松 山 大 学 論 集
第 21 巻 第 4 号 抜 刷
２０１０ 年 ３ 月 発 行

ロボット倫理研究をめぐる批判的・倫理的研究

―― ロボットの「自律性」がなぜ問題になるのか ――

仲　田　　誠

ロボット倫理研究をめぐる批判的・倫理的研究

―― ロボットの「自律性」がなぜ問題になるのか ――

仲　田　　誠

1．は　じ　め　に

「ロボットと人間の相互作用 HRI（Human-Robot Interaction）」，「社会的ロボット」，あるいは「ロボット倫理」をめぐる諸問題について考えていこうというのがこの論文の狙いだが，この論文における考察の大半は，ロボットと人間の関係がそもそもなぜ重要な問題になりうるのか，その点をめぐる考察になる。（あるいはロボット倫理研究がなぜ必要とされているのかについての考察。）HRI や Robotethics（Roboethics），Social Robot は欧米では近年さかんに行われている研究だが[1]，日本では必ずしもそうではない。実際，筆者も欧米の情報倫理の専門家に日本のロボット倫理研究の状況を聞かれて（ヨーロッパ，米国では，情報倫理の研究者がロボット倫理の研究を同時に行っていることが多い），とまどうことが多かった。ロボット倫理の研究というものが筆者の周囲で行われているというケースをほとんど見聞きすることがなかったし，筆者自身なぜそうした研究が真剣な学問的探求の対象になりうるのか，まったく見当がつかなかったからである。一体，ホンダのアシモやソニーのアイボがなぜ，倫理的研究の対象になりうるのか，そのような社会的ロボット，ペットロボッ

1）According to Gianmarco Veruggio, Fiorella Operto, “the name Roboethics was officially proposed during the First International Symposium of Roboethics（Sanremo, Jan/Feb. 2004），and rapidly showed its potential.”（Veruggio and Operto, 2006, “Roboethics : a Bottom-up Interdisciplinary Discourse in the Field of Applied Ethics in Robotics,” *IRIE* 2006 vol. 6（Ethics in robotics），pp. 2-8.）

トと人間との「相互作用」がまじめな研究対象になりうるのか，正面から聞かれて（筆者と交流があるドイツのラファエル・カプーロという研究者にNakada, what do you think of Aibo from ethical perspectives？と聞かれた時は，ほんとうに「藪から棒」，「あるいは藪からロボット」という感じであった）きちんとその場で答えられる人が日本にどのくらいいるであろうか。

Kitanoという日本のロボット学者も，「日本のロボット学の研究者は，人間の社会の中でのロボットの普及や働きによる社会的・倫理的問題の検討に熱心な西洋の学者に較べて，倫理的問題に関心をもつことがほとんどない。日本の研究者は，ロボットのメカニズム・機能の向上という点ばかりに関心を向けている」と筆者と同様の見解を表明している[2)]。

欧米でのロボット倫理研究，社会的ロボット研究の内容に具体的に踏み込んでみると驚きはさらに強まる。たとえば，そこで行われている議論の中心にあるのは，ロボット自身の「主体性」または「自律性」（autonomy）[3)]，さらには「責任」である。ロボットを使用する側，あるいは製造する側の「責任」を問うというのであるならば，話は理解できる。その意味での「ロボット倫理」や「社会的ロボット（の使用・製造）」をめぐる問題，議論の必要性も理解できる。

しかし，「ロボット倫理」や「社会的ロボット」に関する論考で問題になるのは，多くの場合，ロボット使用者や製造者の責任ではなく，ロボット自身の側の「自律性」や「責任」である。

2）Kitano, Naho, “Animism, *Rinri*, Modernization ; the Base of Japanese Robotics,” ICRA’07 2007 IEEE International Conference on Robotics and Automation 10-14 April 2007, Roma, Italy. Full Day Workshop on Roboethics Rome, 14 April 2007（The world’s leading professional association for the advancement of technology）（http://www.roboethics.org/icra07/contributions.html. 2009年8月24日アクセス。）Kitano, Naho, “‘Rinri’ : An Incitement towards the Existence of Robots in Japanese Society,” *IRIE* 2006 vol. 6（Ethics in robotics）, pp. 78-83.）

3）自立性または自律性，どちらのことばを使うかは難しい問題だが，一応，ここでは自律性ということばを使うことにする。ヨーロッパや米国のロボット倫理・社会的ロボット研究者がAutonomyということばを使用するときには，Autonomyにはロボットが自分自身の責任を自覚し，主体的に判断するというニュアンスがこめられていることが多いように感じられるからである。

John P. Sullins という研究者は，ロボットは moral agent と見なしうる（見なしてよい）存在だと考える。ただし，それには，ロボットが，「自律性」だけでなく，「意図」（一定の意図 intention をもって行動する，善や悪をなしたりする）をもつかどうか，「責任感」（他の moral agent への責任感）をもつかどうかという点も考慮にいれなくてはならない。この３つの点をクリアーできれば，ロボットは自律的で責任をもつ（ロボットの責任を問える，あるいは，ロボットに責任を帰属させることができる）存在だと考えて良いということになる，こう彼は断言する[4]。ちなみに肝腎の「自律性」には，「プログラマーやオペレーターから自律していること」という説明しか与えられていない。

ここまで具体的に「モラルロボット」の条件を持ちだすということは，ロボットの自律性や責任能力（責任感），意図をまじめに考えているからに違いない。

Noel Sharkey の一見まともな，なんの反駁もしようもない議論・主張も，もし，その背後に John P. Sullins と同じような議論（あるいはそれと関連する世界観）が潜んでいるのだとすると，私たちにとっては（少なくとも筆者にとっては）彼我の心，精神の間に潜む想像以上の隔たりを垣間見させる理解困難なものになる。（自律的ロボットへの戸惑いはそのまま自律的ロボットの登場を期待する（あるいはまじめに危惧する）ヨーロッパ，米国の研究者への戸惑いである。）

「サービスロボットの使用は予期せぬリスクと倫理的問題をもたらすものになる。倫理的問題の２つの大きな領域は，子供や高齢者の（ロボットによる）世話と軍の自動（自律）ロボット兵器の開発である。…人間との接触抜きで何時間もあるいは何日も（ロボットにまかせて）放っておかれた子供が社会的孤立によって受ける心理的影響については十分にはわかっていない。…自動的に（自律的に）目標を補足して人間の関与（指令）抜きで相手を攻撃するロボットについては，至近距離で相手を見分ける能力をもったコンピュータがないと

4）John P. Sullins, 2007, “When Is a Robot a Moral Agent?,” ICRA 07-Workshop on Roboethics（subimitted paper to this workshop）.

いうことから倫理的問題が生じる[5]。」

ロボット倫理をめぐる問題（わかりにくさ）は，「（ロボットの）自律性」の問題に集約的にあらわれているのだが，一方，そうはいいつつもロボット倫理の問題にどこかで引きつけられていくのは（少なくともこれほど多くの研究者がロボット倫理に関心を寄せる理由は知りたいと思う），ロボットの存在そのものの不思議さによるものだろう。ロボットの「自律性」をめぐる問題に戸惑いやばかばかしさを感じるのは，「自律性」は人間だけに適用されるべき「属性」だと考えるからであり，一方，ロボットには，「自律性」はともかく，つい人間的な視線をそこに向けて，さらには人間的な「属性」をそこに「帰属」させてしまいがちな不思議さがある。ロボットは人間に近いからである。

しかし，同時にロボットは人間とは異種で（機械だから），その意味で人間からは遠い。近くて遠い存在だから，ロボットはさまざまな人間的でかつ機械的な仕事を押し付けられる。

ロボットの「自律性」をめぐる議論も実は「自律性」だけが問題になっているわけではない。「自律的」で人間的なものとされたロボットは（考えてみればわかることだが）同時にかならず人間的でないこと，人間がいやがることを押し付けられている。工場で一日中同じ作業をさせられるとか，戦場で危険な爆弾処理をさせられるとか，そういった仕事だし，いわゆる「社会的ロボット」やペットロボットの場合も，ただ，人間のごきげんとりをさせられるだけで，「相手」に向かって不満をぶちまけたりすることはできない。（工場で一日中同じ作業をさせられることは，人間にももちろんあるが，その場合，人間はそれにみあった報酬を与えられるとか，組織に所属する帰属意識を与えられるとか，あるいは仕事そのものに達成感を覚えたりする。）

人間的でありながら，人間ではない，人間に近いが一方で人間から遠い，そもそもそこにロボットの「存在意義」があるわけだが，いわゆるアシモフの３

5）Noel Sharkey, 2008,“COMPUTER SCIENCE : The Ethical Frontiers of Robotics,”*Science* 19, December 2008 : Vol. 322. no. 5909, pp. 1800-1801.

原則のわかりにくさも，その「遠さと近さ」の感じ方に彼我の間でずれがあることによるものだろう。これは今のところ「直感」でしかないが，「遠さと近さ」のずれがどのような場面で感じられやすいかがたぶん欧米と日本で違うのであろう。

ヨーロッパや米国では，ロボットは人間の友達というよりも，人間の使用人，奴隷的な側面が強いように思われる。人間に近い存在であり，人間と違った存在であるロボットは，人間と同じ仕事を人間から押し付けられ，しかも，ロボットは人間に近い存在であるがゆえに（ロボットが担う苦役は本来人間が担うもの），ロボットの苦役は人間的な意味を帯びたものとして人間からは受けとめられる。

アシモフの３原則，「ロボットは人間に危害を加えてはならない」，「ロボットは人間に服従しなくてはならない」，「ロボットは（人間に危害を加えない，人間の命令に従うという前提のもとで）自己を守らなければならない」はまさに，人間が奴隷的存在に課する命令である。一見，ばかばかしいアシモフの３原則はその大前提を考慮にいれると，ただ，ばかばかしいものというだけではなくなる。（欧米でもロボットと人間の友達関係的な側面に注目する研究者もいる，そのような研究者は，日本の友達ロボット，ペットロボットを事例としてとりあげることが多い[6]。）倫理的問題はロボットが人間と友達であると想定されている場面ではあまり切実な問題としては見えてこないが，ロボットがいわば奴隷として働かされる場面では，倫理的問題は実感として感じられやすいのであろう。

たぶん，われわれにとって，ロボットの近さと遠さが実感されやすいのは，身近な「友達ロボット」（WAKAMARU[7]）が「フェチロボット」（HRP-4C[8]）

6）たとえば，Jutta Weber の研究。Weber, Jutta : Human-Robot Interaction. In : Sigrid Kelsey（ed.）, 2009, *Handbook of Research on Computer-Mediated Communication*. Hershey, PA : Idea Group Publisher.

7）http://www.mhi.co.jp/kobe/wakamaru/english/live/index02.html

8）http://www.aist.go.jp/index_en.html

に変容する場面においてであろう（「フェチロボット」以外のものへの変容でももちろんいいが，この例がわかりやすい）。

「友達ロボット」と近くて遠い「フェチロボット」の写真を同時にながめると，「近さと遠さ」はやはり日本的なロボットの場合でも隠れたレベルで働いていることが感覚的に理解される。

以下，「自律性」の問題，ロボット倫理研究の「わかりにくさ」の問題を読み解く作業をすすめてみたいが，筆者の考えでは，ロボット倫理の研究者の役割は，ロボットの近さと遠さの二面性から生じてくる問題をてんびんにかけて，多くの人が納得するバランス点をみつけて，さらにそれをロボットの機能のなかにもちこむ（そのように技術者，ロボット工学者に提言する）ことにある。この「バランス」とは，ロボットは機械なので，ロボットを動かすプログラムの形で再現させるしかないが，具体的に言えば，「自律的」に行動したロボットが引き起こすかもしれない望ましくない結果（なぜ望ましくない結果が引き起こされるかというと，ロボットは遠い存在として，しかも，人間的な環境の中で＝近い存在として＝「いやな」仕事をさせられるからだ）をあらかじめシミュレーションして，それを避ける機能をロボットに組み込み，「望まし

くないがしてもらわなくては困る」のレベルを一定水準以下に保つように調整するということである。「望ましくない」が一定水準を越えて，法律的な問題になり，人間の側の責任問題になることを避けなければならない。

したがって，ロボット倫理をめぐる問題は，基本的にこの近さと遠さのズレ，あるいは，亀裂のなかにあることになるが，筆者の考えるところでは，ロボット倫理研究の問題点は（わかりにくさ），ロボット倫理研究を成立させるために必要ないくつかのの要素がしばしば，ロボット倫理をめぐる研究，考察の中から消えてしまっていることから生じる。一つは，ロボットが近くて遠い存在であることではじめて可能になる人間の欲望，身勝手さをめぐる問題であり，別の一つは，ロボットの近さと遠さをいっぺんに眺めることができる（あるいはそうしないと問題を設定することが最初からできなくなる）ロボット倫理研究者の特権的な立ち位置であり，さらには，ロボットが機能することを可能にしている人間的な環境の変容である。こういった要素がなぜどのようなかたちで消えているのかについて考えることは（筆者の考えでは），ロボット倫理研究を，たんなる人間の身勝手さの自己弁護的な研究ではなく，人間そのものについて考える一段水準の高いものに変えていく可能性につながることになる。いわば，ロボット倫理に関する倫理的・批判的な研究の可能性がそこで生じることになる。

前置きがながくなったが，以下，具体的なロボット倫理研究，HRI 研究の内容に踏み込みながら，問題点をさらに深く掘り下げてみることにしよう。

2．ロボットの自律性やモラリティをめぐるさまざまな意見

すでに部分的には触れたことだが，ロボット倫理，社会的ロボット研究に関する先行研究の事例を繙いてみて感じることは，そもそもなぜロボット倫理や，ロボットと人間の相互作用を問題にするのか，その点の議論がほとんどない点である。自律性やモラリティをなぜ問題にするのかその点の議論がないま

ま，自律性とは何か，モラリティとは何か，さまざまな説が紹介され，立場の異なる論者の間で論争がされているというのが実態である。

Brian R. Duffy は，ロボット倫理に関する具体的な問題として次のような問題を提起する[9]。「人間との相互作用という文脈のなかで，ロボットが自分の行動を評価する（モラル的な能力として）ようにプログラムされるべきかなのだろうか。」「モラル的であるとは何であるかわかるために，ロボットが人間的能力をもつことは必要なことなのだろうか。」

こうした問いかけは，一見まともで具体的な問題提起であるかのように見えて，実は機械やロボットに自律性，自律的なモラル判断（行為）能力を与えること（を議論する）がなぜ必要なのか，肝心な点についてはまったく議論しておらず，問いかけの前提としての状況設定が見えにくくなっている。

Gianmarco Veruggio というイタリアの研究者（ロボット倫理のロードマップと称する論文を発表して注目を集めている研究者の一人である）は，ロボット倫理の研究は，ロボット倫理はまだ初期の定義づけの段階にあるものであると言う。2004 年に「ロボット倫理」ということばがサンレモの第一回ロボット倫理国際シンポジウムで提唱されたという事実からもわかるように，ロボット倫理はまだ始まったばかりの研究であり，したがって，ロボット倫理はまだその社会的必要性も含めて方向性が模索されている段階である。しかし，このような見解が提示されているにもかかわらず，Veruggio は最初から「ロボットは悪か善か，ロボットは人間にとって危険か」というロボット倫理の議論の出発点を疑おうとしはしていない。したがって，Veruggio がとりあげる情報倫理と

9）Brian R. Duffy, 2006,“Fundamental Issues in Social Robotics,”*IRIE* 2006 vol. 6（Ethics in robotics）, pp. 31-36. なお，この論文の中では，Maturana and Varela の議論を参照しながら，機械と生物の違いについて検討している。～生物と機械の違いは，アロポイエティック（allopoietic 他者創造的）とオートポエイティック（autopoietic）という点にあらわれている。～これ自体は，Maturana and Varela の議論のたんなる紹介にすぎないが，Duffy がこの論文で行なおうとしているのは，ロボットと人間の違いをさまざまな点から検討することである。そのことは評価できるのだが，問題は，なぜ，そこまでして，ロボットと人間の「近さと遠さ」を比較検討しなくてはならないかである。

は，「ロボットについて意識，感情，人格（そしてモラリティ）を語ることは正しいのか」という問いの前提，起源がかき消された問いである。われわれが知りたいのは，ロボットと感情，意識，モラリティを結びつける（一見まともで実は根拠が稀薄な）問いかけがどこからスタートするか，その点だが，それがないため，（問題設定のありかたが問われないから）Veruggio の情報倫理は結局は情報倫理学説史的なものになっていて具体的内容の乏しいものになっている。

Veruggio の論文の中では，「心無きエイジェント（情報エイジェント，モラル・エイジェント）の議論で名高い（あるいは物議をかもしている）フロリディの説[10] がとりあげられているが，これもやはり，情報倫理学説史的な文脈の中で語られていて，なぜその説をとりあげることが重要なのか，あるいは，ロボットの自律性やモラリティがなぜ問題になるのかは問われていない。

（ただ，Veruggio もたぶんまったく問題の所在に気づいていないというわけではないのだろう，そう好意的に推測させるような一文も彼の論文の中には含まれている。「ロボットに倫理（感，意識）を与えることと自律性をもたせること，この二つの必要性はどのように矛盾したことなのだろうか？」）

さきほどとりあげた Sullins の立場も，まず，最初にロボットはモラル的存在なのかどうかという一見明白で実は根拠のない問題設定をたてた上で，さまざまな意見を紹介し，その上で自説を述べるというかたちで議論を展開している。

Sullins によれば，ロボットのモラリティに関する学説は３つに別れる。ロボットがモラリティの意識や感覚をもつことなど全くの幻想だという意見，ロボットは疑似的なモラル・エイジェントまたは不完全なモラル・エイジェントだという意見，ロボットはほんもののモラル・エイジェントだという意見，こ

10）Floridi, L., 1998,“Information Ethics : On the Philosophical Foundation of Computer Ethics,” Ethicomp98, The Fourth International Conference on Ethical Issues of Information Technology, Erasmus University, The Netherlands, 25/27 March 1998. Floridi, Luciano and Sanders, J. W., 2004,“On the Morality of Artificial Agents,” *Minds and Machines*, 14. 3, pp. 349-379.

の3つである。Sullins自身は，「ロボットはほんもののモラル・エイジェントだ」という立場であることを宣言し，「ロボットがモラル・エイジェントでないなら，人間もモラル・エイジェントではない。人間（の信念，目的，欲求）は文化，環境，教育，脳の科学的構造の産物だからだ」という意見を主張している。

このようなSullinsの姿勢はフロリディの「心無きエイジェント」への好意的な視点にもつながる。フロリディは，エイジェントの行為が，環境のなかで相互作用的で適応的なら（state changesやプログラミングによって，プログラムは多少とも環境から独立している），それで，その実体がエイジェンシイであるとみなせる，とこう考える[11]。動物，環境，法的実体（会社のような）のモラルについてもこうした方向性から考えていこうとする議論はそれなりに興味深いが，しかし，ここでもやはり，なぜ，ロボットというエイジェントがなぜ問題になるのか，問題提起の出発点はどこにあるのかは隠されたままである。

Peter M. Asaroの論文[12]では，自律性，責任は権利の問題と結びつけながら提示され，最終的には，「ロボットには法律体系が適用可能かどうか」というばかばかしい問題の提起（これは問題提起というよりそのようなアジェンダを設定しようとする意思表明に近い）のかというかたちで終わる。Asaroもこうした問題が「ばかばかしい」という印象を与えかねないという点は十分に認識しているらしく，「胎児や昏睡状態の患者の法的対応，権利の問題，また，未成年の法的権利を問うことはばかばかしい問題ではないだろう」という論法で，ロボットへの法的適用を肯定的に論じる。

3．隠された動機・メタ認識としてのロボット倫理

ロボット倫理研究，社会的ロボット研究の現実感の乏しさは，筆者だけの個

11）Floridi, Luciano and Sanders, J. W., 2004, 前掲文献。

12）Peter M. Asaro, 2007, “Robots and Responsibility from a Legal Perspective,” a paper for ICRA’07 2007 IEEE International Conference on Robotics and Automation 10-14 April 2007, Roma, Italy（Full Day Workshop on Roboethics Rome, 14 April 2007）.

人的な印象ではなく，筆者が周囲の研究者や大学院，学部の学生（情報倫理や情報社会論を研究したり，それに関する授業を受講している）の意見を参考にして判断する限り，多くの日本の人々に共有されている印象であるように思えるのだが，一方で，ヨーロッパ，米国では，数多くの研究者がこの問題に真剣に取り組んでいるのはたしかである。われわれは，ただたんにロボット倫理研究，社会的ロボット研究のリアリティの稀薄さを指摘するだけでなく，ヨーロッパ，米国の研究者にとって何が研究の現実感を支えているのかを問う必要があるだろう。われわれには見えにくい動機やパースペクティヴをより見えやすいものにするという作業がここでは必要になる。（もっともともと見えにくいものを見る作業なので，以下での考察は本格的な解釈学的考察～異なった地平の融合をめざす～というよりも，より前段階的な仮説提示というかたちをとる。）

1）ロボットによる「決断」の必要性とロボット倫理研究者の特権的立ち位置

ヨーロッパ，米国でロボットの自律性が突然（われわれにはそう見える）アジェンダとして前面に出てくるのは，自律的なロボットの出現を強く求める動機があるからではないか。このことを強く感じさせるのが次の AFP の記事である（2009 年 8 月 17 日付け）。

> 戦争に行くということは，常に命を落とす覚悟を求められるものだが，ロボット兵器の開発が進む中，数世紀にわたる真理が変わろうとしている。
>
> アフガニスタンやイラク，パキスタンで活躍する米軍の無人攻撃機の「パイロット」は，攻撃が実行される戦場から数千キロも離れた場所で操縦かんを握り，危険にさらされることなくミサイルを発射することが可能だ。また，危険なルート上での物資運搬を担当し，敵の戦車を発見すると攻撃するロボットも現在開発中だという。[13)]

ロボットへの情熱が，人間の身勝手さに支えられていることは一目瞭然だが，こうした（身勝手さを象徴する）システムの延長にあるのは，攻撃命令を人間の手でするのではなく，ロボット，兵器自身の手で行うような技術の開発であろう。（自分の手を汚さずに，人間を殺傷するロボットへの隠されたというかほとんどむき出しの欲望。）そこでロボットの自律的判断・決断に関する議論が求められることになる。人間の兵士の場合も，やみくもに「敵」を攻撃，殺害してよいわけではなく，ジュネーブ諸条約など一定の戦時下のルールに従うことになる。ただ，こうしたルールの適用にあたっては判断が難しい場合がある。たとえば，「戦争時に発生した負傷者は，保護の対象になる」というルールだが，手傷を負った「敵」が自爆攻撃をしかけてきたらどうするか，「敵」が味方を人間の楯にして攻撃をしかけてきたらどうするか，[14] そのような場合にどうロボットに自律的に判断させるか，そのことを考える（「自律的」なロボットの判断の結果として発生しうる「望ましくない」事態についても考えておく）ことが（多分）ロボット倫理研究者の重要な仕事になる。

（もちろんプログラムに従ってだが，プログラムに従って行う「行為」がプログラマーやオペレーターからある範囲内で「自律」しているということは想定しうる。その場合は，ロボットの自律性がたしかに問題になるのが，同時にそれはあくまでも人間の側の〈「自律的」ロボットの使用・開発を開始する前の〉倫理的な熟慮，判断が前提になるわけである。）

ロボット倫理研究の重要な仕事の中には，「自律的」なロボットの判断の結果として発生しうる「望ましくない」事態が生じた場合，「だれに責任を負わ

13）APFBB ニュース。（http://www.afpbb.com/article/war-unrest/2631329/4464637）（2009年8月29日アクセス）

14）Patrick Lin らは，米国海軍の公式研究の一環としてこのような想定状況をまとめあげ，報告書のかたちにして公表している。その中では，軍事用ロボットに関する自律性をめぐる問題，戦闘員と民間人の区別，負傷者の見分けかた，戦時ルール・ガイダンスとロボットの「行動」との関係などが「シミュレーション」されている。Patrick Lin, George Bekey, and Keith Abney, 2008,“Autonomous Military Robotics : Risk, Ethics, and Design,”Prepared on : December 20, 2008（*This work is sponsored by the Department of the Navy, Office of Naval Research,under award # N 00014-07-1-1152*）.

せることにするか」という倫理的あるいは手続的判断，決定も含まれることになるが，一方で，倫理的あるいは手続的判断，決定に関して人間の側がそこに深く関わっている（当然のことだが）ということをはっきりと明示するのか望ましいことではない。ロボットになぜ，そのようないやな仕事を押し付けるのかという問いかけが生まれてきた場合の対処に困るからである。そこで，ロボットの「自律性」に関心を向ける（向けさせる）という「倫理的」で「戦略的」な問題設定が行われる。

ロボットを焦点にすえるとこうした事前の倫理的判断，手続きは「見えなくなる」が，この「見えない」部分を増やしていくことこそ，ロボット倫理学者の重要な作業であるかのようにも思える。しかし，同時にこれはロボット倫理研究の前提である，「近さと遠さ」の亀裂の上に立つ，「今」（ロボットが行動するのは「今」である），「過去」（ロボットのプログラムは「過去」にさかのぼって組み込まれている），「将来」（ロボットの行動が「将来」引き起こしかねない事態を想定するのは，人間である）を自己のイマジネーションの中で結びつけるというロボット倫理研究の出発点，問題設定が行われる場所（特権的な場所）そのものをやがて見えにくくしていくようにも思える。とくに，後発的に，すでに設定された問題状況の中に入っていく研究者にとってはそうであろう。やがて，そこから，ロボット倫理研究の仕事が単なるシミュレーション（「自律」に伴って生じるであろう望ましくない事態を想定し，それに対する対処をあらかじめ考えておくこと）に堕するという研究の質的変容が生じる。

ロボットを焦点にすえるロボット倫理研究者たちの「事前」の倫理的判断，手続きは（結果として）「見えにくくする」ことだが（この「見えない」部分，プロセス，人間の関与の度合いが低くなっている部分こそ，ロボットの「自律性」そのものである），逆に言えば，ロボット倫理研究を単なるシミュレーションに堕落させないためには，ロボット倫理研究者たちが，つねに，この「不可視化」と，それとはまったく正反対の「可視化」という作業を並行して進めていく必要があるということになる（これはロボット倫理研究を批判的な立場で

論じる場合だけでなく，ロボット倫理研究そのものに必要な作業だと思える)。

ロボット倫理の研究者の特権的な立ち位置，「メタ・倫理」（ゴフマンがいう「フレーム」のようなものだが，倫理的問題，あるいは法的問題のあり場をロボット倫理研究者が「宣言」することで可能になる議論のフレーム。これをとりあえず，ここでは，暫定的に「メタ・倫理」と称することにする）のあり方を「宣言」する立場，この点を含めてロボット研究全体を見通すことが，本来なら，ロボット倫理研究たちに求められることなのだが，また，それが，そもそも，ロボット倫理研究そのものの成立の前提になっているわけだが，この点まで含めてロボット倫理の問題を議論している研究者は筆者の知るかぎりほとんど見当たらない。

2）人間の側の自律性，責任観・意識の変化

ロボット倫理研究，社会的ロボット研究において，ロボットの側の「自律性」，「責任」が問題になってくる事情の背後には，たぶん人間の側の「自律性」，「責任（観・感・意識）」のありようの変化も関係しているのだろう。このことを気づかせてくれるのが，やはり，ミリタリー・ロボットをめぐる問題である。Times（オンライン版）のインタビュー[15]に答えて，Patrick Lin（注12で紹介した文献の研究者。このインタヴューではPatrick Linはたぶん，注12の報告書の執筆者として応答している）は，ロボットの「自律性」の必要性について次のようにコメントしている。

> 「ロボットに組み込まれたプログラムの内容によって指示される以外のことはロボットにはできない」というのは広く広まっているが誤った考え方だ。そういう考えは時代遅れで，プログラムがたった一人の人間によっ

15）2009年2月19日付け記事（http://technology.timesonline.co.uk/tol/news/tech_and_web/article5741334.ece）（2009年8月29日アクセス）

て書かれ，理解された時代の考えかたである。現在のプログラムは，何百万というライン，コードを含むもので，プログラマーのチームによって書かれている。プログラム全体を見通すことのできるような人間はだれ一人としていない。

「(だから) ロボットの側にプログラムとセットになった〈善悪を知る学習機能〉をもたせなければならない」(Patrick Lin) というのはずいぶん飛躍した議論だが，現代の高度な分業社会，ネットワーク化した世界の中での，人間の側の「自律性」，「責任」をめぐる事情の変化は言われてみればたしかにそうだ。

これがさらに単体のロボットではなく，いわゆるネットワークロボット (“Network Robot Systems” (NRS)) になると，全体を見通す個人の存在というのはますます想像しにくくなる。それと同時に，「自律」，「責任」という人間のありかたに由来することばで事態を把握することも難しくなる。

ネットワークロボットシステムは，たとえば，「身体をもったロボットの存在 (Physical embodiment of robot)」，「ロボットの自律的能力」，「ネットワークを通して，ロボットが人や環境センサーと協力・連携できるようなシステム (の仕組み)」，「物理的なロボット以外にも，環境にセンサーやアクチュエーターが存在」，「人間とロボットの相互作用」といったシステムの内容 (構成要件) を含むものとしてイメージされているが,[16] 常識的に考えれば，システムの複雑化・高度化は，ロボットが「自律的」である必要性を奪っていく。「目」や「耳」が環境の側にあれば，「自律的」ロボットに求められることはその分少なくなっていくと考えられるからだ。

この NRS は，多分軍事目的にも応用可能だが，そこでは，人間の生死が関わる場面だけに，「自律性」，「責任」の問題は避けて通ることができないよう

16) Alberto Sanfeliu, Norihiro Hagita, and Alessandro Saffiotti, 2008, “Network robot systems,” *Robotics and Autonomous Systems* 56 (2008) 793・97.
萩田紀博，2008，「ネットワークロボット概論」，『電子情報通信学会誌』(Vol. 91, No. 5)，346-352 頁。

にも思える。しかし，「戦場の看視」，「敵・味方の認識」，「物理的働きかけ（武器の使用）」という人間，ロボット，環境センサー・アクチュエーターの分業体制が完璧に機能すれば，「自律的」なロボットに「判断」をまかせ，責任をロボットに帰属させるということはやがてなくなるだろう。「主体」，「自律性」，「責任」という概念，それに関わる視点で，このシステムの問題に切り込んでいくことは一層むずかしくなる。その意味では，そこで，ロボット倫理は一旦終焉する。少なくとも，個々のロボット（スタンド・アローン型ロボット）の「善悪」の判断や「能動性」がロボット倫理の主たるテーマではなくなるであろう。

Veruggio によれば，ロボットの将来の方向性の一つとして，ネットなどを通じて，ロボットが相互に連絡しあい，情報を共有しあう，そのような状況が想定される[17]。ネットワークロボットの発展型なのかもしれないが，しかし，そうなれば，ますます，ロボット倫理は変容する。倫理や責任，判断は人間の側に戻ることになる。「心無きエイジェント（あるいはペイシェント）」であっても，それが「エイジェント」，「ペイシェント」として受け入れられるのは，それが単体（独立した個体，組織）として存在しているからであろう。「心」を身体の中にもたないロボットは，もはやロボットと呼ぶべき存在ではないように思える。ロボットの存在を背後で支えていた「遠さと近さ」の感覚がそこでは稀薄になっていかざるをえない。

3）メタ認識と社会的ロボット

ロボット倫理研究は，「不可視化」と「可視化」のせめぎ合いだと言ったが，このことを別のかたちで理解させてくれるのが，「心の理論」のロボット研究への応用（最近これがはやっているようだ）に関わる事態である。

17）Gianmarco Veruggio, 2005, “ROBOETHICS : a new Ethics for Humans and Robots,” Italy-Japan 2005 Workshop, The Man and the Robot : Italian and Japanese Approaches. (Tokyo, Sep 7, 8, 2005).

「心の理論」とは，近年注目されているという心理学の理論だが，この「心の理論」とは，「ある状況に置かれた他者の行動を見て，他者の考えを予測し，解釈することができる」という心の働きをさす[18]。もともと，霊長類研究から始まった研究で（ただし，メタ認知，メタコミュニケーション，状況意味認識という言葉の使用が示す通り＝「心の理論」はこうした言葉と結びつけられることが多い＝，「心の理論」およびその中核に位置する「再帰的認識」に類する議論，理論は心理学，メディア研究，哲学，精神医学，認知科学の領域などでずいぶん前から盛んであった＝著者（仲田）注＝），「心の理論」という用語を最初に使ったのは，Premack, D. & Woodruff, G.[19]（1978）だとされる。筆者は，「心の理論」のポイントは，「再帰的な心理状態」による他者理解，あるいはそうした「再帰的な心理状態」の理解モデル・フレームの共有による「私たちの心」モデルの成立・共有にあると考えるが，人間的なロボットを見ただけでわれわれが心の中にあれこれとイメージを思い描き，さまざまなことを感じてしまうメカニズムは，たしかに，このような「心の理論」，あるいはメタ認識モデルでかなりの部分が説明がつく。ちなみに，林は「再帰的認識」について次のように説明している。

> 再帰的な心的的状態とは，「メアリーはアイスクリーム屋さんが公園にいると考えているとジョンは思っている」というように，「考えている」，「…したい」といった信念や願望などの心的状態が入れ子構造であらわれる思考を指す[20]。

18）立田幸代子，2005，「「心の理論」の獲得過程と象徴遊びの発展について－幼児と自閉症児の比較分析－」，『立命館人間科学研究』，第8号。
19）Premack, D., & Woodruff, G., 1978, “Does the Chimpanzee Have a Theory of Mind ?,” *Behavioral and Brain Sciences, 1 (4)*, 515-526.
20）林　創，2001，「「心の理論」の二次的信念に関わる再帰的な心的状態の理解とその機能」，『京都大学大学院教育学研究科紀要』，47：330-342

問題は，「再帰的認識」が，「～が～と考えている～ということを～が考えている～ということを彼は考えているのだろう」と再帰のレベルが高次になっていくことであるが，この点と関連して言えば，ロボットについての「再帰的認識」は一般にどのような構造をもっているかについて考えることが重要なポイントになる。これは，ロボットがペットやコミュニケーションの相手として認識される場面では，どのような「再帰的認識」（内容，レベル）が働いているかがHRI，人間とロボットの交流の意味を考える上で重要だということである。この入れ子細工の構造の次元の違いが意外と深刻な問題をもたらすかもしれないという点をわれわれは考慮にいれる必要があるだろう。少なくともそれはロボット倫理の研究対象になりうる（すべき）問題である。この点について考えるため，再び，林の論文を引用する。

> 被験者が「ジョンはアイスクリーム屋さんが公園にいると考えている」と表象できるかどうかは「一次的信念（first-order belief）の理解」の問題で，被験者が「メアリーはアイスクリーム屋さんが公園にいると考えているとジョンは思っている」と表象できるかどうかは「二次的信念（second-order belief）の理解」の問題となる（この段階が再帰的な心的状態の理解に相当）。一般に，二次以上を高次（higher-order）の心的状態と呼ぶ。[21)]

自閉症や認知症の患者（そもそも患者と言っていいのかどうかわからないが），あるいは幼児とロボットのコミュニケーションや心的交流がHRI，社会的ロボット研究の一つの重要なトピックとなっているが，[22)] ロボット開発者，HRI研究者と「患者」，「幼児」の再帰的認識の次元が異なっていることで（多分），予想されないような状況がそこでは生じているのかもしれない。たとえば，幼児は，「ワカマルちゃんは～と考えている」という単純な心的意識の帰

21）林　創，2001，前掲論文，330頁。

属をロボットに対して行っているのか，あるいは，「ワカマルちゃんは～と考えている～と私が考えている（ことが楽しい）」という初期的なメタ認識のレベルで行動したり，判断したりしているのかもしれない。一方，ロボット研究者たちは，「「ワカマルちゃんは～と考えている」幼児がいるから，「ロボットには本当の心がある」のだ」とロボットに心を帰属させる理由として幼児の初期的な「心の理論」を使っているのかもしれない。あるいは，再帰的認識が HRI の核心にあるかもしれないのに，「ロボットと人間の心の交流」があったと，再帰的認識によって生じた状況を別の問題にすりかえてしまうのかもしれない。

また，再帰的認識はそれだけだと，あくまでも一人の人間の心の中で生じた孤立した現象であって，再帰的認識をまじめに考えていくと，それを他者がどう認識するかという問題が当然ここで生じてくるが，この場合，この「被験者」の「心の理論」に基づく「帰属」を観察している研究者の意識，「心」自体はどこにあると想定されているのだろうか。独我論を避け，共同主観的な論点に立って考えてみるならば，しかも，「心の理論」の図式に従って考えてみるならば，研究者，観察者の意識，判断，行為も，「観察者は～していると考えている～と観察者以外の誰かが考えている」という入れ子細工の過程の中に入っ

22）たとえば，産業技術総合研究所が行っているパロに関する研究。産業技術総合研究所のホームページでは，「2004 年 9 月 17 日発表」のアナウンスメントとしてパロに関する実験結果が発表されている。「世界一の癒し効果，アザラシ型ロボット「パロ」，いよいよ実用化」という「キャッチコピー」付きの PR で，以下はその内容の一部抜粋である。

～高齢者向け施設での，ロボット・セラピーの効果の評価に関しては，心理的効果，生理的効果，社会的効果についての評価を行った。心理的効果については，POMS（複数の項目のアンケート），フェイススケール（笑顔から泣き顔までの絵で気分を表現），GDS（うつ状態の評価方法）などの主観評価，生理的効果については，尿検査により，2 種類の尿中ホルモン（17-KS-S および 17-OHCS）の測定による評価，社会的効果については，ビデオ撮影により被験者のコミュニケーション量の評価と介護者からのコメントによって評価を行った。これにより，パロとの触れ合いによって，心理的には，気分が向上したり【図 3 参照】，活気が出たり，「うつ」の改善効果があった【図 4 参照】。生理的には，ストレスが低減した【図 5 参照】。社会的には，高齢者同士および介護者との会話が活発になり，雰囲気が明るくなった。～

（http://www.aist.go.jp/aist_j/press_release/pr2004/pr20040917_2/pr20040917_2.html）（2009 年 8 月 30 日アクセス）

てこざるをえない。

観察者の対象の被験者，被験者が「心の理論」によってある事柄を帰属させている「ある人物」，それと，それ自体観察の対象となるべき観察者自身は，「心の理論」が観察者の勝手な妄想でないかぎり，権利的には平等である。もしすべての現象が，「心の理論」の適用も含めて，独我論的に，観察者，研究者の「心」の内部で生じているのだとするならば，この孤立した主観の想像はずいぶん手の込み入った想像である。

こうした問題点に加えて，「心の理論」と「心」とは別のものだということも（忘れられがちだが）重要な問題点である。この点に関しては，長井志江と浅田稔の説明がわかりやすい。

> 人間と共生し日常的に活動するロボットにとって，重要な機能の一つにコミュニケーションの能力が挙げられる。一方で，人間は他者とのコミュニケーションを実現する手段として，「心の理論」をもつことが知られている。心の理論とは，自己や他者の行動をその人がもつと予測される内的表象（知識や信念，欲求など）に帰属させて考える仕組みであり，この理論をもつことで他者の行動予測やふりの理解などが可能になる。[23)]

つまり，「心の理論」とは，他者，相手が行った行動を相手の立場（心，気持ち，意図）に立って考えるということであり，「心の理論」をもつことがそのまま「心をもつ」ことではないわけである。実際，「心の理論」の未発達な子供でも，「心」は当然ある。ところが，「ロボットの心」に関する研究の中では，しばしば，「心の理論」をもつことが「心をもつこと」だとされてしまう。あるいは，「帰属」が「帰属の際のてがかり獲得（利用）」の問題におきかえら

23）長井志江，浅田稔，2001，「「心の理論」に基づくヒューマン－ロボットコミュニケーション－共有注意のための発達的モデル－」
（Proceedings of the 19th Annual Conference of the Robotics Society of Japan, pp. 117-118, September 2001.）

れたりして，「心の理論」の混乱が生じる。鳴海と今井の研究[24]は，ロボットから発せられた人工的言葉で人間の反応が影響されるという，それ自体は興味深い研究だが，しかし，この研究から「ロボットに心がある」とか「ロボットと人間の間でコミュニケーションが成り立っている」とかいう結論を引きだせるわけではない。鳴海と今井が行った研究のポイントは以下の通りである。

［仮説］

ロボットとの対話によって「ロボットと共感した」と感じられる場合，実験群の被験者はロボットとのコミュニケーションを作為的にとらえず，コミュニケーション自体に集中できる。

［予測１］

いくつかの演出を行った後，ロボットがお菓子を差し出し「食べてみてね」と発話する演出を行う。実験群の被験者がそれまでのロボットの感覚表現の発話によって「ロボットと共感した」と感じられる場合，ロボットから渡されたお菓子を実際に食べる確率は，情報表現の発話を行った対照群の被験者と比較して増加する。

この実験から実権者たちが，引きだそうとする結果は，「人間は感覚器官を用いて実世界の物体を認識すると同時に，無意識のうちにその物体に対する感覚を得る。ここでロボットが同じ物体に関する感覚表現（例：おいしいね！）を発話すると，人間はロボットに対してマインドリーディングを行って本来感覚を持たないはずのロボットの感覚を無意識に察してしまう。この他者の感覚を読むという行動によって関係が形成される。」である。

これは「心の理論」を直接的に応用した実験ではないし（「心の理論」ということばは論文中では使われていない）が，「心の理論」モデルを使って「ロ

24）鳴海真里子，今井倫太，2003，「演出を用いたヒューマン・ロボットインタラクション」，『情報処理学会研究報告』，2003（100）pp. 67-74

ボットに心を持たせようとする」タイプの実験ときわめて近い位置にある。実際，この実験では，「被験者がロボットに感情を帰属させる」結果，ロボットと人間の間に「関係が形成される」と結論づけられているのである。

長井らの論文では誤解を招くような表現がもっと直接的なかたちで使われている。

> （心の理論とは＝筆者注＝）他者の信念や欲求といった内的表象を推測し，それらを統合的に関係づけて他者を理解する（理論である）。心の理論をロボットに実装することによって，人間とのコミュニケーションの実現を目指した研究がいくつか行われている。小嶋は，心の理論を発達的に獲得する手法として，共有注意と模倣を出発点とするモデルを提案したが，具体的実現にはまだ至っていない[25)]。

このように表現されると，「ロボットも心（心の理論）を持ちうる」と考える人が出てくるはずである。筆者自身には，この文はそのような意味をもつものであるように読めるのである。

ここでも，やはり，見えるべきもの（たとえば，ロボット研究者，ロボット倫理研究者の心，帰属のプロセス）が「不可視」なものになったり，「不可視」であるもの（帰属の水準のずれ）が「可視化」されたり（たとえば〈ロボットとの交流の過程で生じるとされる〉「癒し」が，血液中の化学物質の増減という〈目にみえる〉指標によって可視化される）するという状況が生じているのだが，この「見えたり見えなかったりする」人間の側の事情に注目しているロボット倫理研究者は，今のところ，ほとんどいないように思われる。

25）長井志江，浅田稔，2001，前掲論文。

4．結論・今後の課題

人間のような形をしたものが，倫理的対象になるということは（原理的に考えると不思議なことだが），経験的には理解しやすいことである。日本のマンガに氾濫する（たぶん，筆者はそうしたマンガには関心がなく，確かめたことがないので，「多分」）ポルノは，人間のような形をした線と面（マンガのキャラクターを純粋なマテリアルとして表現するとそうなる）が二次元上に登場するだけのものだが，倫理的問題の対象になるということは納得できる。

その意味で，日本でもロボット倫理研究は，やがてその必要性が認識される時が来るのかもしれないが，しかし，その時が実際到来する前に，われわれは，ロボット倫理，社会的ロボット研究のレビューも含めて，さらにもう少し詳しくヨーロッパや米国(一部，日本)の先行研究を眺める必要があるだろう。「なぜロボットは人間に近く，また，人間に遠い存在でなくてはならないのか，なぜそもそもロボットはそのような存在であることを求められているのか」などというロボット研究，ロボット倫理，社会的ロボット研究の核心に位置する疑問はまだよくわからないままだし，どのような「状況設定（研究の前提になるような，研究を動機づけるような)」があれば，日本でもロボット倫理，社会的ロボット研究がリアルになるのか，そうした点もまだよく見えてこないからである。こうした点は先行する研究内容を検討しながら考えてみる必要がある。

こうした作業と並行して，ロボット（倫理）研究における「見えないもの」をよりはっきりと見るようにこころがけておくことも必要だろう。少なくとも，問題を設定していた当初の時点では見えていたロボット（倫理）研究の特権的な立ち位置が隠されたまま，日本に持ち込まれたのではたまったものではない。その場合，われわれに求められるのは，たんなる「自律」と「責任」の帳じり合わせ的なシミュレーション作業だけになる可能性が大だからである。

5．文　　献

（参照文献に関しては，スペースの関係でここではとりあげない。本文中の脚注参照のこと。）