくするかにかかっているのである。道徳的行為者の義務とは、情報圏全体の持続可能な繁栄に対する貢献という点から評価されるべきである。そしてまた、情報圏全体に対して負の影響を与えるプロセスや行為や出来事が悪の事例と見なされるのである。

2. 倫理学における道徳的受容者

続いて、道徳的受容者に関して、倫理学ではどのように考えられてきたかについてまとめる。

近代以降の西洋倫理学においては、いわゆる「自律的な人間」というのが道徳的受容者と考えられてきた。この考えは、Kant の「君は、みずからの人格と他のすべての人格のうち存する人間性を、いつでも、同時に目的として使用しなければならず、いかなる場合にもたんに手段として使用してはならない」(Kant, 1994 [1785] : 52) という議論と結びつけられてきた。Kant の言う「目的」として扱うということは、その相手の意志を尊重することである。Kant の考えに従えば、自由な意思を備えた相手として扱うということが、その人を尊重すること、つまり、その人の「尊厳」を守ることの基盤になっている。

自由な意思を備えた存在であることを道徳的受容者の要件とする考えは、道徳的行為者と道徳的受容者を表裏一体に捉える考え方である。この考えは、近代以降の我々の社会において一定の役割を果たしてきた。また、現在でもそうであると言ってよいであろう。しかしながら、20 世紀後半以降の応用倫理学の分野では、我々が配慮すべき存在の範囲はより広く考えられるようになっている。

例えば、生命倫理学においては、厳密な意味では「自律的な人間」とは言いえない人間存在に対して

どのように振る舞うべきかがしばしば重要なテーマとなる。意思決定能力を失った患者や十分な意思決定能力をもつにいたっていない子どもの治療停止に関する議論はその1つの例である。人間社会には、倫理的な関係性を通じて、道徳的行為者とは言えないが、道徳的受容者として位置付けられている存在があるのである。

生命倫理学の分野では、この道徳的受容者には、＜個体として生存しうる人間＞とすることはできない「胎児」、および、「ヒト受精胚」も含まれるようになっている。日本におけるそうした存在の位置付けを例に取って見てみよう。まず、出生前の胎児については、墮胎罪の規定によって、出生後の人と同程度ではないが、刑法上の保護の対象となっている。また、総合科学技術会議は、2004年に「ヒト胚の取扱いに関する基本的考え方」を取りまとめ、「ヒト受精胚」を次のように位置付けた（総合科学技術会議、2004：5）。すなわち、「これまでの社会実態を踏まえて定められた我々の社会規範の中核である現行法体系は、ヒト受精胚を『人』として扱っていない。ヒト受精胚を『人』として扱う考え方を採用することは、この現行法体系を大幅に変更し、受精胚を損なうことを殺人と同義に位置付けることを意味するが、人工妊娠中絶手術が行なわれ、また生殖補助医療において余剰胚等の一部の受精胚を廃棄せざるを得ない現在の社会実態を踏まえれば、そのような制度変更は現実的とは考えられない。また、そのような制度変更について社会的合意を得る見通しもない」とする一方で、ヒト受精胚は「人」へと成長し得る「人の生命の萌芽」として位置付けられ、「通常のヒトの組織、細胞とは異なり、特に尊重されるべき存在」とした。ヒト受精胚は人そのものではないが、人間の尊厳という社会の基本的価値を維持するために特別に尊重されるべき存在とされたのである。

こうした総合科学技術会議の位置付けの背景にあるのは、2003年の生命倫理専門調査会の中間報告書で示された考えである。そこでは、「胚は人ではないとしても『人の生命の萌芽』としての尊重を軽視すれば、それが成長した存在である人に『人の尊厳』があるとする我々の理念の維持は難しくなる」（総合科学技術会議、2003：1）と論じている。

さらに、生命倫理以外の応用倫理学の分野では、道徳的受容者の範囲は「人間」を超えて広がっている（神崎、2017：92-94）。

例えば、Peter Singer は、少なくとも人間以外の動物の一部は道徳的受容者であり、それらの動物の利害は人間と等しく配慮されるべきと主張した（Singer, 2002 [1975]）。

功利主義者である Singer にとって、道徳的に重要なのは快楽と苦痛である。それゆえ、道徳的に配慮される道徳的受容者であるための条件は、快楽や苦痛を感じる能力をもつかどうかである。そして、少なくとも一部の動物は、快楽や苦痛を感じる能力をもっており、その意味で利害をもっている。かくして、それらの動物は人間と等しくその利害を配慮されるべきなのである。Singer は、生物学的に人間に属している個体のみを道徳的な配慮の対象とし、その以外の動物の個体を道徳的な配慮の対象に含めないのは、人種差別と同じく非難されるべき態度であるとし、そのような態度を「種差別」と呼んだ（Singer, 2002 [1975]：6）

一方、環境倫理学もまた、人間だけを道徳受容者であるとするこへ異議を唱えてきている。例えば、

Aldo Leopold の「土地倫理」は、生態系を道徳的な配慮の対象とした (Leopold, 1968 [1949])。

Leopold の言う「土地」とは、ある空間や地域にすむ動物や植物だけではなく、土や岩や水などの無機物なども含めてすべてのものによって構成される「共同体」としての生態系のことである。Leopold はこの意味での土地こそが最も重要な道徳的受容者であるとする立場を表明している。そして、「物事は、生物共同体の全体性、安定性、美観を保つものであれば妥当だし、そうでない場合は間違っているのだ」(Leopold, 1968 [1949] : 224-225) と論じている。

3. クリティカルなポストヒューマニズムからの考察

次に、筆者がこれまで注目してきたクリティカルなポストヒューマニズムの立場から、道徳的受容者について考えてみたい。

例えば、Nayar のクリティカルなポストヒューマニズムによれば、人間の主体や身体は、置かれた環境において機械や他の生命との相互的な関係性によってうみだされるものであるが、その一方でまた機械や他の生命との相互作用は形成された主体や身体によってうみだされていく (Nayar, 2014)。つまり、主体と身体は機械や他の生命と複雑に関わり合うことにより「生きられた経験」を通じてうみだされるとともに、「生きられた経験」は機械や他の生命と相互に作用し合う主体と身体によって作りだされるのである。

Nayar のクリティカルなポストヒューマニズム (Nayar, 2014) は、人間の主体と身体を上述のように捉え直すことにより、「啓蒙主義的なヒューマニズム」の「人間」の概念は排他的なものであり、その概念は、ある種のグループに属する人、機械、動物などその概念から排除されてきた者を下位におき抑圧してきたと見なす。

こうした考えに基づき、Nayar は、生物、非生物いずれもを包含するアプローチとしての「Species Cosmopolitanism」という概念を提示する (Nayar, 2014 : 150-156)。Nayar によれば、「Species Cosmopolitanism」は、ジェンダーであれ (男性/女性)、生の形式であれ (動物/人間、人間/植物)、構成要素であれ (有機体/非有機体)、安易な二項対立を拒絶しているため、クリティカルなポストヒューマニズムの究極の形である。クリティカルなポストヒューマニズムは、あらゆる生の形式を含むマルチな市民権を提唱している。

Nayar などクリティカルなポストヒューマニズムの論者は、サイエンス・フィクションの分析を通じてその思考を鍛え上げてきている。そうした動向を受け、本稿筆者は、アニメーション映画『イノセンス』(2004年、監督:押井守、制作:プロダクション I.G) にもそうした「クリティカルなポストヒューマニズム」的な思考が認められることを明らかにしてきた (根村, 2016)。

『イノセンス』の、バトーと草薙、そして、助け出された少女の次の会話を見てみよう。

バトー「犠牲者がでることは考えなかったのか。人間のことじゃねえ・・・魂を吹き込まれた人形がどうなるかは考えなかったのか」

少女「だ・・・だって・・・だって、私は人形になりたくなかったんだもの」

草薙「鳥の血に悲しめど魚の血に悲しまず、声あるものは幸いなり、人形たちにも声があれば、人間になりたくなかったと叫んだでしょうね」

バトーの言葉や草薙の言葉に見て取れるのは、〈啓蒙主義的なヒューマニズム〉が提示する〈人間〉の概念とは異質な存在を、そのままの形で敬おうとする〈敬意〉とも呼ぶべきものである。『イノセンス』における人間以外の存在への配慮は、同質なものへの共感ではなく、異質なものへの〈敬意〉に基づくものと考えられるのである。

『イノセンス』に見られる〈人間性〉を中心とした価値付けを回避しようとする姿勢、その姿勢に基づく異質な他者への〈敬意〉は、Nayar の言うクリティカルなポストヒューマニズムとしての「倫理的プロジェクト」(Nayar, 2014 : 30-31) にほかならないと言えよう。

4. 範囲は定められるのか

これまでの考察で、道徳的受容者の範囲と道徳的行為者の範囲は必ずしも一致するものと考えられてはこなかったことは明らかであろう。したがって、ロボットを道徳的行為者と見ることができるかどうかという問題について一定の見解に達することができていなくとも、ロボットを道徳的受容者として見ることができるかという問題について考えることは十分に可能である。

では、道徳的受容者に関してどう考えるべきなのか。神崎宣次は、W. H. Hunt が「単なるモノは道徳的に考慮されるか」という論文 (Hunt, 1980) でこの問いに対する究極的な回答を与えたと論じている (神崎, 2017 : 96-97)。Hunt は、Singer が「生物学的に人間である」ことを基準とする立場を種差別としたことに対して、「苦痛を感じる能力」も恣意的に選ばれた基準にすぎず、本当に恣意的ではないのは「生きている」ことなのだと主張した Kenneth Goodpaster を批判する。すなわち、Hunt は、その Goodpaster の主張もまた同様に恣意的なものに過ぎず、「存在する」を基準にしてもかまわない、と指摘する。つまり、これまで提案されてきた様々な基準はそれぞれの支持者が恣意的に選んだ条件を主張しているという点で大差はないと言う。そして、特定の条件を道徳的に配慮すべき範囲の境界線を引く基準とするには哲学的正当化が必要だということを主張するのである。

しかしながら、神崎は、そのような哲学的な正当化によって唯一の基準を示すことができるかについて懐疑的である。少なくとも、現状の議論を見る限り、唯一の基準を示すことに成功した倫理学者はいないとする。そして、異なる角度から問題を見ることを提案する。すなわち、「理論的に考えれば道徳的被行為者の範囲を定める条件はこうであるから、これは道徳的被行為者に含まれるというのが話の順序なのではなく、ある特定の対象を道徳的被行為者に含めざるをえなくするような倫理的問題が登場した状況において、その状況の改善につながる条件が基準として主張されるのだと」(神崎, 2017 : 97)。

こうした見方は、上述のような主張をうんできた応用倫理学分野の成り立ちを考えるならば頷けるものである (神崎, 2017 : 97-98)。すなわち、人間による環境破壊が深刻になった時期に環境倫理学が登

場し、生態系は道徳的受容者であると主張されるようになった。また、動物が苦痛に満ちた扱いをされている状況を改善しようとする意識が高まり、動物解放の主張がなされるようになった。つまり、道徳的受容者の範囲をめぐる問いは実践的関心に基づいて提起されてきているのである。

そう考えるとすれば、ロボットを道徳的受容者と見なすべきかどうかという問題は、ロボットを道徳的受容者として考えることが要請されるような状況がどれだけ認めうるかということに依存することになるであろう。神崎は、将来においてロボットが道徳的受容者としてみなされる可能性がないことを意味しないとしつつも、現時点ではそのような状況にはないと論じている（神崎、2017：98）。

しかしながら、ロボットは、「自律的に判断して動き、言葉や時にはジェスチャーを用いて人とコミュニケーションできる人工物全般」（今井、2018：692）と定義されるようになっている状況を考えるならば、現時点でも、すでにロボットを道徳的受容者と見なすことが要請される状況が生じつつあると言えるのではないだろうか。

V. ロボット倫理はどこに向かうのか

1. ロボット倫理に向けての第一歩

ここまで考察をまとめるならば、そもそも、人間社会が想定してきた「道徳的行為者」「道徳的受容者」という概念は、先だって存在している何らかの<実体>に拠って立つものではなく、<有機体も非有機体も含む諸々の存在が相互に作用し合う具体的状況において構築されてきたもの>と考えることができるであろう。そして、そう考えることができるならば、ロボットを道徳的行為者として見なすべきかという問いにせよ、また、道徳的受容者として見なすべきかという問いにせよ、その答えは、ロボットが人間社会を含む周囲の環境とどのように関わっているかに依存すると考えることになるであろう。

その意味では、ロボット倫理を考えるための第一歩は、今現在、ロボットが人間社会を含む周囲の環境とどのように関わっているかをできるだけ具体的に把握することだと言えよう。そして、この第一歩を踏み出した際に目を向けざるをえないのは、やはり、ロボットを「自律的に判断して動き、言葉や時にはジェスチャーを用いて人とコミュニケーションできる人工物全般」（今井、2018：692）と定義する動きが出てきていることである。

これまでも述べてきたように、人間とのコミュニケーションを主眼とするロボットの捉え方に拠って立つならば、ロボットを道徳的行為者と想定する可能性が開かれることになる。しかしながら、啓蒙主義的な人権思想に基づいて諸制度が設計されている社会においては、一定の条件を満たした人間のみを道徳的行為者と見なしているというのが現状であろう。

それに対し、啓蒙主義的な人権思想に基づく社会においても、人間とコミュニケーションができる動物を配慮すべき存在と見る見方が浸透しつつあり、道徳的受容者に関しては、その範囲が広がりはじめていると見ることができる。とするならば、人間とのコミュニケーションを主眼とするロボットの捉え方は、すでに、既存の人間以外の道徳的受容者と照らし合わせつつ、ロボットを道徳的に位置付けるこ

とを要請ははじめていると考えられるのである。

このように見てくると、Frolidi が、人間と機械が融合するシステムがますます増大する情報社会における倫理の出発点として、道徳的受容者に直目したこと（Frolidi：2010）は、妥当性のあるアプローチと言えるであろう。

2. 無自覚の人間中心主義の回避

さて、ロボット倫理をより具体化していくに際して、今一度向き合わなければならないのは道徳的受容者の範囲を定めようとする議論が含みもつ問題点である。これまで道徳的受容者の範囲を定めるための唯一の基準を提示することに成功した者はいないことについては先に論じた通りであるが、そもそも、そのような唯一の基準を設定しようとする考えることには大きな問題が内在するのではないだろうか。

西垣通もこの問題を取り上げる（西垣，2007：18-23）。西垣は、「情報圏においては、さまざまな存在物に対応する情報オブジェクトを具体的にいかに定めるかが決定的に重要である」が、実際には「人間の分類した概念構造が用いられている」にもかかわらず、この点に関して、Frolidi の「反省的な洞察はかならずしも十分ではないように思われる」と論じている（西垣，2007：21）。そして、Frolidi の情報倫理は、「理性ある人間はとくに選ばれ、神からの管理者としての使命を与えられている。したがって、生物も含めて万物のあり方を正しく保護管理し、調和ある繁栄をもたらさなければならない」と考える西洋流エコロジー思想をデジタル情報圏に展開したものにほかならないとしている（西垣，2007：20）。つまり、人間が唯一の基準を設定しようとする考えることは、人間を「神から使命を与えられた特別な代理者」の地位に置くものであり、人間中心主義を超克しきれてはいないとするのである（西垣，2007：23）。

これまでの考察では、倫理学者らの議論の現状を受けて唯一の基準を提示することに成功していないと論じてきた。しかしながら、西垣の言うように、無自覚の人間中心主義という陥穽に陥ってしまうことを避けるためには、特定の研究者が唯一不変の基準を設定しようとする考えること自体を避けねばならないと言うべきであろう。そして、道徳的受容者の範囲に関する問いが一義的な解答に行きつくことがないことを前提とするならば、その問いは常に問われ続けるものとならざるをえないであろう。このことはまた、どういう存在を道徳的行為者として見なすべきかという問いに関しても同様である。特定の研究者が唯一不変の基準を設定しようとする考えることは避けねばならない。その答えが一義的な解答に行きつくことはないがゆえ、その問いは問われ続けなければならないのである。

ところで、無自覚の人間中心主義を避けるためとして、西垣は伝統的・人文主義的な西洋哲学とは異なる視点から倫理を構想しようとする（西垣，2007：25-30）。そこで提示されるのが、人間が生物の一員だという自然科学的観点をより徹底させるオートポイエーシス理論である。ここで大きな疑問がうまれてくる。このオートポイエーシス理論に基づいて、生物と機械を峻別できるとすることは、この理論に拠って立つ者を無自覚に神の視座に置くことになっているのではないだろうか。そして、ある範囲の設定をすることにより神の視座に立つことに対して無自覚なことはまさに西垣自身が批判しようとして

いることではないのか。

西垣の議論を見ていると、オートポイエーシス理論を無意識の人間中心主義から免れているものと考えるのは、それが「自然科学的観点」に基づいたものであるからのように思われる。

しかしながら、科学論者の中には、何らかの自然科学言説が〈ありのままの事実〉を映し出している、すなわち、〈真理〉を言い当てていると考えることに異を唱え、自然科学も、時代や地域といった社会・文化背景から切り離すことはできないという認識を提示する者が少なからずいる（Haraway, 1999, 2000；Hacking, 2006；Gergen, 2009, 2015）。

このような自然科学に対する考え方は、社会構成主義と言われる立場に属している。その立場は、Keneth J. Gergen を援用するならば、次のように論じることができるであろう。すなわち、その考え方は、何かを命名し記述する自然科学者の言葉には存在する何かを写し出す特権があり、それゆえ、自然科学の言葉が私たちに唯一の現実を教えてくれると考えることに対して危惧を示し（Gergen, 2015：231-232）、自然科学が「それ以外のあらゆる声を沈黙させる独裁者」（Gergen, 2015：22）とならないようにと警鐘を鳴らす。そして、自然科学者が自らの説明を〈ありのままの事実〉、〈真理〉と考えることにより、その説明を反省的に問い直すことができなくなっていることを批判するのである。

本稿では、先に実践的関心がどのようなところにあるのかを明らかにすることを第一歩とすると述べたが、無自覚の人間中心主義を回避するためにも、特定の研究者が唯一不変の絶対的な境界線を設定することを避け、常に変化する現実の実践的関心を探ることに基づいて暫定的な境界線を設定するというアプローチを取るのがロボット倫理の構築にはふさわしいと言えるであろう。

このようなアプローチは、結局は人間がその関心に基づいて境界設定をしているという意味では、「人間中心主義」を完全に超克しているとは言えないかもしれない。しかしながら、このアプローチは人間の実践的関心を探ることによって基準を設定せざるをえないということを深く自覚するアプローチである。それは、人間のうみだす知はどれも絶対的基準を設定しえないと認識するがゆえに、人間を特別な地位に置くことが出来ないことを前提としたアプローチであるという意味では「人間中心主義」は回避しえていると言えるであろう。

3. 実践的関心をいかにして探るか

それでは、実践的関心を探るには、どのような方法があるのだろうか。その1つとして挙げられるのは、本稿筆者がこれまでサイバー・カルチャー研究において取ってきた現象学的方法である。この方法は、フッサール現象学の、普遍的なものである本質を把握する方法としての現象学的方法ではなく、経験的事実的構造を明らかにしようとする方法である（根村, 2017）。〈本質〉を取りだそうという態度は現象を何らかの世界の現れとして捉えることであり、現象をその現象たらしめている起源を想定することである。しかしながら、経験的事実的構造を明らかにしようとする現象学的方法は、そうした起源も基礎も想定しない。現象を何らかの世界の現れとしてではなく、現れそのものとして捉えようとする試みは、「現象学のラディカル化」、あるいは、「現象学に対するラディカルな批判」と見なされても

いる (Colebrook, 2002 : 6).

また、現象学的方法は、多くの場合「解釈学的方法」を要求する (リー, 2009 : 29-30). 経験は、その研究に携わる者が自己反省を通して直接的に把握しうるものと、内省を通じて直接には把握できないものの2種類に分けられるからである。第1の種類の経験の場合は、内省という方法により研究に携わる者に直接的に把握されるため、解釈学的方法が要請されるということはない。一方、第2の種類の経験は、研究に携わる者が直接には把握できない他者の経験や自身の経験であっても過去のものとなっている経験などがそれに当たる。第2の種類の経験は直接的に内省によって把握されるものではなく、解釈によって間接的にのみ把握されるのであり、解釈学的方法が必要となる。この場合、解釈学的方法は現象学的方法にとって不可欠な要素となる。解釈学のアプローチが現象学的研究を遂行するために必要とされるのである。

この方法に関連して、知能研究の新しい方法として「一人称の研究」が提案されていることにも触れておきたい。その方法は、「数少ない被験者の個別具体的状況に（一人称的な $N = 1$ の事象に）面白い現象を見出し、それを基に知の面白い側面について仮説を立て、その仮説を検索クエリーにして、同じ現象が他者にも成りたちはしないかと探す」（諏訪, 2015 : vi）ことによって普遍を探究しようとするものである。

人工知能研究の新たな潮流である「一人称研究」という方法と本稿筆者が提示する方法とは相通じるところが多いと言えよう。本稿が提示する方法も、ごく個別的経験を基に、ロボットと人間のコミュニケーションのうちにどのような倫理が立ち現れつつあるのかを探り、そこから仮説を立てていくことになるであろう。この仮説は唯一で普遍的な基準となりえないものである。本稿筆者の方法論においては、それは状況依存的なものにならざるをえず、どのような倫理が立ち現れているのかという問いは常に問われ続けなければならないものとなるであろう。諏訪正樹は、「一人称研究」においては「個から普遍をみる」とは、普遍的な知見にたどり着くということそのものではなく、個々の研究者が自分なりの仮説を立て普遍的知見を指向する探究のプロセスを意味するのかもしれないと論じているが（諏訪, 2015 : 41-42）、本稿筆者の方法においてもまた、仮説そのものにおいてではなく、その探究のプロセスにこそ普遍を見ることができるとも思われる。

VI. 結び

本稿では、〈人間とコミュニケーションをとる機械〉としてのロボットが社会や生活の中に入り込みつつある状況を背景に、ロボットを、「自律的に判断して動き、言葉や時にはジェスチャーを用いて人とコミュニケーションできる人工物全般」（今井, 2018 : 692）と定義し、議論をはじめた。

一方、本稿では、現代哲学・倫理学、とくに、クリティカルなポストヒューマニズムの議論を参照しつつ、人間社会が想定してきた「道徳的行為者」「道徳的受容者」という概念は、何らかの〈実体〉の存在に拠って立つものではなく、〈有機体も非有機体も含む諸々の存在が相互に作用し合う具体的状況

において構築されてきたもの>と考えるに至った。そして、そうした考えに基づき、ロボットを「道徳的行為者」として見なすべきかという問い、「道徳的受容者」として見なすべきかという問いのいずれも、その答えは、ロボットが人間社会を含む周囲の環境とどのように関わっているかに依っていると結論付けた。

こうした哲学・倫理学の立場から、あらためてロボットが<人間とコミュニケーションをとる機械>としての存在感を高めている状況を考慮し、すでにロボットを道徳的受容者と見なすことが要請される状況にあると考えるのが本稿の立場である。人間はロボットに対してどのような配慮をすべきなのかを、経験の構造を探る現象学的方法を用いて探究するのは次の課題としたい。

注

- (1) 西垣通は、accountable を「説明できる（説明義務がある）」と訳しているが、本稿では「結果責任がある」という訳を当てている。accountability も同様に「結果責任」という訳を当てた。
- (2) 元々は R. Hepburn の議論である（Hepburn, 1984）。

引用・参考文献一覧

- Anderson, Suzan Leigh (2012) "Machine Metaethics." In Michael Anderson and Suzan Leigh Anderson (eds.), *Machine Ethics*. Cambridge: Cambridge University Press, pp.21-27.
〔邦訳: クレア・コールブルック（國分功一郎訳）『ジル・ドゥルーズ』青土社、2006年〕
- Flolidi, Luciano (2010) "Information Ethics: Its nature and Scope." In Jeroen van den Hoven and John Weckert (eds.), *Information Technology and Moral Philosophy*. Cambridge: Cambridge University Press, pp.40-65.
〔邦訳: ルチアーノ・フロリディ（西垣通訳）「情報倫理の本質と範囲」西垣通・竹之内禎編著訳『情報倫理の思想』NTT出版、2007年、pp.47-98〕
- Foucault, Michel (1966) *Les mots et les choses. Une archeologie des sciences humaines*. Paris: Gallimard.
〔邦訳: ミシェル・フーコー（渡辺一民・佐々木明訳）『言葉と物—人文科学の考古学—』新潮社、1974年〕
- Gergen, Kenneth J. (1999) *An Invitation to Social Construction*. London: Sage.
〔邦訳: ケネス・ガーゲン（東村知子訳）『あなたへの社会構成主義』ナカニシヤ出版、2004年〕
- (2009) *An Invitation to Social Construction*. Second edition. Los Angeles: Sage.
- (2015) *An Invitation to Social Construction*. Third edition. Los Angeles: Sage.
- Hacking, Ian (1999) *The Social Construction of What?* Cambridge, Massachusetts: Harvard University Press.
〔邦訳: イアン・ハッキング（出口康夫・久米暁訳）『何が社会的に構成されるのか』岩波書店、2006年〕
- Haraway, Donna, J. (1991) *Simians, Cyborgs, and Women: The Reinvention of Nature*. New York: Routledge.
〔邦訳: ダナ・ハラウェイ（高橋さきの訳）『猿と女とサイボーグ—自然の再発明—』青土社、2000年〕
- (2000) *How Like a Leaf: An Interview With Thyrza Nichols Goodeve*. New York: Routledge.
〔邦訳: ダナ・ハラウェイ+シルザ・ニコルズ・グッドイヴ（高橋透・北村有紀子訳）『サイボーグ・ダイアローグズ』水声社、2007年〕
- Hepburn, R. (1984) *Wonder and the Other Essays: Eight Students in Aesthetics and Neighbouring Fields*. Edinburgh: Edinburgh University Press.
- 平井成興・水川真・林原靖男 (2014) 「第1章 ロボットについて」『ロボット白書 2014—社会を変えようとするとき、そこにロボット技術がある!—』新エネルギー・産業技術総合開発機構。
- Hunt, W. H. (1980) "Are Mere Things Morally Considerable?" *Environmental Ethics* 2.1, pp.59-65.

- 今井倫太 (2018) 「なぜロボットを使うの？」『情報処理』(情報処理学会), 第 59 巻 8 号, pp.692-697.
- 人工知能学会 (2017) 「チューリングテストと中国語の部屋」
<https://www.ai-gakkai.or.jp/whatsai/AITopics3.html>, 最終アクセス日: 2019 年 3 月 27 日.
- Kant, Immanuel (1994) *Grundlegung Zur Metaphysik Der Sitten* [1785]. Hamburg: Meiner.
〔邦訳: イマヌエル・カント (中山元訳) 『道徳形而上学の基礎づけ (1785)』光文社 (光文社古典新訳文庫), 2012 年〕.
- 神崎宣次 (2017) 「この映画の撮影で虐待されたロボットはいません—道徳的被行為者について考える—」久木田水生・神崎宣次・佐々木拓『ロボットからの倫理学入門』名古屋大学出版会, pp.83-101.
- 久木田水生 (2009) 「ロボット倫理学の可能性」『Prospectus (京都大学文学部哲学研究室紀要)』第 11 巻, 京都大学, <http://hdl.handle.net/2433/71114>, 最終アクセス日: 2020 年 6 月 15 日
- 久木田水生 (2017) 「機械の中の道徳—道徳的であるとはそもそもどういうことを考える—」久木田水生・神崎宣次・佐々木拓『ロボットからの倫理学入門』名古屋大学出版会, pp.3-21.
- リー, ナミン (2009) 「現象学と質的研究の方法」『死生学研究』第 12 号, 東京大学大学院人文社会系研究科, 吉田聡訳, pp.8-34.
- Leopold, Aldo (1968) *A Sand County Almanac and Sketches Here and There: With Other Essays on Conservation from Round River* [1949]. New York: Oxford University Press.
〔邦訳: アルド・レオポルド (新島昭訳) 『野生のうたが聞こえる』講談社 (講談社学術文庫), 1997 年〕
- Nayar, Pramod K. (2014) *Posthumanism*. Cambridge: Polity Press.
- 根村直美 (2016) 「『イノセンス』に見るポスト・ヒューマニズムとく身体>の構築主義」『社会情報学』(社会情報学会), 第 5 巻 1 号, pp.73-88.
- (2017) 『ポストヒューマン・エシックス序説—サイバー・カルチャーの身体を問う—』青弓社.
- 西垣通 (2007) 「普遍倫理への模索」西垣通・竹之内禎編著『情報倫理の思想』, NTT 出版, pp.7-45.
- 真部雄介・細野繁・今井晴基 (2018) 「編集にあたって」『情報処理』(情報処理学会), 第 59 巻 8 号, pp.692-697.
- Singer, Peter (2009) *Animal Liberation* [1975]. New York: Harper Perennial Modern Classics.
〔邦訳: ピーター・シンガー (戸田清訳) 『動物の解放』改訂版, 人文書院, 2011 年〕
- 総合科学技術会議・生命倫理専門調査会 (2004) 『ヒト胚の取扱いに関する基本的考え方 (中間報告書)』内閣府.
- 総合科学技術会議 (2004) 『ヒト胚の取扱いに関する基本的考え方』内閣府.
- 諏訪正樹 (2015) 『一人称研究のすすめ—知能研究の新しい潮流—』近代科学社.
- ウォーバートン, ナイジェル (2010) 『哲学の基礎』(栗原泉訳), 講談社.

【謝辞】

本研究は, 日本大学経済学部・グローバル研究センターの助成を受けて行われたプロジェクト研究「ロボット倫理はどこに向かうのか」(研究代表者: 根村直美, 平成 30 年度 - 令和元年度) の成果の一部である。ここに謝意を表したい。