

日本におけるロボット倫理学

岡本 慎平

はじめに

21世紀に入って既に十年以上が経過した今、かつてはSF物語のガジェットとしか認識されていなかった先端テクノロジーの問題が、現実を生じうる喫緊の倫理問題として表面化してきた。ロボット工学の発展に伴って近年予見されている問題もその一つである。既に諸外国では、人間の管理監督を伴わない自律的ロボットが引き起こす倫理問題について軍事面や医療面を中心に様々な議論が活発に行われ、こうしたロボットの開発はどのような目的をもって、どの程度まで進めるべきなのかが議論されている。このようなロボットに関する倫理問題を扱う領域は、ロボット倫理学 (Robot Ethics) やロボエシックス (Roboethics) という名で形成されつつある。

ロボット倫理学の各論を見渡すと、とりわけ重要な争点となっているのはロボット兵器の問題である。いくつか代表的なものの概観すると、ロボット兵器の開発状況やその実用化の現状を広く知らしめたSinger (2009) やKrishnan (2009) に始まり、2012年11月には国際的人権団体のヒューマン・ライツ・ウォッチが人道的見地から自律兵器使用の国際的規制を要請する報告書『失われつつある人間性：殺人ロボットに反対する根拠 (*Losing Humanity: The Case against Killer Robots*)』を提出し、国際的な反殺人ロボットキャンペーンを開始している。それに対し、ロボット兵器は人間の兵士に比べて戦争犯罪をおかす可能性や誤射・誤爆の可能性も低く、むしろ積極的に開発・導入を進めるべきだと擁護するArkin (2009) など肯定的議論も存在する。もちろんロボット倫理学の議論はロボット兵器の問題に尽きるものではない。デイヴィッド・レヴィは人間の性的・恋愛パートナーとしてのロボットの可能性を示唆し⁽¹⁾ [Levy 2007]、ウェンデル・ウォラックとコリン・アレンは認知工学と倫理学諸理論の双方の側面から、人工知能

(1) 恋愛パートナーとしてのロボットの可能性をめぐるレヴィらの議論の詳細な検討は、本特集の西條玲奈氏の論文を参照せよ。

にはいかなる「道徳」が実装可能なのかを探求した⁽²⁾ [Wallach and Allen 2008]。そして2011年にはパトリック・リンらによって、これらの議論を包括的に概観した論文集『ロボット倫理学 (*Robot Ethics*)』が出版されている [Lin, et al 2011]。

こうして欧米を中心にロボット倫理学が盛り上がりの兆しを見せている一方で、この領域における日本の存在感は未だ非常に薄いと云わざるをえない [Kitano 2005、仲田2010]。例えば先述のパトリック・リンも、ウォール・ストリート・ジャーナルでのインタビューにおいて *Robot Ethics* 所収論文に日本人研究者による論考が存在しないことに触れ、「日本はロボットの分野が世界的にも盛んなので、ぜひ日本の学者、業界の専門家の論文を掲載したかったが、うまく見つからなかった」と述べている⁽³⁾。日本の研究者は、ロボットの倫理問題に対して沈黙を保ったまま何のリアクションもとってこなかったのだろうか？

もちろん日本人がこれまでロボットについて何も考えてこなかったわけではない。むしろ世界的にもトップレベルの水準にあるロボット工学技術や、マンガやアニメに登場するロボットたちの社会的影響を背景にして、これまで多くの日本の研究者がロボットの倫理問題を論じてきた。そしてこれらの議論の中には、検討に値する重要な論点も多数含まれている。そこで、これまでなされてきたロボット倫理学の議論の確認と、これからのロボット倫理学の展望を見定めるため、本稿では日本におけるロボット倫理学の諸議論を概観したい。

1 日本におけるロボット倫理学、その黎明期

1.1 吉田夏彦——ロボットの哲学

日本において産業用ロボットが急激に普及しつつあった1970年代初頭、ロボット工学の発展やその社会実装について哲学的・倫理的な観点から考察する試みは既に出現していた。吉田夏彦による『ロボットの哲学』[吉田1971]は、ロボット工学の発展が人間社会に与えるインパクトを様々な方面から素描したものである。

吉田によれば、ゲーテの『ファウスト』に登場するホムンクルスに代表されるように、人間は自分自身の手によって「生物、特に人間に似た機械」であるロボットを作り出すことを夢見てきた [吉田1971、2頁]。そして同時に、創造主である人間に対してロボットが「叛乱をお

(2) 「人工知能への道徳の実装」を扱う領域は「機械倫理学 (Machine Ethics)」とも呼ばれ、哲学的観点から最も活発に議論が繰り広げられている問題の一つである。こうした研究を包括的に扱った論文集に Anderson & Anderson (2011) がある。また、Gunkel (2012) は Allen らや Anderson らに批判的な立場から、これらの議論における道徳的行為者性 (moral agency) や道徳的被行為者性 (moral patiency) といった概念を検討している。

(3) 「【日本版コラム】無人爆撃機の利用は米国憲法に違反するか? ロボット倫理の専門家に聞く」ウォール・ストリート・ジャーナル日本版 2012年7月26日 http://jp.wsj.com/public/page/0_0_WJPP_7000-484015.html

こして、人類をみなごろしにしてしまう」[7頁] チャベックの悪夢を恐れてきた。さて、科学技術の発展によって我々の社会にロボットが登場したといっても、当時の日本で普及しはじめた産業ロボットは、我々とコミュニケーションをとりうる鉄腕アトムのような「人間の代用品」ではなく、工業の省力化といった目的のため人間の行動の一部を補佐する「マン・マシン・システム」である。しかしながら、産業ロボットの目的は人間の労力の省力化であるため、いずれは「人間を不可欠の要素としていた仕事を、この無人化工場が単独でなしとげる」[12頁] という、ある意味SF作品で描かれてきた「ロボットの奉仕する社会」と同様の社会になりうる可能性も十二分にあると吉田は論じる。

このような観点から、「人間とロボットの共生」をSF的な夢物語ではなく近い将来現実に起こりうる問題として扱い、そこに生じる様々な哲学的問題を概観しようというのが吉田の議論の骨子である。吉田のロボット哲学は倫理的・社会的問題にかぎらず、論理学や言語哲学を含めて非常に多岐にわたっており、その全てをここで紹介することは出来ない。だが第七章「人間とロボット」は、人間そっくりのロボットの登場によって社会がどのように変化するのかを論じている点でとりわけ注目に値する。吉田は、ロボットの登場が我々の社会に与える影響について三種類の可能性を提示している。第一に、科学革命において天動説の崩壊が宇宙^{マクロコスモス}における超越界の否定を伴ったように、人間と同じように喋り考えるロボットの登場が人間の心における超越界の否定を伴ってしまうという可能性である [220頁]。換言すれば、「心をもったロボット」が完成するということは、人間の心が科学の自然主義的世界観において十全に理解されるということでもある。もし人間の心が機械の心と同様のものであれば、宗教や美など超越的なものの価値が薄れてしまい、「ロボットとの共生により、一切の超越者がわれわれの視界から消えてしまう」[221頁] かもしれないと吉田は懸念する。これが第一の可能性である。第二に、ロボットの普及によって生産活動の省力化が究極的に進展することで、社会を円滑に運営するマネジメント能力を備えた少数の管理者層と、管理者による支配を受けるだけの大多数という形で、人類が二極分化してしまう可能性である [224頁]。そして第三に、第二の予測と同様に人間が単純労働に従事する必要がなくなることで、全ての人間が知的生産活動に従事し「創造的活動におけるプライオリティをめぐってはげしいあらしが展開される」[226頁] という「厳しい競争社会」の到来という可能性である。

これらの未来像の中には、比較的この21世紀の現状に当てはまるものもあれば、まったくの杞憂に終わったものもある。しかしここで重要なのは、少なくとも1970年代にはロボットのもたらす問題についての議論が存在したというその先駆性である。また吉田はロボット工学者の吉川弘之と批評家の立花隆による対談本『ロボットが街を歩く日』[吉川・立花1987]にも「ロボットと人間」という論考を寄せており、ロボット工学の発展段階から長年にわたりロボットと哲学の関係について議論を深めていたことが分かる。

1.2 黒崎政男——哲学者クロサキはアンドロイドの夢を見たか？

黒崎政男もまた、ロボットの社会的・倫理的・哲学的影響について比較的早期から考察した哲学者の一人である。黒崎は『哲学者はアンドロイドの夢を見たか——人工知能の哲学』[黒崎1987]において、人工知能開発と近代哲学における記号論の歴史的展開を参照しつつ、ロボットの哲学的問題についての議論を展開した。そして黒崎は、知能とは「ある存在者（人間やコンピュータ）そのものに内在している性質」ではなく「他の存在者とのかかわりの場において成立する事態」である、という立場（関係論的把握）をとる。これにより黒崎は、ロボットの心という問題は単に計算機能の観点からではなく、我々とロボットが社会においてどのような関係をもつのかという観点から考察しなければならないと指摘している [黒崎1987]。1998年の『となりのアンドロイド——哲学者クロサキの憂鬱』[黒崎1998]においても、黒崎（1987）と同様の立場である「知能の関係論的把握」がより一般的に敷衍されながら述べられている。

さて、黒崎は黒崎（1998）において、「社会で彼ら〔ロボット〕が何か問題を起こしたとき、どのような形で罪を問うのか、罰するのか」[黒崎1998、173頁]というロボット倫理の重要性を認める。しかし黒崎は、それをアイザック・アシモフが提示した「ロボット工学の三原則」に代表される「ロボットを主体とする倫理」として理解してはならないと強く戒める。というのも、たとえロボットが原理的には人間と同じような心をもちうるとしても、少なくともこの数十年の間には、責任能力を備えたロボットを作り出すことが不可能だからである。したがって、黒崎は以下のように明言する。

ロボットは主体にはなりません。あくまで、それらは人間が作ったものであって、もし倫理が必要であるとすれば、ロボットを製造する倫理、ロボットを使う倫理であって、決してロボットの倫理ではありません。倫理が必要なのは、〈作り・使う〉人間の側です。[170頁]

例えば、猛犬が人を襲えば責任を問われるのはその飼主であり、自動車が人をはねれば責任を問われるのはその運転手である。これらの問題と同じように、ロボットの倫理を考える際には使用者や製造者の責任という形で、人間を責任主体とする問題として考察しなければならないというのが黒崎のロボット倫理に対する態度である。

1.3 人工知能と心の哲学からのロボット倫理学

吉田や黒崎の議論に見られるように、たしかにロボットの社会的・倫理的問題も検討されていたものの、ロボットと哲学を結びつける話題として最も頻繁に論じられたのは「ロボットは心を持ちうるか」という問題だったという点は指摘しておくべきだろう⁽⁴⁾。

例えば大森荘蔵は『流れとよどみ』において、「ロボットにも心とか意識といったものがあるか」という問題を、他我論一般に拡張する形で議論している [大森1981]。そして大森は、「ロボットに心があるのか」という問題を、我々は人間に対するのと同じようにロボットに対して魂を持ったものとして振る舞いうるのか、言い換えればロボットは我々の「アニミズム」の範囲に入りうるのかという問題として捉え直すことを提言している [大森1981]。同じく80年代や90年代には先述の黒崎政男によって大森とほぼ同様の立場と思われる「知能の関係論的把握」が提起され [黒崎1987、黒崎1998]、2000年代に入ると分析哲学における心の哲学の議論を視野におさめながら、柴田正良による『ロボットの心』 [柴田2001] や、柴田を含めた心の哲学者たちによる『シリーズ心の哲学Ⅱ ロボット篇』 [信原2004] など、人工知能と心の哲学の問題が本格的に論じられてきた。哲学者側からのアプローチに限らず、谷口忠大の『コミュニケーションするロボットは創れるか』 [谷口2010] など工学者からの哲学的な心の問題へのアプローチも存在する。

こうした「ロボットの心」の議論の中でも、柴田(2001)や井頭昌彦の「[心を持ったロボットを作る]というプロジェクトはどのようなものでありうるか？」 [井頭2012] は、その倫理学への射程という点で特筆すべきだろう(柴田(2001)については第三節で後述する)。井頭は、「心を持ったロボット」の設計への切り口として、「痛みを感じられるロボット」の設計という課題を設定する [井頭2012, 394-5頁]。さて、言語をその使用から理解するアプローチをとると、「痛い」という語は「我々はどのような場面で「痛い」という語を使うべきか」という公共的側面から理解される。するとロボットに「痛み」を実装するためには、我々の言語実践において「ロボットは痛みを感じている」という言明が適切になるような振る舞いをロボットに実装することが求められる。このようにして、「心を持ったロボットを作る」という課題を「解釈主義的アプローチ」から試みようというのが井頭の提言である。しかしこのアプローチでは、たとえロボットがそのように振舞ったとしても、「痛みがあるように見えるだけで本当は違うのではないか」という疑問が生じる可能性(「チートの可能性」)があることを井頭は指摘する [407頁]。井頭はこの問題の解決案として、もし我々が痛みロボットを道徳的配慮の対象とみなすようになれば、我々は自然とそのロボットに対して心を持っているかのように扱うだろう、という道

(4) 例えば吉田夏彦を含む多くの参加者が「機械」とは何かをめぐって討論を重ねた心身問題シンポジウムの記録である『人間と機械』(1965年)では、ギルバート・ライルの『心の概念』を引き合いに出しながら、行動主義的アプローチによってロボットにおいても人間とまったく同等に心を認めようという主張が論じられている。

徳的配慮実践による直観の改定可能性を示唆している。そのため「心を持ったロボットを作る」ためには「我々の道徳的配慮を惹起する振る舞い」を実装することもまた重要であると指摘する [408頁]。

このように「ロボットの心」という哲学的関心からも、社会实践における問題としてのロボット倫理学へと向かう議論が展開されはじめている。

2 ロボット工学の発展とロボット倫理学の導入

2.1 産業ロボットからソーシャル・ロボットへ

21世紀に入ってからロボットへの注目が集まった理由には様々な要因が考えられる。例えば、ロボットの応用範囲が産業ロボットだけでなく、一般家庭や社会環境の中で人間とともに活動する「ソーシャル・ロボット」へと広がってきたことを挙げることも出来るだろう⁽⁵⁾。90年代から本格的に始まったソーシャル・ロボット研究、1996年に発表された本田技研の二足歩行ヒューマノイド「P2」、1999年に販売が開始されたSONYのペットロボット「AIBO」といった革新によって、人間とロボットが共生する社会の登場がいよいよ間近に迫った問題として認識されはじめた [井上2004, 30頁]。言い換えれば、ロボットが直接的に人間の福祉と関係をもつ状況が間近に迫ったのである⁽⁶⁾。

また経済産業省は平成16年に次世代ロボットビジョン懇談会の報告書「2025年の人間とロボットが共存する社会に向けて」を提出し、(1) 少子高齢化によるマンパワー不足の補助、(2) 大規模災害への対処や治安維持、(3) ゆとりある社会の実現、といった目標のため2025年を目処にロボット技術のさらなる飛躍を目指しはじめた⁽⁷⁾。こうした議論の中で必然的に、人間の福祉のためにどのような分野においてロボットが開発されなければならないか、その安全性はどのように確保するべきか、そしてロボットの登場による社会の変化の予測が急務の問題となった。

また21世紀初頭には諸外国においてロボット倫理の重要性が認識されはじめた時期でもある。とりわけイタリアのロボット工学者ジャンマルコ・ヴェルージオは2004年にロボエシックス

(5) 様々な工業生産に従事する産業ロボットに対し、1989年に新井らはソーシャル・ロボットを「多数の人間と共存」し、「不特定多数の人間の中を動きまわり、不特定多数の人間の指令に従う」ロボットとして定義した [新井1994, 24頁]。

(6) より人間の福利に関係するロボットとして、柴田崇徳らが1993年から開発を開始したセラピー用ロボットを挙げることも出来るだろう [柴田2006, 柴田・和田2011]。アニマル・セラピーのロボット版としてのロボット・セラピーについては、浜田 (2004) を参照。

(7) 「2025年の人間とロボットが共存する社会に向けて「次世代ロボットビジョン懇談会」報告書」<http://www.meti.go.jp/report/downloadfiles/g40302d32j.pdf>

国際シンポジウムを開催し、「ロボット工学の倫理学」としてのロボット倫理学が本格的に始動した⁽⁸⁾。

2.2 石原孝二——先端テクノロジーの倫理問題としてのロボット倫理学

こうした状況に伴い、またヴェルージオの「ロボエシックス」構築プロジェクトと連動しながら、日本においても哲学者・倫理学者によるロボット倫理学への本格的な取り組みも現れ始めた。例えば2009年、石原孝二は国内外のロボット工学者の協力の下でロボット倫理学の研究プロジェクトを開始し⁽⁹⁾、各地でロボット倫理学に関するワークショップを開催して議論を喚起した。またその一環として、2010年には*Accountability in Research*誌においてロボエシックス特集を編集し、彼自身も富士珠美との共著で、その特集のイントロダクションを執筆した。

この特集において石原らは、ロボット倫理学が近年論じるに値する論題へと至った原因として、二つの要素を挙げている。第一に、近年の科学技術の倫理学が、既に社会的に受容された問題だけでなく、急成長中の新たな科学技術の問題へも向けられはじめた点である。第二に、ロボット工学自体が神経科学やナノテクノロジーに比する形で急成長を遂げ、もはやSF的な絵空事ではなく、現実に生じつつある倫理問題として考えなければならなくなった点である。こうした点において、ロボット倫理学は神経科学を対象とする神経倫理学やナノテクノロジーを対象とするナノエシックスと同様の「先端テクノロジーの技術倫理」であると石原らは主張する [Ishihara and Fukushi 2010, pp. 273-4.]。

もちろん、これだけではロボット倫理学を既存の技術倫理学と区別し、新たな領域とみなす必要はない。だが石原らは、ロボット工学、とりわけソーシャル・ロボットの開発には二つの点でこれまでの技術倫理学の範囲外となる問題が存在することを指摘する。一つはロボットが人間社会に参入することによる社会構造の変化を予測し評価するという問題であり、もう一つはロボットの開発によって我々自身の人間理解にどのような変革が起こるのかという「ソーシャル・ロボットの研究が人間理解に及ぼすインパクト」[p. 275]の考察という問題である。こうした石原らの概要は、ロボット倫理学という新領域の実体を捉えるため非常に有益である。

2.3 ロボット倫理学の日本における紹介

また同時期に、諸外国のロボット倫理学への取り組みの国内への紹介も始まった [木村2007、仲田2010、虫明2010]。仲田誠の「ロボット倫理研究をめぐる批判的・倫理的研究」[仲

(8) これらロボット工学の技術倫理学としてのロボット倫理学の状況の概観は、本特集の本田康二郎氏の論文を参照せよ。

(9) 2009年度学術研究助成基金助成金採択「ロボエシックス基礎理論の構築」(代表者：石原孝二、研究課題番号21520004)

田2010]や「自律的ロボットと「場」」[仲田2012]は、ヴェルージオらによって展開されている諸外国のロボット倫理学の研究動向を幅広く、かつ詳細に紹介した論文である。またとりわけ仲田(2012)では、西洋のロボット倫理学の議論はロボットの「自律性」という問題に偏重しており、ロボットと人間が出会うところとしての「場」の多層性」という重要な観点が抜け落ちていることが指摘されている。

虫明茂の「機械じかけのエチカ——ロボエシックスの諸問題」[虫明2010]も同じく海外でのロボット倫理学の取り組みを紹介したもののだが、(1)ロボット開発の現状、(2)ロボットに対する文化的背景、(3)ロボット技術のガバナンスの問題、という形で内容の区分・整理を試みている点は注目に値する。また冒頭にも述べたロボット倫理学危急の問題であるArkin(2009)らロボット兵器の議論を検討した木村武史の「アメリカにおけるロボット技術の軍事利用に関するロボエシックス的一考察」[木村2011]は、アメリカにおけるロボット兵器開発の背景としてアメリカ文化での「武器」の価値などを指摘している点で、国外の議論には見られない独自の視点を確立していると言える。

3 日本におけるロボット倫理学の展開

前節で見たように、諸外国で生じたロボット倫理学の議論に応じて、日本においてもこの領域を構想する試みが現れ始めた。そのおおまかな枠組みは以上のような形で概観することが出来ただろう。だが時期は多少前後するものの、こうした導入以前からも、個々の日本人研究者はロボット倫理についての独自の議論を展開している。そこで、本節では2000年代を中心として、日本人研究者がどのような論を展開してきたのかを概観したい。

3.1 柴田正良——善悪のクオリアと自閉症のロボット倫理学

1.2節で概観したように、黒崎政男はロボットを責任主体とするロボット倫理学の不可能性を強調した。それに対し柴田正良は『ロボットの心』[柴田2001]の最終章で、「ロボットに善悪の倫理的判断をさせるにはどうしたらいいのか」[柴田2001、229頁]という問題に、言い換えれば責任主体としてのロボットはいかにして可能かという問題に取り組んだ。

柴田は、(1)善悪を感知するセンサーの開発、(2)すでに人間がおこなっている状況判断からの道徳判断の導出、という二つのアプローチはともに道徳判断の機能の実装として不十分であると指摘する。というのも、道徳性は物理的対象であるとは考えにくいため前者のアプローチではセンサーで感知する対象を決定できず、我々の道徳判断は行為への動機を内在しているのに対し後者のアプローチでは判断と動機の結びつきが行われないからである。そこで柴田は、「ある場面や行為に対して人がもつクオリアや判断から生ずる」[柴田2001、238頁]二階のクオリアとして、「善悪のクオリア」を想定する。もしロボットにクオリアを実装すること

が可能であれば、この「善悪のクオリア」も同様に実装することが可能だろう。だが二階のクオリアは状況の組み合わせ次第でどのような場合にも生じうるため、ロボットの道徳判断はロボットにとって主観的なものにならざるをえない。そこで柴田は、ロボットに「善悪のクオリア」を実装する際には、「道徳的判断を客観的な事実の反映であるかのごとく受け入れる」[柴田2001、241頁]という一種の「幻想」を組み込むことで、クオリアと実践の結びつきを確定させる必要があると主張する。

このような「善悪のクオリア」を素描するだけに留まった柴田(2001)に対し、2010年の‘Towards robot ethics through the ethics of autism’[Shibata 2010]や「異世界の者たちの倫理」[柴田2010]などの論文では、道徳的コミュニティの構成という観点からロボット倫理学が構想されている。

これらの論文は、銀河の深宇宙にただ一人だけで生きる「実際にも可能的にもあらゆるコミュニティのメンバーではない」孤独な人物「新クルーソー (Neo-Crusoe)」には行為の影響を受ける範囲内に一切の他者が存在しないため、彼は一切の倫理的責務に服さない、という「新クルーソーの思考実験」から議論が始まる。柴田によれば、こうした新クルーソーは、自閉症スペクトラム障害(ASD)患者やフルスペックの心を実装した自律的なロボットのアナロジーとして考えることが出来る。ASD患者は、「心の理論」を欠いているため、他者が自分と同じような道徳的義務を負い道徳的価値を有する「人間」であることを理解できず、倫理的な反応的態度をとることが出来ない。そのためASD患者は我々との間に権利と義務を相互に負う道徳的関係を作り出すことができず、「確かに生物学的には人間だが、道徳的な意味でのパーソンはなく、我々の道徳的コミュニティに位置づけることもできない」[Shibata 2010, p. 355] ことになると主張する論者もいる、と柴田は指摘する。これに対して柴田は、H. L. A. ハートが提示した道徳的コミュニティ成立のための諸条件を考慮に入れることでASD患者の道徳的コミュニティからの疎外を防ぐことが可能となる、というデボラ・バーンバウムの反論を妥当なものとして援用する。

だがバーンバウムの議論でも、「自らの内部状態から意図的判断を下すことが可能であり、いかなる規則も疑い、拒絶することができる」ようなロボットを我々の道徳的コミュニティの一員とみなすことは困難であると柴田は主張する。というのも、道徳的コミュニティを築くために必要な、誕生、死、認知能力、身体能力、概観、生殖などの、我々とASD患者の場合では共通していた存在条件が、ロボットの場合では我々と全く異なっているからである [p. 359]。

しかし、ロボットが我々と同等の権利・義務を負うことができないからといって、我々とロボットとの間には一切の倫理的関係が存在しえないというわけではないことを柴田は強調する。自閉症の倫理学を通したロボット倫理学の考察からは、例えばロボットが我々のコミュニ

ティに参入するためには「心の理論」の実装が必要不可欠である点など、人間とロボットの未来の倫理的展望に向けていくつかの論点を提示することが出来るというのが、柴田の結論である⁽¹⁰⁾。

3.2 久木田水生——総合倫理学としてのロボット倫理学

久木田水生は、ロボットが責任主体となりうるかという問題やロボットが我々と同じ道徳的コミュニティの一員になりえるかという問題とは異なった方向から、ロボット倫理学に対するアプローチをおこなった。2005年以降久木田は継続的にロボット倫理学の論文を発表しているが、端緒となった論文は「フィクションの中のロボット」[久木田2005]である。

この論文において久木田は、ホワイダニット（なぜ犯人は事件を起こしてしまったのか？という動機を謎とするジャンル）のミステリー小説には、人間の動機はあまりに突飛なものでありうるため合理的推論が結論を正当化するとは限らないというジレンマがあると指摘する。こうしたホワイダニットの難点とミステリー小説の醍醐味である数学的な美しさを両立させている稀有な作品として、久木田はアイザック・アシモフの『われはロボット』を挙げる。というのも、アシモフの小説に登場するロボットたちは、「ロボット三原則という公理・公準からの論理的な帰結であるがゆえに、読者はその帰結に納得せざるをえない」[久木田2005、60頁]からである。そのため、「公準からの証明」によって「ホワイダニット」の謎を説明することが可能となる。そして、ロボット三原則が公準としての説得力を持っているのは、この三原則が「人間の倫理体系の一つのモデル」[63頁]となっているからだとして久木田は指摘する。三原則のどの規則も、我々人間が従っている行動原理を抽象化したものである。つまりアシモフの小説の魅力は、ロボットという理想化された人間を描くことを通じて、「規則に縛られるがゆえに困難な状況に直面して、戸惑う人間の姿」を描くことに成功していることだというのが久木田の主張である。

久木田（2005）以降の論文、「ロボット倫理学の可能性」[久木田2009]や「総合倫理学（synthetic ethics）に向けて」[久木田2010]では、ロボットを通じて人間の倫理を理解しようというアプローチがさらに展開されている。久木田（2009）では、まずロボット倫理学の扱うべきテーマが以下のように区分された。[久木田2009、2頁]

- (1) ロボットを製造する際の倫理
- (2) ロボットの守るべき倫理
- (3) ロボットに対する倫理

(10) 柴田（2011）でも、簡潔にはあるがこの観点が要約されている [柴田2011、131-3頁]。

これらの問題群のうち「(1) ロボットを製造する際の倫理」は、どのような目的のために、どのような手法をとって開発していくべきなのかを問題とする技術者・開発者の倫理を問うものであるため、既存の技術倫理学とほぼ共通の問題圏にある。ロボット倫理学にとっての独自の問いとなるのは「(2) ロボットの守るべき倫理」と「(3) ロボットに対する倫理」である。(2)の問題を考察するためには、ロボットに「倫理」を実装するということがどのようなことなのか、そしてそもそもロボットが「心」を持つとはどのようなことなのかを明らかにしなければならない。そして(3)の問題には、ロボットが我々の社会に参入することにより、我々の道徳観や社会観はどのように変化するのか、そして我々はロボットに対してどのように振る舞うべきなのかが問われなければならない。

さて久木田は、ロボットが道徳的行為者 (moral agent) や道徳的受容者 (moral patient) になりうるのかという(2)や(3)の問題に対して、「理論的に不可能であるとは言えない」までも、「実現される可能性は低いと言わざるをえない」と主張する(久木田2009, 5頁)。というのも、現状の人工知能には行為者や受容者として必要な「意図や目的を持つ」能力が備わっていないからである。ロボットは道徳的な動機や責任を負うための自発性を持たないため道徳的行為者となれず、ロボット自身は自分の利害関心を持たないため道徳的受容者となることも出来ない。しかし久木田は、こうした実現不可能性という困難を抱えながらも、(2)の問題に付随する「ロボットに守るべき倫理を実装する」試みには、探究すべき固有の意義があると述べる。というのも、人工知能やロボットに道徳性を実装するという試みは、我々人間にとっての道徳がどのような過程を経て発生しどのように発展するのかをシミュレートすることによって、我々自身の道徳性の理解を促す可能性があるからである。久木田(2010)ではさらに、コンピュータによって進化や言語発達など実際には観察不可能な現象を再現することでそれらの現象を理解しようとする総合動物行動学とのアナロジーによって、人間の道徳性を理解するためのシミュレーションとしてのロボット倫理学が探究され、久木田はこれを「総合倫理学 (synthetic ethics)」⁽¹¹⁾と名付けている[久木田2010]。

3.3 北野菜穂——西洋と日本のロボット観の比較研究

北野菜穂は日本と欧米でのロボット倫理学に対する観点の違いを明らかにするため、日本と欧米でのロボット観の相違を比較検討した。北野によれば、ヨーロッパのあるロボット工学者は日本の現状を評し、「日本はロボット工学の最も優れた国の一つだが、日本のロボット工学研究者は自分自身の研究がもつ倫理的問題について妥当な意見をまったく持っていない。これ

(11) こうした、ロボットの開発を人間自身の理解を促進するための手掛かりとする久木田の立場は、ロボット工学において「自然システムのある側面を人口システムによって複製すること」で理解しようとする「構成論的アプローチ (Constructive Approach)」[谷口2010, 29頁]を倫理学において展開しようとする試みだと言っても良いだろう。同様の方向性を示唆する文献として浅田(2010)がある。

が日本の欠点だ」と述べたという [Kitano 2005, p. 93]。Kitano (2005) の目的は、ヨーロッパ人の考えているような形でのロボット倫理学が日本に存在しないのは「日本の欠点ではなく、日本と西洋の社会的・文化的、工業的な相違」[p. 93] にすぎないことを明らかにすることである。

北野はこれら三つの観点から、日本文化におけるロボットの扱われ方を検討した。第一に、社会的な観点である。日本においてヨーロッパの科学技術は、明治時代の文明開化によってもたらされた「啓蒙」という積極的な側面が重視されたため、例えばイングランドでのラッダイトのような反テクノロジー思想が日本に根付かなかったのだと北野は主張する [p. 98]。第二に、サブカルチャーの影響という観点である。日本で広く受容された『鉄腕アトム』や『ドラえもん』、『機動戦士ガンダム』といったマンガやアニメなど、様々な形で「幼少期からロボットに親近感を抱かせるメディアの力」[p. 99] が働いた。それに対し欧米の小説・映画に登場するロボットは、カレル・チャペックの『R. U. R』に代表されるように人間に対して反旗を翻す存在として描かれてきた点を、北野は日本と欧米の思想的風土の相違点として捉えている。そして第三に、経済的・工業的な観点である。戦後の日本は工業立国を目指し、アメリカやドイツの三倍近い世界最大の産業ロボット市場を開拓した。そのため、日本では欧米に比べてロボットに対する親近感が高いのだと北野は主張する。

北野によれば、これらの差異が、日本においてヨーロッパ的なトーンでのロボット倫理学が活発に論じられていない要因である。ヨーロッパでは、人間のコントロールを外れたロボットが暴走してしまう危険性が常に問題視されたため、危険なロボットをどのように制御するのかという観点からのロボット倫理学が主流となった。それに対し、北野は日本独自のロボット倫理学の観点として、たとえ人工物であっても万物に魂が宿りうるという「アニミズム」に基づくロボット倫理学の可能性を主張する⁽¹²⁾。この観点では、魂を持った存在として扱われるロボットが我々との間にどのような共生関係を築いていくのかという問題として、ロボット倫理学が位置づけられるのである⁽¹³⁾。

こうした北野の主張は、道徳的コミュニティの不可能性を主張する柴田正良の議論とは対照的に、アニミズムの観点を考慮に入れると人間と「ロボット」の間でも一種のコミュニティが成り立ちうることを強調するものである。そしてロボットと人間が共生する未来社会を考える上で、「人間とロボットのコミュニティ」を基調とする北野の論点は、「ロボットの制御」とは異なる側面からのロボット倫理学の構築につながる可能性もあるだろう。

(12) ロボットも心を持つかという問題に対するアニミズムの観点からの議論として、1.2節や1.3節で検討した大森 (1981) や黒崎 (1998) らの議論を指摘できる。また日本のヒューマノイド研究における「ロボットに魂を宿す」という潜在的な意図を指摘したものに木村 (2007) がある。

(13) Kitano (2006) ではこれらの相違点に加えて、コミュニティの規律を重視する日本と個人の判断を重視するヨーロッパでは「倫理」そのものの捉え方が異なっているという主張がなされている [Kitano 2006]。

3.4 高橋透——人間・機械・サイボーグ

最後に、ロボット技術による我々の社会实践の改定可能性を指摘した高橋透の議論を参照したい。高橋は『サイボーグ・エシックス』[高橋2006]において、現在様々な形で進められているロボット技術による人体のサイボーグ化の試みを概観し、サイボーグの登場が我々の素朴な人間理解や「人間の尊厳」という概念に対して重大な挑戦状を叩きつけていると主張した。高橋によれば、人間の尊厳という概念は人間の生命が人間以外のものとは異なった独特のものであるという考えを前提としている。しかしながら現在の遺伝子工学やロボット工学の発展は、人間自身も人為的に介入し、再現しうるものであることを示している。すると「尊厳」を保証するための人間の独自性・純粋性が次第に薄れていってしまうのではないかと高橋は主張する。

また高橋は、現在の科学技術は人間だけでなく自然環境にも様々な形で介入しているため、厳密に「人為の存在しないもの」という意味での「自然なるもの」は既に殆ど存在しないと述べる。そして、こうした「人為」と「自然」の二項対立の人間中心主義を克服するためにも、人間とそれ以外の垣根を攪乱する「サイボーグ」という事象に目を向ける必要があるという観点から、高橋は次のように主張する。

このように考えるならば、サイボーグ化の時代に必要なのは、人間の純粋性・純血性に依拠する、人間の尊厳という考え方であるというよりも、むしろ人間をサイボーグ、つまり人間と人間以外のものとのハイブリッドと本質規定するエシックスである、ということになるだろう。[高橋2006、153頁]

高橋によれば、ロボット技術に限らず、現在の先端テクノロジーは人間/機械/動物という区別を攪乱するものである。そして、この人間や自然と人為的なものの境界の攪乱によって、「尊厳」という倫理的概念の内実を改定する必要が出てきている。ロボット技術の応用の拡大により伝統的な倫理観が根底から覆される可能性を指摘している点など、高橋の主張には注目すべき論点が存在する。

おわりに

以上、管見の限りの紹介ではあるが、少なくとも、日本においてもこれまで多くの論者がロボット倫理学を議論の俎上に載せてきたことは分かるだろう。これらの議論を発展させ、また実際のロボットが引き起こす倫理問題を見定めることなど、検討すべき我々の課題は多い。日本におけるロボット倫理学の展開を活かした議論や社会实践をなしうるか否かは、まさにこれからにかかっていると言っても過言ではない。

参考文献

- 新井民男 (1994)「新世代ロボット」、吉川弘之 (編)『ロボット・ルネサンス——人間とロボットは共存できるか』三田出版会
- 浅田稔 (2010)『ロボットという思想——脳と知能の謎に挑む』NHK 出版
- 井頭昌彦 (2012)「「心を持ったロボットをつくる」というプロジェクトはどのようなものでありうるか?」『思索』第45巻第2号
- 井上博允 (2004)「ロボットの進歩と課題」、井上博允・家出武雄・安西祐一郎・瀬名秀明『岩波講座ロボット学1 ロボット学創成』岩波書店
- 大森荘蔵 (1981)『流れとよどみ』産業図書
- 木村武史 (2007)「ロボティクスの神話とロボエシックスの萌芽」松村一男・山中弘 (編)『神話と現代 (宗教学論書12)』リトン
- 木村武史 (2011)「アメリカにおけるロボット技術の軍事利用に関するロボエシックス的一考察」『筑波地域研究』第32号
- 久木田水生 (2005)「フィクションの中のロボット——ホワイダニットのジレンマとロボット工学三原則」『Prospectus』第8号
- 久木田水生 (2009)「ロボット倫理学の可能性」『Prospectus』第11号
- 久木田水生 (2010)「総合倫理学 (synthetic ethics) に向けて」『Prospectus』第12号
- 黒崎政男 (1987)『哲学者はアンドロイドの夢を見たか——人工知能の哲学』哲学書房
- 黒崎政男 (1998)『となりのアンドロイド——哲学者クロサキの憂鬱』NHK 出版
- 柴田正良 (2001)『ロボットの心——7つの哲学物語』講談社現代新書
- 柴田正良 (2010)「異世界の者たちの倫理——自閉症者の倫理・ロボットの倫理に向けて」『哲学・人間学論集』創刊号
- 柴田正良 (2011)「ロボットの哲学」『応用哲学を学ぶ人のために』世界思想社
- 柴田崇徳 (2006)「メンタルコミットロボット・パロとロボット・セラピーの展開」『日本ロボット学会誌』第24号3巻
- 柴田崇徳・和田一義 (2011)「アザラシ型ロボット「パロ」によるロボット・セラピーの効果の臨床・実証実験について」『日本ロボット学会誌』第29巻3号
- 高橋透 (2006)『サイボーグ・エシックス』水声社
- 谷口忠大 (2010)『コミュニケーションするロボットは創れるか——記号創発システムへの構成論的アプローチ』NTT 出版
- 仲田誠 (2010)「ロボット倫理研究をめぐる批判的・倫理的な研究——ロボットの「自律性」がなぜ問題になるのか」『松山大学論集』第21巻4号
- 仲田誠 (2012)「自律的ロボットと「場」——ロボット倫理研究をめぐる問題点についての批判的・存在論的考察」『筑波大学地域研究』第33号
- 信原幸弘 (2004)『シリーズ心の哲学II ロボット篇』勁草書房
- 浜田利満 (2004)「ロボット・セラピー・システム」桑子敏雄 (編)『いのちの倫理学』コロナ社
- 虫明茂 (2010)「機械じかけのエチカ——ロボエシックスの諸問題」『就実論叢』第40号
- 吉川弘之・立花隆 (1987)『ロボットが街を歩く日——剛い機械から柔らかない機械へ』三田出版会
- 吉田夏彦 (1971)『ロボットの哲学』日本経営出版会

- M. Anderson and S. L. Anderson (eds) (2011) *Machine Ethics*. Cambridge University Press.
- Arkin, Ronald C. (2009) *Governing Lethal Behavior in Autonomous Robots*. Chapman and Hall.
- Barnbaum, Deborah R. (2008) *The Ethics of Autism: Among Them, but Not of Them*. Indiana University Press.
- Gunkel, David J. (2012) *The Machine Question: Critical Perspectives on AI, Robots, and Ethics*. MIT Press.
- Kitano Naho (2005) Roboethics: a comparative analysis of social acceptance of robots between the West and Japan. 『社会学研論集』 第6号
- Kitano Naho (2006) ‘Rinri’: An Incitement towards the Existence of Robots in Japanese Society. in *International Review of Information Ethics*.
- Krishnan, Armin (2009) *Killer Robots: Legality and Ethicality of Autonomous Weapons*. Ashgate Pub.
- Ishihara Kohji and Fukushi Tamami (2010) Introduction: Roboethics as an Emerging Field of Ethics of Technology. in *Accountability in Research*. Vol. 17. pp. 273–277.
- Levy, David. (2007), *Love and Sex with Robots*, Harper Collins
- P. Lin, K. Abney and G. A. Bakey (eds) (2011) *Robot Ethics: The Ethical and Social Implications of Robotics*. MIT Press
- Shibata Masayoshi. (2011) Toward robot ethics through the ethics of autism. in J. L. Krichmar and H. Wagatsuma (eds) *Neuromorphic and Brain-Based Robots: Trends and Perspectives*. Cambridge University Press.
- Singer, P. W. (2009) *Wired for War: The Robotics Revolution and Conflict in the 21st Century*. Penguin Press. (小林由香利訳 (2010) 『ロボット兵士の戦争』 NHK 出版)
- W. Wallach and C. Allen (2008) *Moral Machines: Teaching Robots Right from Wrong*. Oxford University Press.