

美しさを作る・心を作る

ロボットについて哲学的に考えるということは幾つかの側面を持つが、ここでは、そのうちの二つに絞って話をすることにしよう。一つ目は、「ロボットの機能について考えることが人間の能力について考えることになる」という側面であり、もう一つは、与えられた紙幅の関係でごくわずかしか触れられないが、「ロボットの存在そのもの、もしくはロボットと人間の共存について考える」という側面である。ロボットの哲学が応用哲学の一つでもある理由は、前者においては「ロボットについての知見を人間の理解に生かす」という意味でロボットを人間の鏡とするからであり、また後者においては、「人間についての議論をロボットの状況に拡張する」という意味で、人間をロボットの鏡とするからである。

ただし、どちらの議論においても真に考察の対象とすべきであるのは、たんなる工場ロボットや探査ロボットではなく、人間と同じような心的機能をもつ自律的なロボットである。ここで、「自律的」というのは、少なくとも、任務の遂行において人間を含めた他の存在から「指令」を受けることがなくとも自分の判断でその任務を遂行できる、ということの意味

1

するが、この「自律性」の最終的な形態は、もちろん、自分の生存に関して人間と同じように「自分の意志」をもつということであろう。したがって、そのようなロボットがまだ実現されていない以上、この哲学的な議論の大部分は、いわゆる「思考実験」にならざるをえないという宿命をもつ。

さて、ロボットはふつう人間が「作る」ものだと考えられている。われわれの世界（現実世界）では、「何かを作る」とは何をすることなのだろうか。ちょっと考えてみると、「何かを作る」ことは、その「何か」がたとえいかに抽象的、一般的、関係的なものであろうと、物理的な物を組み合わせてその「何か」を作る以外にはない、ということが分かるであろう。〈美しさ〉を作るためには、美しい絵画や美しい庭園など、美しい〈何か〉を作る他はない。〈正しさ〉にしても、例えば正しい法律制度を作る必要があり、そのためには法律文書や人的組織や裁判所などを作らなければならない。では、ずばり、ソクラテスの「心」はどうだろうか。それを作るためには、何をどう作ればいいのか？ ソクラテスの心がソクラテスの存在や活動と不可分であることを考えると、ソクラテスその人に似た「もの」を作らずして、ソクラテスの心に似た「もの」を作ることができるだろうか？ われわれの世界が超自然的な奇跡を許さない限り、どのような人の「心」であれ、それを作るためにはそれを宿している「もの」、つまり人間を作らなければならない。しかもそれは、身体の損傷や病気に関するわれわれの経験からすれば、人間の手や足ではなく、脳（と、それにつながる中枢神経系）である。

2

人間も進化が作ったロボットだ

ここから、われわれは、かなり重要な教訓を得ることができる。いろいろと細かな議論を省略してそれを述べれば、「心や魂は何らかの物理的な物（物質）の組み合わせによって初めて存在する」ということである。物質を介さずに〈心そのもの〉をいきなり作ることができないからといって、そのことは必ずしも、「物質に依存しない心そのもの」（例えば、死後の魂のような）がこの世界に存在しえないということの証明にはならない。われわれに作ることができなくとも、存在することが不可能とは直ちに言えないからだ。しかし、人類が蓄積してきた経験や獲得してきた科学知識などのすべてと整合的に考えるなら、この世界において心がいかなる物質にも依存せずに単体で存在する可能性はない、と言ってよいであろう…他の可能世界においてはともかく。この主張を「素朴な物理主義」と呼ぶことにしよう[柴田 二〇〇一 一五—三〇]。この主張を心に関してもう少し正確に述べればこうなる（もちろん、完全に正確というわけではない）。

「心は、現実世界において、物質に依存しない個体としては存在せず、心のさまざまな性質が他の物質個体のもつ性質として存在する。」

もちろん、「素朴な物理主義」には敵がいるに決まっている。それを「素朴なメンタリズム」と呼ぶことにしよう。どちらも「素朴」という言い訳めいた限定がついているのは、それらが、存在論の正統的な主張というよりもわれわれの常識の（重要ではあるが）背景思想

だという意味である。つまりそれらは、哲学者のものであるよりは世間のものなのだ。その「素朴なメンタリズム」は「素朴な物理主義」を否定する。

「心は、現実世界において、物質に依存しない個体として存在し、心のさまざまな性質がその心的個体のもつ性質として存在する。」

ロボットの哲学が興味深いのは、われわれの常識がこれらの相容れない二つの背景思想・世界観の激突の上に危うく成り立っている、ということこそが露わにするからである。例えば、「ロボットは人間のような心を持つことができるか？」という問いを常識に投げかけてみよう。われわれの常識は、すべての現象の根底には物質的な基盤とその変化があるはずだという確信と、人の心は単なる物質の塊には還元されない崇高な存在だという確信の真二つに分かれるであろう。われわれは、この二つの相反する世界観を人の生死や治療の現場で、あるいは行為の選択や責任の判断において都合よく使い分けている。しかし、今後、様々な形でのエンハンスメント（能力増強）や人間のサイボーグ化の問題が差し迫った課題となり、ますます自分たちの身の回りにロボットが氾濫するようになれば、われわれの態度もいままでのように曖昧なままではいられなくなる。ロボットの哲学は、このような意味で、この二つの世界観が真っ向からぶつかりあうテーマであり、またそれに留まらず、未来のわれわれの世界観を示す道標ともなっているのである。

さて、ロボットは人間が作るものだとすれば、人間は誰が作ったのだろうか。進化の歴史が人間を作った、というのが素朴な物理主義の答えだ。もちろん、進化はローカル・ルール

の盲目的な積み重ねによって [ドーキンス 二〇〇九 三二一—三二六]、ロボット工学は意図的なトータル・デザインによって、というように作り方に大きな違いはある。しかし、どちらも自然法則を侵犯するような奇跡に頼ることもなく、自然界に見出される素材以外のいかなるものを用いることもなく〈心〉を作る、という点では同じく自然からの創造である。したがって、人間はこの意味で進化が作ったロボットだと言ってもよいであろう。

心を理解する二つの見方（機能とクオリア）

心を人工的に作ろうとするとき最も問題となるのは、われわれが心というものを二つの側面から理解しており、それがしばしば概念的混乱を引き起こすということである。その一つは機能であり、もう一つはクオリアだ。機能は、ふつう知覚や記憶の「働き」と言われるときに念頭におかれている、それらが果たす何らかの「因果的な役割」のことだと考えればよい。例えば、酸素を用いて糖をゆっくりと燃焼させることによってエネルギーを得る生物体では、ヘモグロビンが行う働きは体内の各組織への酸素の運搬である。つまりそれが、ヘモグロビンが体内で果たす機能である。ここには、素朴な物理主義にとって重要な論点がもう一つある。それは、エビ、カニ、イカ、タコのような節足もしくは軟体動物では、このヘモグロビンの機能をヘモシアニンという物質が代わりに果たすという点だ（そのおかげで、それらの「血液」は青みがかっている）。このように、マクロな観点で同一と見なされる〈機能〉は、それより下のレベルの複数の異なった種類の物質やプロセスによって同じように実

現される。これは、一つの機能が下位レベルの因果的メカニズムによって多重に実現されるという意味で、〈機能の多重実現〉と呼ばれる。〈飛ぶ〉という機能を考えてみよう。アゲハチョウは大きな羽根をゆっくり動かして飛ぶ。スズメは小さな羽根をせわしなく羽ばたかせて低空を飛び、タカは羽ばたかずに滑空して高い空を飛ぶ。飛行機は推力の作り出す揚力で飛び、ロケットは推力そのもので飛び、気球は空気に対する浮力で飛ぶ（タンポポの種は？もちろん風まかせで飛ぶ）。ロボットが心を持つことを可能としているのは、この機能の多重実現という原理である。つまり、心は、人間においても、まだ見ぬ宇宙人においても、そしてロボットにおいても、異なった素材と異なった因果的メカニズムによって多重に実現されるであろう。

しかし、多重実現をこのようにストレートに主張できるのは、心の機能に関してである。もう一つのクオリアはそう簡単にはいかない。機能が客観的に観察可能な三人称的現象であるのに対し、クオリア (qualia) は、「感覚質」とも訳されるように「感覚された限りでのその感じ」であって、根本的に、体験している当人にしか観察（接近）できない一人称的現象だ。ロボットにも心を持つということに最も深い疑念が生ずるのは、機械の内部にこのような一人称的な体験、つまり主観的な経験が生ずるとはとうてい思えない、という点である。例えば、味センサーの機械仕掛けがいかに複雑になろうとも、味を見分ける機構の内部に「桃の甘さ」や「ミカンの甘さ」の感覚が実際に生じて、その「感じ」をセンサーが報告しているということはないだろう。問題は機械仕掛けの複雑さではない、というように思わ

れるだろう。生物以外にそもそもこのような主観的経験は可能ではない、したがって電子工学のいかなる最高傑作であっても、そこにクオリアの経験が生ずる余地はない、というわけだ。

しかし、この疑念のポイントがあくまで「主観的経験がそこに生じている」ことへの疑いであるなら、実は、他人もロボットと同じ認識論的な位置にあることが分かる。他人に私と同じ意識や経験が生じているのかという認識論的な疑いは、いわゆる他我問題という古典的な哲学問題であり、ロボットのクオリア問題は本質的にこれと同じ論理構造を持っている。例えば、あなたが紺碧の海の青さに感動しているときに、傍らにいる親も兄弟も、もちろん恋人も友人も、あなたが実際にどんな青のクオリア（色感覚）を感じているのかは原理的に知りえない。どんなクオリアが他人に生じているのか、それどころかそもそもクオリアが生じているのかどうかさえ私には原理的に分からないのだから、それを根拠に、他人のクオリア経験に関する懐疑論を私が主張することもそれなりに理にかなっているし、他人はクオリア経験を一切欠いた「ゾンビ」（いわゆる「哲学的ゾンビ」）だ、と主張することも可能であろう。しかし、心の哲学におけるこの「ハード・プロブレム」[チャーマーズ 二〇〇六]に深入りするの控えて、ここでは、他我問題からの教えを確認するだけにしておこう。それは、見かけも機能も人間と区別がつかないロボットをあなたがゾンビとして扱うなら、あなたは他人もすべてゾンビとして扱わなければ筋が通らない、ということである。たとえ素材の違いがクオリアの有無を左右しているという可能性を排除できないとしても、何十年もロ

ットだと分からなかった友人にあなたがクオリアや意識を帰属させてきたということは、少なくとも懐疑論が問題となる文脈以外では、「クオリア」はすでにそのロボットのものだ、ということを示しているだろう。そして、懐疑論が問題となる文脈においては、「クオリア」を懐疑論者のほしのままにさせておけばいいのだ。なぜなら認識論的な問題としては、この懐疑は原理的に解けないからである。われわれは、この懐疑論に対して最終的には常に次のような存在論的主張を断固として優先させており、それにしたがって他人と共に生きている。つまり、われわれの住む現実世界は、他人であれロボットであれ宇宙人であれ、一定の条件を満たした認知機能主体にはクオリアが法則的に生じるような可能世界なのである。

ロボット制作が人間の認知について教えること

だが、感情はどうだろう。ロボットは感情をもちうるのだろうか？ ここでも感情という心的現象は、機能とクオリアの両方から理解される。そしてわれわれは通常、感情を「恐れ」や「喜び」や「悲しみ」といったクオリアとしてもっばら理解する。それゆえ感情が機能とは無縁な「認知の最終結果」だと見なされてきたことと相まって、感情は、ロボットが人間のような心を持つことができない最大の理由とされることが多い。ここにある混乱は、対応する機能が存在しないのにこれらのクオリアだけをどうやってロボットの中に生じさせればいいのか、という当惑の内によく現されている。

だが、感情には立派な機能がある。ダマシオによれば、感情は身体状態を刻々と脳に伝え

る、生命管理のための重要なメッセンジャーだ。それどころか、感情はわれわれの合理的な判断を支え、未来の出来事を予測する。ダマシオのソマティック・マーカー仮説[ダマシオ二〇〇〇]にここでは深入りしないが、感情が、認知を混乱させるだけの負の要因などではなく、身体と脳をもっぱら神経化学的ルートでつなぐ積極的な役割を担っていることは確かなことである。

例えば、感情機能の確かさを逆説的に示す現象がある。ダマシオが報告している一つの事例では、パーキンソン病を治療するための電極の挿入がたまたま女性患者に劇的な変化をもたらした。「患者は進行中の会話を唐突に止め、目を伏せて右にやり、ついで、ちょっと体を右に傾けた。そして、彼女の情動的表情は悲しみのそれになった。数秒後、彼女は突然泣き始めた。涙が流れ落ち、彼女の全体的な様子は深い苦悩のそれだった。」[ダマシオ 二〇〇五 九九]。ところが、医師が慌てて電流を止めると、とたんに彼女のすすり泣きもピタリと止まり、あっという間に彼女は陽気な状態に戻った。「一体あれは何だったの?」、とあとで彼女は医師に聞いたそうである。これは、脳幹内の微妙な箇所への電気刺激が意図せざる部位に流れ、いわば感情の表層的メカニズム、つまり悲しみのクオリアに対応する部分だけが発動し、感情の深層メカニズムの作動なしに患者が悲しみのクオリアを実際に経験した例である。彼女の悲しみの感じはウソではなかったが、その背後にはこのとき何もなかった。しかし、これが患者本人にとってさえ異常なことと思われたのは、それが、原因となる裏打ちのある本物の感情ではなかったからである。つまり、感情の本体は、そう思われるほどに

9

は感情のクオリアではなく、感情が果たす機能なのだ。

しかし、ロボットに話を戻せば、人間において感情が果たす機能はロボットにおいて何をなすのだろうか? これは妙な言い方である。むしろ、こう問うべきだろう。ロボットに人間と同じ感情は必要なのか? ここにおいてロボットの哲学は、ロボット制作という現場から最も重要なアイデアの一つを得ることになる。例えば、「 218×150 」という単純な計算問題をロボットの頭脳たるコンピュータにやらせるとき、それが古典的なプログラム型であるのか、それともニューラル・ネットワーク型であるのかにしがたがって、異なった計算メカニズムが働くことになる。つまり、「かけ算をする」という機能が多重に実現されるとき、それを実現する複数のメカニズムは、素材や条件や環境によって互いにまったく異なったものとなることがある。これは、ここで問題となる〈機能〉に、われわれの目的関心に沿った大ざっぱな特徴づけが与えられる場合にしばしば起こる。例えば、「子育てをする」という、きわめて「ゆるい」仕方で記述された機能が、様々な動物たちの間でどのような仕方で実現されているのかを考えてみてほしい。したがって、このことは、感情という機能のみならず機能一般に関して、人間とロボットの間に言えることである。つまり、同じ機能を果たしているからといって、同じ仕掛け、同じメカニズムが働いているとは限らないのだ。すでに生物進化のストーリーの中に、われわれはこの特徴を見出すことができる。

人体を「生きるための機能」の集合体と見るとき、それは余りにも「不完全」である。これは、ドーキンスによれば、「最初の誤りを製図板に戻って正しくやり直すより、むしろ事

後訂正というやり方で埋め合わせ」た結果であり、それは進化において不可避のことであった[ドーキンス 二〇〇九 五〇八]。なぜなら、進化が課す条件はあたかも、飛行機の製造技術者に白紙の製図板に戻ることを許さず、「祖先の」プロペラ・エンジンを用いて「子孫の」ジェット・エンジンを作らせるようなものであり、しかもその中間型がすべて飛べなければならず、なお悪いことに、どの一つもその先行型よりわずかな改良でなければならない、というようなものだからである。こうした条件下で「進化」させられたジェット・エンジンは、「あらゆる種類の歴史的な遺物、変則性、不完全さを背負い込んでいるだろう……」[ドーキンス 同]。

したがって、白紙の製図板から設計されたロボットが、人間と同じ認知機能を果たすために、人間と同じメカニズムを用いるとは限らない。いやむしろ、進化の事実や素材の違いを考えれば、そうでない場合の方が多いだろう。したがって、ある意味で残念ながら、人間と同じ認知機能を果たすロボットを作ったからといって、人間が用いるメカニズムを突き止めたことにはならないのである。

すると、感情の機能が生命管理というもっと大きな機能を果たすためのものであり、ロボットにおいてはそれが不要でないとか、あるいは人間とまったく異なるメカニズムによって果たされるということが判明する可能性もあるだろう。いずれにせよ、それは個別の機能の経験的な探究とロボット制作の現場から明らかになる問題であるが、われわれとしてはここで、一般的に、ロボットを人間の〈鏡〉として用いることの限界を心にとめておけば十分で

ある。機能の多重実現は、ロボットの実現方式と人間の実現方式が異なることを許すからである。

ロボットの倫理

最後に、ロボットの哲学においては「ロボットの倫理」という重大な問題が残されている、ということを中心に指摘しておくことにしよう。ロボットの倫理において最も有名なのは、アシモフの「ロボット三原則」である。

- 第一条 ロボットは人間に危害を加えてはならない。また、その危険を看過することによって、人間に危害を及ぼしてはならない。
- 第二条 ロボットは人間にあたえられた命令に服従しなければならない。ただし、あたえられた命令が、第一条に反する場合は、この限りでない。
- 第三条 ロボットは、前掲第一条および第二条に反するおそれのないかぎり、自己をまもらなければならない[アシモフ、二〇〇四、五]。

しかし、正確にはこれは倫理というものではなく、せいぜい、ロボットが人間のよりよき道具となるための「仕様書」にすぎない。例えば、かなりの自律性をもたせられた軍事ロボットが、味方を今始まる敵の攻撃から守るために、丘の上の家に潜む敵の小隊を壊滅するよう命令され、

ロケット砲を打つ寸前でその家に子供たちが囚われていることを知ったとき、彼はどうすべきだろうか。またそれと知りつつロケット砲を発射して子供たちも殺害してしまったとき、この軍事ロボットを使用し攻撃を命令した人間と、このロボット自身にはどのような責任が生じるのだろうか [Bekey, 2010]。こうした問いに、アシモフの三原則だけで十分に答えることはできないだろう。

現にあるロボットのようにもしロボットの自律性が低ければ、その分だけロボットが行ったことについて使用者や所有者の責任は重くなり、ロボットの自律性が高くなればそれだけ使用者や所有者の責任は軽くなるのかもしれない。極端な場合、われわれが想定するような「完全に自律的」なロボットにおいては、行為の責任は、大人の間と同じように考えられるのかもしれない。しかし、「まったくの道具」と「完全な自律性」の間で、ロボットとそれを使用する人間にどのような責任が発生するのかは、今後検討を要する重要な倫理的問題である。そもそもロボットに責任を負わせるということがどういう意味をもつのか、中間的な場合でさえ、まだ判然としないところが多い。それは、幼児や精神耗弱者や動物などに生じる責任と同じように扱うべきなのだろうか。中間的な自律性を持つロボットにわれわれが倫理的な態度を取るとき、ロボットがどのような倫理的な存在なのか明確にならなければならない。しかし、「完全な自律性」を持ったロボットという仮想的場合でさえも、「ロボットはわれわれと同じだ」と単純に言うことはできない。なぜなら倫理の基盤である、誕生（製造）・死（破壊）・認知能力・身体能力・外観・生殖（再生産）といった多くの基礎的な存在条件が人間とロボットの間で大きく異なるからである。

このような場合、本当にわれわれとロボットが同一の権利・義務をもって共生することは可能だろうか [Shibata, 2010]。

ロボットの倫理は、ある意味で人間の側の倫理的洞察をロボットという新しい存在、またそれらを含む新しい「人間・ロボット共同体」に応用する試みであるが、しかしまた同時に、それらの議論を通して「物理主義的世界観を受け入れた時代」における新しい倫理を模索する試みでもある。ロボットの倫理は、必ずしも愉快ではないかもしれない。しかし、必ず考えておかなければならないことなのである。

参考文献

- アシモフ、アイザック、二〇〇四『われはロボット』小尾英佐訳、早川書房。
石黒浩、二〇〇九『ロボットとは何か…人の心を映す鏡』講談社現代新書。
柴田正良 二〇〇一『ロボットの心』講談社現代新書。
柴田正良 二〇一〇「異世界の者たちの倫理」金沢大学哲学人間学研究会編『哲
学人間学論叢』創刊号一七一三七頁。
ダマシオ、アントニオ 二〇〇〇『生存する脳』田中三彦訳、講談社。
ダマシオ、アントニオ、二〇〇五『感じる脳』田中三彦訳、ダイヤモンド社。

チャーメーズ、デイヴィド 二〇〇六『意識する心…脳と精神の根本理論を求めて』林一
訳、白揚社。

月本洋 二〇〇二『ロボットのこころ—想像力をもつロボットをめざして』森北出版。

ドーキンス、リチャード 二〇〇九『進化の存在証明』垂水雄二訳、早川書房。

戸田山和久・服部裕幸・柴田正良・美濃正編 二〇〇一 『心の科学と哲学』昭和堂。

長滝祥二・柴田正良・美濃正編 二〇〇八『感情とクオリアの謎』昭和堂。

信原幸弘編 二〇〇四『シリーズ心の哲…ロボット篇』勁草書房。

Bekey, George A, Patrick Lin and Keith Abney 2010 “Ethical implications of intelligent robots,”
in Krichmar J. L. and H. Wagatsuma (eds.) 2010 *Neuromorphic and Brain-Based Robots:
Trends and Perspectives*, Cambridge U. P.

Shibata, Masayoshi 2010 “”Toward robot ethics through *the Ethics of Autism*,” in [Krichmar J.
L. and H. Wagatsuma (eds.) 2011