

## 0章 プロローグ-----サラの話

サラはその機械に入る前に一瞬ためらった。サラの生まれ育った街角のパンを焦がしたような懐かしい匂いや、人々の夕方のざわめき、そして何よりそこに残してきた娘の顔が心に浮かんできた。そこは地球だった。

…ネフィー、お前にもうすぐ会えるわ。もしこの機械がうまく作動しないとしたら、わたしは地球でよみがえることはない。しかし原理上は完璧なはずだ。わたしはこの機械の中で分子単位にまで分解され、身体の隅々まで何一つ漏れなくわたしの「設計図」は読みとられる。わたしはそのまま分子の煙となってここで消滅するが、地球に転送された「設計図」はその周囲の分子から<わたし>を再び創り上げる。身体に残る人生の痕跡ばかりか、記憶に刻まれた苦い思い出さえもまったく同一の、もう一人の完璧にわたしである<わたし>を…

もうサラがためらっている時間はなかった。地球からこんなにも遠いこの星の運命は、あと数分で尽きようとしていたのだ。窓の外では巨大なオーロラの牙が凄惨な美しさをなげかけ、獣の断末魔に似た悲鳴が辺りの空間を充満していた。サラは、ゆっくりと転送機の扉を開け、裸身のままその中に入っていった。

溶けずに残っていた意識の何かが、波のように退いていくまどろみの中からぼうっと姿を現し、それが自分になっていくのをベッドの上のサラはぼんやりと感じていた。

…わたしは、わたしだ。そこにネフィーがいて、すっかり年老いた番犬のモンクがいる。そして、わたしは、わたしは、サラ…

さて、もはやお馴染みとなったこの手のストーリーは、しかし、心というものに関するわれわれの二つの強い確信の対立、というかひょっとすると二つの根深い希望の対立をたやすく呼び出すことができる、という点でなお哲学者にとって重宝なものである。地球でよみがえったサラの心は、転送機の前でためらっていたサラの心と同じ一つの心なのだろうか？ われわれがこの話のサラに身をおいてみるなら、明らかに「同じだ」と言いたくなる。しかし、この話がサラに別人の記憶を与えるような結末であったとしたら、われわれはそう言うことに躊躇したであろう。というのも、その場合、サラは、ネフィーに対する何の特別の感情も自分の中に直ちに見出すことはないだろうからである。しかし、それとも、「そのような想定」は「このような想定」の下では不合理(!)なのだろうか。つまり、身体(したがって脳)の分子レベルでの再生が実現されたにもかかわらず、記憶が異なっているというような想定は。

しかしそもそも、いったい、サラは一度死んだのだがその完璧なレプリカが別の素材から作り上げられた、と言うべきなのか？ それとも同一のサラの本質がたまたま二度具体化されたにすぎない、と言うべきなのか？

もしサラは一度死んだとするなら、よみがえった現在のサラは以前のサラとは別人である。たとえ現在のサラがネフィーに対して以前とまったく同じ愛情を抱いているとしても、それは厳密には「まったく同じ愛」であることはできないだろう。なぜとって、「愛」が、愛する人と、愛される人と、その二人の関係から成り立つものなら、確かにこの二つの場合では、愛する人が別人だからである。つま先から髪の毛の一本に至るまで二人のサラはそっくりだとしても、サラであるということは本物のサラだけが満たすことのできる性質である。それゆえに、もしネフィーがサラその人の生存を求めていたのだとしたら、ネフィーの祈りは神に届かなかったことになる。たとえネフィーが、サラそっくりの別人の存在にどれほど

心をなごませようと。いやむしろ、「この人がお母さんその人であってくれさえすれば」とネフィーがどれほど悔しがろうと。

しかし他方、同一のサラがああ星とその後の地球とで二度にわたって存在したのなら、確かに二人のサラは同一人物である。たとえ二人のサラを形作る物質の流れには大きな断絶があるとしても、二人とも、サラをサラたらしめる何かを持っているはずである。幼い少女であったサラ、あるいは小説家となっていたかもしれないサラ、あるいは不幸にしてこの世には生まれなかったサラ、それらすべてのサラを一人の同一のサラとなすく何かがあるはずである。いったい、そのく何かはああの「設計図」の転送によって保存されたのだろうか。それともむしろ、ああの「設計図」こそがそのく何かなのだろうか。

## 0章の1 ロボットの心-----サラとは？ わたしとは？

本書のテーマは一言でいうと、ロボットに心がもてるか、ということである。この質問をいきなり大学生にすると、学生の大半はあまり迷いもせず「No」と答える。そこで、その理由は何か、とたたみ込んで尋ねると実にさまざまな答えが返ってくる。曰く、

「ロボットには計算ができて、人の気持ちは分かるはずがない」

「ロボットはプログラムされたこと以外のことをする創造性をもっていない」

「ロボットに意識はない」

「ロボットがセンサーをもっていたって、実際に何かを感じることはできないはず」

「ロボットが誰かを愛するなんて考えられないよ」

「心とは人間の本質だ、それをロボットがもったらそれはもう人間だ、だから定義によりロボットは心をもてない、証明おわり・・・」

「人間は自分たちの心を科学的に完全には解明していない、だとしたら、よく分かっていないものをどうやってロボットに作ってやることができるのか？」

「科学ですべてを理解できるというのは思い上がりだ、人間の心は科学の手の届かないところにある」

・・・

そこで彼らの言い分をひとしきり聞いた後で、「じゃ、ドラえもんには心がないわけ？」と反撃(?)すると、彼らは一様にのけぞって、「えっ、そりゃ、ずるいよ」といわんばかりの顔をする。要するに、鉄腕アトムや『スター・ウォーズ』のC3POや、『ブレードランナー』のアンドロイドや『ターミネーター 2』のシュワちゃん扮するロボットなどは、もしそれらがく本当に人工物であって、しかもく本当にくストーリー通りに振る舞ったとしたら、それらに心を認めないのはこれまた変だ、と学生たちも感じているわけである。

しかし、本当のところはどうなんだろう。現在、身の回りにはロボットはほとんどが流れ作業で活躍している産業用ロボットで、そのアームを振り回している「機械的な」姿からは、どんなに未熟な心であれ(あるいは、どんなに労働に疲れた心であれ?)彼らの心のようなものを想像することはできない。ごく最近話題になったある電機メーカー作成のロボット犬「アイボ」も、産業用ならぬ愛玩用ロボットである点では感情表現を豊かに示すことができるけれども、生まれだての赤ん坊ほどの心の実在性も感ずることはできないだろう。しかし、だからといって、ロボットが心をもてないと決まったわけではない。それは現在の技術的な限界であって、原理的な限界ではない、ということだってできる。むしろ、現在のロボット状況をたてにして永遠にロボットに心を拒むのは誤りだ、と主張する方が理屈としてはまともである。

では、本当のところ、ロボットが心をもつというのは原理的には可能なのだろうか。これからの話を面白くするために、私は、「可能だ」という陣営に身を投ずることにする。そしてそ

の上で、哲学者や科学者たちが「可能だ」「不可能だ」と主張する論拠を、少し丁寧に検討してみよう。そうすれば、「不可能だ」と答えた多くの学生たちだって、それなりに深い真実をつかんでいたことが分かるだろう。

ロボットが心をもてるか、とはどういう問いか

さて、目標ははっきりした。それは、「ロボットが心をもつことは可能か」という問いに答えることである。しかし、何をどんな風に示したらこの問いに答えたことになるのだろうか。ドラえもののビデオを示しても、産業用ロボットを分解しても、この問いに答えたことにならないのは分かり切ったことだ。この問いは一方では、神経科学や認知科学や人工知能研究などの成果に関わるものであるが、他方では、心や思考や感情や身体や人格といったわれわれの概念の境界域への冒険である。SF や哲学者の好む思考実験がそうであるように、それは例えば、心といった概念がそんな思いもよらない状況の中でもまだ信頼のおける内容を保っていられるのか、ということの探求である。

そこでまずは、われわれがロボットの心を考えるときにわれわれを暗に導いていると思われる、世界についての二つの大きな描像、というと大げさだが、要するに今日のわれわれの常識を形作る二つの大きな枠組みを押さえておこう。やがて明らかになるように、われわれの問題が悩ましいのは、この二つの世界観の激突する場所(あるいは曖昧に交じりあう場所)にそれが位置しているからだ。その二つの描像をかいま見るならば、われわれの問いにはどんな二次問題が含まれているのか、そして何を示したらそれに答えたことになるのか、ということがぼんやりとでも見えてくるだろう。そのためには、遠回りに見えるかもしれないが冒頭のサラの状況に戻ろう。いったい転送の前と後で二人のサラは同一のサラなのか？

しかし、サラの同一性の問いに答えるためには、ぜひとも知っておかなければならないことがある。それは、そもそもサラとは何者なのか、ということである。いや、「何物」なのか、と言うべきかもしれない。というのも、知りたいことはサラがどんな人物か、ということではなく、サラという存在とは何なのか、ということだからだ。それは、サラがどうなったときに存在し始め、どうなったときにその存在を終えるのか、と問うことに等しい。そしてサラの同一性の問いの効能はここにあるのだが、われわれが思いつく二つのそれらしい答えをよく眺めてみるなら、その二つとは、まさに今述べたばかりの二つの世界描像に他ならないことが分かる。それらは互いに対立しながらも、そのどちらもがわれわれにとって親しく重要であり、かといって、それら単独ではわれわれを満足させてくれそうもない。よく言えば、われわれはそれら二つの答えを場合によって使い分け、それらに応じて世界を棲み分けているともいえるが、思想的潔癖さ(?)を身上とするなら、それは間違いなくダブル・スタンダードである。それはたとえば、動物愛護団体の会合で動物の権利の根源性について感動的な熱弁を振った後で、行きつけのレストランのピフテキに舌鼓を打つような人物の「くだらかな良心」に少し似ているかもしれない。

## 0章の2 素朴な物理主義

さて、その答えの一つとは、サラはそれを構成している物質の集まりにほかならないというものである。物質の塊、それ以上でも以下でもない。だからサラは、現在のわれわれの常識からいえば、サラの両親の精子と卵子が結合した瞬間から以降のいずれかの時点から存在し始め、サラの体細胞のすべてが分解して「大いなる物質循環」に入り込むまでのいずれかの時点で存在を止めたことになるだろう。そして、サラの思考も、喜びも、苦痛も、希望も、すべてサラを造り上げている物質が造り上げた現象である。あるいはそれらは、サラを造り上げている物質の配置や過程がもつ「高次の性質」であるのかもしれない

い。しかしいずれにせよ、科学の成果と一体となった現在の世間的常識によれば、この宇宙の全存在を構成しているのは、遺伝子のような微細なサイズの化学物質から、裏庭で寝そべる猫のような中間サイズの柔らかな生き物、果ては星雲のような巨大でとらえどころのないものに至るまで、最終的には中性子や陽子や電子といった素粒子の複合物にほかならない。われわれ人間の精神の尊厳をただの素粒子群の位置や配置に還元してしまうこの〈素朴な物理主義 folk physicalism〉は、しかし、一見したほどバカげたものではない。それどころか、われわれがいまだ未知ではあるが切実に制御したいと思っている現象、例えばさまざまな遺伝病やエイズのような感染症に立ち向かうとき、われわれの探求を最底辺で支える希望の砦となっているのは、この〈素朴な物理主義〉なのである。たとえば大げさに聞こえるかもしれないが、しかし、それらの病いの発症や治療が物理化学的なメカニズムではなく祈禱師のまじないや教祖の霊力による、などと本気で信じていたとしたら、誰も、そうした気の遠くなるような探求の労苦をあえて引き受けようとは思わないであろう。

病気と治療というモデルにそって自分たち自身を考えるとときにはわれわれは喜んで〈素朴な物理主義〉を受け入れている、と言ってもいいだろう。たとえば、どんなに個性的な人物がどんなに個性的な病気に悩んでいたとしても、それが個性的な肝臓や細胞(?)どころか、個性的な原子や電子(??)を持っているせいだ、と考える人はいないはずだ。われわれの科学の概念的枠組みは、素粒子の一つ一つに「カズヤ」や「チカ」のような名前をつけてその個性を研究する、というようなものではないからである。われわれの科学、少なくとも自然科学は、むしろ個性的対象を、没個性的な多数の構成要素の組み合わせとして理解する。なぜそうなのか、ということには興味深い点があるのだが、今のわれわれにとっての要点は、その結果として、他の人間にとって有効であった治療法や薬が基本的には自分にも有効だと確信してかまわない、ということである。こんな場面でわれわれは、自分が他に比類なき独特の存在でありたいとは望まないだろう。それどころかわれわれは、不幸にして病気や事故で失った身体機能を取り戻すために、損傷した体の部分を丸ごと新たに分子レベルで複製して、それを不具合な部分と取り替えたいと願うかもしれない。事実、脳死患者からの臓器移植や、クローン技術が目指す胚性幹細胞からの臓器生成は、このような願いの部分的な実現である。

#### 物質ゆえの不死？

このように考えるなら、〈素朴な物理主義〉はわれわれにとって歓迎すべきものでもあろう。われわれの体細胞は新陳代謝によって日々新たなものに更新され、何ヶ月間の間にはわれわれを造り上げている物質は新しい一組にそっくり入れ替わってしまう。その昔にわれわれの一部であった分子の集団は、今は散り散りとなって、風にそよぐ緑の葉の一部となっているかもしれないし、大洋の中をさまようプランクトンの餌、あるいは土中のモグラの尻尾の一部となっているかもしれない。しかし、このようにわれわれを構成している物質が岩や水の場合ほどにも安定的にわれわれの内に留まりえないのだとしても、われわれの同一性が脅かされることはない。10年前のわたしは、そのような物質の交代に逆らい続けながら、あるいはそのような分子の流れに支えられながら、今日のわたしと同じわたしである。とするならばわれわれは、逆説的に聞こえるかもしれないが、〈物質の塊にほかならないがゆえの不死〉というものを手に入れたことになるだろう。というのも、わたしがわたしであるためには、厳密に同じ分子集団がわたしを構成し続ける必要はなく、同じ種類の分子たちが同じ結合の仕方をしていただけで十分だとするなら、この今のわたしがいずれ朽ち果てるとしても、今のわたしと同じ種類の分子たちを後でかき集め、それらを同じように組み合わせることによって、わたしはそこに文字通り余すところなく存在することになるからである。かくしてわたしは、いつでもどこでも時と場所を越えて、わたしを構成するだけの分子素材と技術がある限り存在しうることになる。つまり、この「存在しうる」とい

うのは、別の時代と場所に自分が生まれていたかもしれないという反事実的で空虚な(?)想定などではなく、資源と技術に依存してのこととはいえ、その時空領域でのまともな(?)物理的出来事としてわたしの再現はありうるということである。

### 0章の3 素朴なメンタリズム

しかし、この<素朴な物理主義>はそれほど歓迎すべきものだろうか。サラという存在とは何なのか、と先ほど問うたとき、喉にまで出かかっていたもう一つの答えが反撃する。それを<素朴なメンタリズム folk mentalism>と呼ぼう。一体、わたしが<物質の塊にほかならないがゆえの不死>を持っているならば、また同じ理由で<物質の塊にほかならないがゆえの任意の再現性>も持っているはずである。つまり条件さえ整えば、わたしの眼前で、文字通りのわたしが造り上げられ、このわたしと対面することがありうるのだ。しかも、そのようなわたしがたった一つしか造られえない、と考えるべき理由が何かあるだろうか？ ちょっとばかりタガの外れた科学者が、豪勢な打ち上げ花火のように同時に幾つものわたしを出現させたとしても、理解しがたいのはその科学者の意図の方であって、何人ものわたしが同時に一つのテーブルを挟んで向かい合っている事実の方ではないだろう。問題は、<素朴な物理主義>にしたがえば、新たに出てきているその幾人ものわたしが、すべてわたし本人であってわたしのまがい物やコピーなどではない、ということである。というのもそれらを本物のわたしのコピーと呼ぶとしても、年老いた本物のわたしだって、少年期のわたしから「自然な仕方」で造られたコピーにすぎないのであって、そこに本質的な相違があるとは思えないからである。少年期のわたしの「連続的で、ゆっくりとした、自然な」コピーは、現在のわたしの「断続的で、突然の、人工的な」コピーと、わたしであるという点で何が異なるのだろうか。これは確かに、広い庭園の中に落ちていどんなにそっくりの二枚の落ち葉でさえも同一の落ち葉ではありえない、と主張したライブニッツの形而上学的洞察に戦いを挑むような話ではない。<素朴な物理主義>も、この砂粒とあの砂粒は、それらを構成している分子の種類と数と結合の仕方が同一であるがゆえに一つの同じ砂粒なのだ、と主張しているわけではないからだ。しかし、それらが確かに一つの同じ個体でなくとも、二つの個体を構成している分子の種類と数と結合の仕方が同一であるならなおさらに、「これとあれは同一の何かだ」、たとえば「同一のチョコレートだ」、と言うことには意味があるだろう。同じ菓子メーカーの、あるいは同じ工場の、あるいは同じ新シリーズの、あるいは同じミルク味の、といったうちのどれが意味されているにせよ、その二つはその意味で同一のチョコレートであったりなかつたりする。これは個体の同一性ではなく、種類の同一性である。<素朴なメンタリズム>がいだく疑念は、<素朴な物理主義>によって、わたしの同一性は個体の同一性ではなく、種類の同一性に成り下がってしまったのではないか、ということである。たとえわたしの同一性が、同一メーカーや同一工場のチョコレートとは比べものにならないほど厳しい基準、つまり「それを構成している分子の種類と数と結合の仕方の同一性」にまで高められたとしても、それはなおこの砂粒>という個体の同一性ではない。というのもわたしは、この今のわたしを構成している特定の分子集団とともに存在し消滅するのではなく、そのような分子集団の転変にもかかわらず、あるいは分子集団の流れの上を漂いながら生き延び、存在し続けるからだ。

だがしかし、わたしの同一性は実は<この>チョコレートの同一性ではなく、<このような>チョコレートの同一性なのだ、つまりもっと露骨に言えば、わたしとは個体的存在ではなく、類種的存在なのだ、ということを受け入れることができるだろうか。それは変だ、とわれわれの<素朴なメンタリズム>は語る。だって、そうではないか。わたしとは何よりも、このわたしという感じ、わたしという視点、ここにこうして座り、あたりを眺め、つい昨日の飲み過ぎたパーティを思い出し、胃のあたりにちょっとした不快感を残しながらあなたと喋っている、といった内面から感じられる意識の全体であろう。この内面から感じられる意識は、

不分明な仕方で現在の感覚や気分や感情や思考と溶け合いながら、わたしという唯一無二の存在を造り上げている、いや、そのような存在そのものだ、と思われるだろう。そのような意識であるわたしが原理的には二人ありうる、それどころか複数ありうる、というのはまさしく個、個体、個人という概念からして本末転倒ではないか。そもそも個体という概念は、わたしという意識的存在、記憶と感覚の広がりの中で唯一の点を占める視座、歴史の中で二度と繰り返しようのない特異性、世界という絵巻物がなだれ込んでくる消失点、つまり、そのようなものが二つとしてあるということの不可能性を表現するために編みだされた考えなのではないのか。とすると、個体というものに最も適格な存在があるとすれば、それは、木の葉や庭石やベンチといった物体ではなく、まさにこのわたしという意識的存在、このわたしという心であるほかはないだろう。とすればさらに、そのわたしがおよそ小石や木の葉ほどにも個体的でないというのは、＜素朴な物理主義＞が何か根本的な点を取り違えているからに相違ない。わたしが複数個存在することになるような「わたしの捉え方」は、それだけで誤りであることを証明されたようなものであろう。

#### 心ゆえの不死？

＜素朴なメンタリズム＞のこの異議申し立ては、＜素朴な物理主義＞とは別な意味だが、やはりわれわれ人間の一つの希望の表明である。それは、生まれつきの身体や治療の失敗というモデルによって、われわれに親しいものであることが分かる。完璧なボディ、非の打ち所のない身体がどのようなものであれ、それをまどって（正確には、まどうべく？）生まれてきたと豪語できるような人はまずいないだろう（いやもちろん、いくつかの例外らしき存在も頭をよぎるけれど）。だが、そんなことは大したことはない。完璧ではなくとも、そこそこに機能してくれるなら、自分の体にどんな文句のつけようがあるか。わたしはそれでも、このわたしの手足、顔かたちなのだ。時たま不満を感じるのは、たかだか、体育祭でみんなの注目を集められる選手になれなかったとか、自分より容姿の優れたライバルに恋人の座を取られたとかいった、ごくありふれた人生の「冷厳な事実」に直面したときくらいであろう。しかし、程度問題が度をすぎれば本質的問題となるのもまた、われわれ人間に特有の状況である。「個性的な相違」というやや気楽な他人との比較も、ヘレン・ケラーのような生まれてすぐの深刻な身体障害を前にすれば、思わずたじろぐに違いない。また、事故や病気によって歩く機能や、しゃべる機能や、ものをつかむ機能が自分から永遠に奪われてしまうという事態を想像するとき、われわれは、自分の身体が今度は自分の存在そのものを幽閉する牢獄となったことに気づき、こう言いたくなるのではないか。わたしはもはや、このわたしの手足、顔かたちなのではなく、と。

それではわたしとは何なのか？ ＜素朴なメンタリズム＞によれば、それはわたしの心である。それは、わたしの身体にもぐり込むというよりは身体に漂着し、身体を所有するというよりは身体に絡め取られ、身体を操るというよりは身体に拘束され、ようやく死に至って身体から解放される何かである。それはこう主張する。わたしがこの手足、この顔かたち、この体であるのは、わたしが王家の滅びた時代に生まれ、銀行家の娘でも農家の跡取り息子でもなく、海岸の町で育ち友人の一人があなたであるように、わたしにとって全く偶然なことである。しかるに、わたしがわたしであるのは必然的であって、これらの偶然的な事柄をすべてわたしから剥ぎ取ったとしても残留する何かである。それは、他人がたいていは誤解し、わたし自身もが時として見誤るわたしの内面性、わたしの隠された本質、つまり一切の偶然性の衣を脱ぎ捨てた裸のわたしの心に他ならない、と。

この観点からすると、わたしとはわたしを構成する物質ではありえない。むしろ、わたしはまずわたしの心なのだから、わたしが存在するためには本質的には身体も不要であろう。試しに、わたしからわたしの身体の一部を徐々に取り除いていってみるがいい。髪や爪や皮膚の一部がわたしから取り除かれても、もちろん、なお依然としてわたしはわたしである。こんなことは日常茶飯事だ。しかしたとえ、指や腕や脚がなくなっても、わたしの身

体から何かが失われたのと同じ意味で>わたしの心から何か失われたと主張する人はいないだろう。こんな仕方ではわたしの心は分割されない。ではもっと重大な何か、そう、わたしの脳が取り去られたとしたらどうか？ だが、<素朴なメンタリズム>は、脳にさえも格別の地位を認めようとはしないだろう。要するに、心は身体・物質とは全く別の存在秩序に位置するというわけだ。でないとしたら、身体を通してやってくる、心にとっての一切のまがまがしき偶然性をいかにわれわれは理解できるというのか。純真な子供時代に彼らを襲う無慈悲な事故や病魔は、心である彼らの存在をいささかも傷つけることはできないはずだ。…かくして、<素朴なメンタリズム>が心に特有の<死>を認めるのでない限り、心は<不死>である。それはむしろ、同一の心としてあれやこれやの偶然的物質の塊、いくつもの身体を遍歴しながら<生きながらえる>のかもしれない。ちょうどサラの心が、たまたまあの惑星と地球とで、物質の流れとしては断絶しながら二度にわたってサラの身体として実現されたように…

#### 0章の4 再び、ロボットの心とは？

やれやれ、どうして哲学者の議論はいつもこんな風なのか。これでは、ロボットの心の問題に行きつくどころか、その前に山と積まれた屁理屈のおかげでわれわれはへこたれてしまうのではないか。だいいち、サラの同一性の問題も、またロボットの心の問題も、<素朴な物理主義>と<素朴なメンタリズム>を持ち出すことでかえってこんがらかってしまっただけなのではないか。いったい上の議論は、そのどちらが正しくて、だからサラの同一性がどうなるのかについて、何を示してくれたのか。そして、そもそものわれわれの問題について…

いや、まさしく、良質のとは言えないとしても、哲学者の気がかりとはこんなものである。言い訳めかしてもう少しつけ加えるなら、哲学の問いが、「異論の余地のない一組の前提」から「異論の余地のない推論」によってすっきりと導かれてくる、などというようなことはほとんどありえない。たいていは、一つの問題は、それに関連した多くの他の問題を巻きぞえにして、拡大しながら、螺旋状に展開されていく。そこでは例えば心という概念は、わたしの同一性として問題にされた後、再び螺旋階段を一巻き登ったところで身体との相互作用として問題にされ、さらに一巻き登ったところでロボットの心として問題にされるという具合である。しかも、前の主張を次々に修正しながら。だから、高校時代の私なら勢い込んで「これぞ弁証法！」とでも叫んだことだろうが、要するに、議論の筋道は舗装のない石ころだらけの悪路を行きつ戻りつすることだ、と諦めてもらわなければならない。

だが、それでもわれわれは少しは前進したのではないだろうか？

#### ロボットの心と二つの<主義>との関係

まず第一に、ロボットに心があるか、あるいはミミズには心があるか、と考えるときにわれわれを引っ張る二つの背反する力、<素朴な物理主義>と<素朴なメンタリズム>の正体がかなり分かりかけてきた。それらは、ミもフタもない言い方をすれば、「この世にあるのは物質だけなんだ」という主張と、「それだけじゃなく物質とは別に心もあるぞ」という主張である。それらは一見して、相互に矛盾しあう主張であるのに、われわれはそれを時と場合によって使い分けている。アジの生き作りを食べるときにはアジの心のことなんかこれっぽっちも考えないが、逆に、人間の死体は文字通りに心なき物体なんだから来るべきの食糧危機にはそれを活用しよう、などと提案する人もまれだろう。その使い分けが人生を曖昧にも、なめらかなにもしているのかもしれない。その当否はともかく、そうだとすると、もし<素朴な物理主義>と<素朴なメンタリズム>のいずれかの立場を純粋に貫くなら、かなり面倒な摩擦をあちこちで引き起こす「とんがった主張」をすることになるだろう。だから、

ロボットの心の問題に曖昧でなくすっきりとした解決を得たいと思うなら、一時的にもせよ、この「とんがった主張」を自分に引き受けなければならないように思われる。

第二に、ロボットが心をもつ可能性に関して、問題の二つの〈主義〉と具体的に絡ませれば何か言えそうである。もし(1-1)この可能性を認めて〈素朴な物理主義〉をとるなら、その人はずばり、心とは物理的なものによって、しかも脳とは違った物理的素材によっても実現可能だという主張にコミットしたことになる。しかし、(1-2)ロボットに心の可能性を認めて、なおかつ〈素朴なメンタリズム〉をとった場合、その人は、心がどこからロボットにやってくるのかを説明するのに少し苦しまなければならない。というのも無理のない前提として、われわれは、「物理的なものではありえない心そのもの」を自分たちで作り出すことができる、という話をこれまで聞いたことがないからだ。われわれはいつかそんな技術を持つようになれる、というのだろうか？ もしそんなことは不可能だというなら、心は、そのようなロボットを造ればなぜかそこに宿るということになるだろう。その場合、(1-2-1)心は偶然そこに宿るのか(神さまのイタズラによって？)、(1-2-2)常に必ずそこに宿るのか(自然法則的に？)、のいずれかのこととなる。前者の場合には、「われわれ」がロボットに心を作ったと言うのは僭越でもあろうし、後者の場合には、物理的な素材によって心が技術的に実現可能だという点では(1-1)と変わらない主張となるだろう。さて、(2-1)ロボットが心をもつ可能性を否定し、〈素朴な物理主義〉をとるなら、その人は、ヒトの脳とその進化的な(?)類似物だけが心を実現できるという主張にコミットしたことになる。それは、中世の錬金術師たちの期待も空しく「金である」という性質が原子番号 79 の元素 Au によってしか実現されえなかったように、「心をもつ」という性質も脳という物理的素材によってしか実現されえない、と主張することである。最後に、(2-2)ロボットが心をもつ可能性を否定し、しかも〈素朴なメンタリズム〉をとるとするなら、その人は、ヒトの脳とその進化的な(?)類似物には心が〈ある〉ことも〈ない〉こともあるが(神さまの気まぐれによって？)、ロボットに心があることは〈ありえない〉(再び、神さまの気まぐれによって?)という主張にコミットしたことになる。それは、この宇宙の物理的に正確な複製ができあがり、そこに当然、この地球の複製がおかれ、さらにあなたの正確な複製が今とまったく同じようにそこに住んだとしても、そのあなたのドッペルゲンガーは心をもたないことも〈ある〉、と主張するだけでなく、ロボットはそのようないかなる想定の下においても心をもつようなことは〈ありえない〉、と主張することである(しかしこの主張をする人は、神さまの出来心をいかにして知りえたのだろうか?)。

私はすでに、ロボットは心もちうるという立場に荷担している。さらに私は、「何者かがこの世界に1セットの物体の集まりと別に1セットの心の集まりをも造った」、という話をすみずみまで理解するにはそれこそ何セットもの追加的説明が必要になる、と考える方に傾いているので、結局、上の1-1の主張にコミットしたことになる。ところで、では、これまでわれわれの議論を引っ張ってきたサラの同一性の話はどうなったのか？ 実は、その話は、本書でのわれわれの問題よりも手強いかもしれない。ロボットの心の可能性がどのように決着しても、心の同一性の問題は片づいていないかもしれない。しかしたとえ片づかないとしても、ロボットの話によって、そこへいたる螺旋階段の必要な一巻きを登ったことにはなるだろう。肩すかしのようだが、この見通しもまた、われわれのなした最後ではあるがささやかな一歩なのである。

さて、抽象的な話はこれくらいでいいだろう。これからはもっと具体的な問題をネタに、もっと具体的な思弁を(いやはや、それでも?)展開することにしよう。それらは、コンピュータが考えているかどうかを判定するテストだったり(1, 2章)、小部屋に閉じこめられて未知の言語と格闘するあなただったり(3章)、常識問題に悩むロボットたちだったりする(4章)。肩の力を抜いて、秋の日の散歩のような気楽な気分で、電腦時代のパンドラの箱を

少し覗いてみよう。