『哲学』日本哲学会　第 70 号 (2019 年 4 月):51-68

人工知能と現代哲学

ハイデガー・ヨーナス・粘菌

森岡正博

＊【数字】の箇所で、印刷頁が変わります。数字はその箇所までの頁数です。

１　フレーム問題

本論文では、人工知能が知能を持つとは哲学的に何を意味するのかについて、これまでになされたいくつかの研究を紹介し、今後の議論のためのひとつの見取り図を提示する。

人工知能とは何かについての一般的な定義はない。マーガレット・ボーデンは近著『ＡＩ』（２０１６年）の中で、「人工一般知能 artificial general intelligence」の概念について語っており、それは推論や知覚を行なうことのできる一般化された力を持つだけでなく、それに加えて言語を操り、創造性を持ち、感情を持つはずだとしている（１）。

これはジョン・サールが１９８０年に「強いＡＩstrong AI」と呼んだものに近い。サールによれば、「弱いＡＩ」とはあたかも賢く考えているかのようにふるまえるコンピュータであり、「強いＡＩ」とは本当に人間のように考えているコンピュータのことである。サールは次のように書いている。「強いＡＩの考え方によれば、・・・・適切にプログラムされたコンピュータは正しい意味で精神 mind なのであ

り、言い換えれば、正しくプログラムされたコンピュータはまさに理解力とその他の認知的状態を持っていると言ってもかまわないのである」(2)。「強いＡＩ」については、２０世紀にさかんに論じ【51】られた。しかしながら「正しい意味で精神である」と言えるためには、様々な難問を解決しなければならないことが明らかになった。もちろん技術的な難関もあるが、哲学的に見たときに最大の課題となったのがいわゆる「フレーム問題 frame problem」である。

フレーム問題とは、人工知能が問題解決を行なおうとするときに、何が自身にとって重要なファクターで、何が自身にとって無視してもよいファクターであるのかを、自分自身で自律的に判断することができないという問題である。これは人工知能型ロボットを現実世界で実際に動かそうとするときに直面してしまう難問である。１９６９年にジョン・マッカーシーとパトリック・ヘイズによって提唱された。その原因を単なる技術的な面だけに還元することはできず、問題の本質は哲学的な面にあると考えられている。ボーデンも２０１６年の段階で「悪名高いフレーム問題が解決されたとするのは非常にミスリーディングである」と書いており、フレーム問題は未解決であるというのが現在でも少なくない専門家の意見だと考えられる(3)。

ところで「フレーム問題」とはいったいどういう問題なのかについて、専門家のあいだに意見の一致があるとは言えない。しかし大きく捉えれば、人間なら誰でも知っている「暗黙知」をいかにして人工知能に覚えさせることができるのか、という点にかかわる難問だとみなしてよいだろう。たとえば給仕ロボットというものを考えてみる。そ

のロボットは、店内の客に食事を給仕するために作られた。ロボットは、給仕に必要な動作や知識をひととおり修得しなければならない。では、給仕ロボットが実際の店内できちんと機能するためには、いったいどのくらいの知識のセットを与えておく必要があるのだろうか。まず、グラスに水を入れすぎると溢れるという知識は必須である。そして、グラスの載ったトレーを動かすと、トレーの上に乗っているグラスも一緒に動くという知識も必須である。なぜなら、その知識がないと、使用済みのトレーとグラスを一度に片づけることができないからである。さらには、それに加えて、グラスを動かすとグラスの中の液体も一緒に動くということをもロボットに教え込まないといけない。しかしながら、トレーを動かしたときの摩擦熱で液体が蒸発する恐れはないという知識をロボットに覚えさせる必要はない。それは正しい知識ではあるが、給仕の仕事にとってまったく重要性のない知識だからである。

さらに考えてみれば、「テーブルの上のトレーを引くこと【52】で壁の色は変わらない」という当たり前の知識を書き込む必要はまったくないが、「夜になったら窓からの自然光がなくなるのでテーブルの上にある物の色調は変化する」という知識は書き込む必要があると思われる。あるいは、ロボットが何かの行為を行なうことによって、その場の状況自体が変化するわけだが、そのような自己言及的な状況変化が起きることをあらかじめ想定して知識のリストを作成することは必要なのだろうか。

このようなことを考えてみれば、給仕に際してロボットが覚えなければならない重要な知識のリストはほとんど無限にあることになるし、給仕に関連するけれどもロボットが覚える必要のない知識もまたたくさんあることになる。では、いったい誰がどのようにして重要性のあるリストを作成し、ロボットに覚え込ませるのか。そのようなことは、そもそも可能なのだろうか。

なぜこのような問題が起きるのかと言えば、給仕の途中でそれまでにない新しい事態に直面したときに、何に対してどのように対処することが自分にとって重要であり、何が重要ではないのかを、ロボットが自律的に判断することができず、適切な解決行動をとることができないからである。面白いのは、人間の場合、この種の問題があらかじめ回避されているように見える点である。高校生になれば、ほとんどの人間は喫茶店で給仕のアルバイトをすることができるはずだ。接客についての簡単な教育を施しておけば、その後の給仕の作業自体に大きな問題が生じることはないであろう。グラスを片づけるときには、とくに何も熟考することなく、液体の残りの入ったグラスを、トレーごと厨房にさっさと引き揚げてくるはずだ。たとえこれまでになかった新しい事態が生じたとしても、人間なら臨機応変に知恵をめぐらして解決行動を模索するだろう。

この点に関して、松原仁は１９９０年に次のような興味深い考察を行なっている。「この一般化フレーム問題は有限の情報処理能力しか持たない主体（人間やコンピュータはもちろんこれに含まれる）には決して完全解決（complete solution）はできないこと、それにもかか

わらず日頃人間はあまりフレーム問題に惑わされていないように見えること、の二点から、人工知能研究においてフレーム問題について考えなくてはいけないのは、人間はあたかもフレーム問題を解決しているかのように見えるのが多いのはなぜであるかという問題（これを疑似解決（quasi-solution）の問題と名付ける）で【53】ある」と言うのである(4)。これは、あらかじめ想定していなかった出来事が生起する可能性のあるオープンコンテクストの状況において、新しい出来事が生起したときに、なぜ人間の知能がそれに対応できるように見えるのかという問いでもある。

もちろん、人工知能がフレーム問題によって立ち往生しないような工夫をすることはできる。たとえば、人工知能が直面する状況を、あらかじめ人為的に狭めておけば、人工知能は想定していなかった出来事に直面することがないので、フレーム問題に対処する必要がなくなる。自動掃除ロボットなど多くの機械はそのように設計されているはずである。あるいは力業で無数の状況設定や知識を埋め込んでおき、例外がほとんど生じないようにしておくことも可能である。もしそのような例外に直面してロボットがフリーズしたとしても、人間の使用者がその結果を許してあげれば、実際上は問題にはならない。このようにして人工知能型のロボットを現実世界で実用することは可能である。しかしこれはフレーム問題が解決されたことを意味しない。既述したように、ボーデンが、フレーム問題が解決されたとするのはミスリーディングであると言うのは、この意味においてである。

ロボットと比較したとき、人間は、新しい状況が生じたときに、（１）行為を遂行するために必要な知識の完全なリストを持っていなくても、何かの暗黙知を利用しながら妥当な判断を行ない行為することができる【暗黙知】、（２）新しい事態に直面したとき、何に対してどのように対処することが自分にとって重要なのかを自分で適切に判断することができる【重要性判断】、（３）いくつかの可能な選択肢を暴力的に無視して行為することができる【無視】、という点において特徴があると言うことができるだろう。この三点に限るわけではないが、これらは人間に目立つ特徴である。人工知能はこれらに原理的に対処できないように見える。フレーム問題は、このことを含み込んで成立している難問である。

この哲学的問いに対して、人工知能研究と哲学の境界領域から注目すべき考察がなされてきている。本論文では二つのアプローチを紹介して吟味したい。ひとつはヒューバート・ドレイファスの「ハイデガー型人工知能」の問題提起である。もうひとつはハンス・ヨーナスの「代謝型有機体」の概念に基づいてなされた問題提起である。前者は「現存在」と「身体」に注目したものであり、現象学の現代的展開と関連している。後者は「生命」に注目したものであり、現代の生命の哲学や人工生命論と関連している。【54】

２　ハイデガー型人工知能

ドレイファスはハイデガー研究者であると同時に、米国における人工知能研究の初期から、人工知能に対する批判的考察を行なってきた哲学者である。『コンピュータには何ができないか』（１９７９年）な

どの著書で著名であるが、２００７年に刊行されたその集大成とも言える論文「なぜハイデガー型人工知能は失敗したのか、そしてどのようにその失敗を修復すれば人工知能はもっとハイデガー的になっていたのか」(5)をここでは取り上げて、その主張を検討してみたい。

ドレイファスもまた、人工知能がフレーム問題に直面してしまうのは、人工知能が、ある与えられた状況においてどういった事実を知ることが重要なのかを理解していないからだと考える。ものごとや出来事はつねに具体的な状況の中に置かれてはじめて意味を持つからである。

ところが、旧来の人工知能研究は、意味を欠いたものごとから成る世界に対して、主体たる精神 mind が価値を付与するというデカルト的なモデルで行なわれてきた。しかしこの方法では、人工知能は人間の知能に近づくことはできないし、フレーム問題を解決することもできないとドレイファスは言う。そこで彼はハイデガーの「眼前性（手近にあること） Vorhandenheit」と「手許性（手許にあること） Zuhandenheit」の区別(6)に着目する。たとえば目の前のハンマーは、デカルト的視線によって対象化されて現われれば眼前存在者であるが、そのハンマーが、手に握られて、釘を打って、ものを作るという一連の指示の連関に埋め込まれたありようで、すでに身近に出会っている存在として現われれば手許的存在者である。私が日常の生活世界においてすでに出会っている道具たちは、そういった指示連関の中に埋め込まれた手許的存在者として立ち現われている。ドレイファスは前者の「眼前性」のことを「presence-at-hand」、後者の「手許

性」のことを「readiness-to-hand」と表現する。旧来の人工知能に欠けているのは、この手許性の次元における理解能力である。私が世界の中に存在するときに、私はつねにすでにこの手許性の意味連関に包み込まれてあるという、人間ならば誰であっても了解できる事態を、旧来の人工知能は実装することができなかった。その次元が実装されていないから、旧来の人工知能は、何が自分にとって重要なのかを決めなければ【55】ならない状況で、判断根拠の無限後退を起こしてしまうのである(7)。

したがって、旧来の人工知能がフレーム問題を解決して、真の人工知能になるためには、人工知能が手許性の次元を実装した「ハイデガー型人工知能 Heideggerian AI」になる必要があるとドレイファスは考えるのである。そして彼は実際にその方向を目指したと思われるいくつかの研究を検討し、それらの試みはいずれもハイデガー型人工知能には達していないと断言する。

彼はまず、ロドニー・ブルックスの試みたロボットを考察する。ブルックスは、昆虫にも似た階層型で分散型のサブサンプション・アーキテクチャの開発者であり、その技術は掃除ロボットのルンバなどにも応用されている。ブルックスのロボットは、内的な世界モデルを参照して行動するのではなく、みずからのセンサー入力を絶えず参照することによって行動する。しかしドレイファスによれば、ブルックスのロボットは環境の固定的な諸性質に対して反応するのみであり、ロボットの置かれたコンテクストや、変化していく重要性に対しては反応しなかった。したがってフレーム問題は解決されなかった(8)。

彼は次に、フィル・アグリとデイヴィッド・チャップマンの作成したプログラム「ペンギ Pengi」を考察する。これは人間のプレイヤーのアバターとペンギンのキャラが氷の塊を投げ合うゲーム「ペンゴ Pengo」となった。アグリが説明するところによれば、このゲームのプログラミングにおいて、彼らはハイデガーの『存在と時間』を参照しつつその「手許性」に対応する「直示的志向性 deictic intentionality」の概念を導入し、その志向性は特定のオブジェクトを指し示すのではなく、エージェントとその環境の相互作用の経時的なパターンにおいてオブジェクトが果たすことのできる役割を指し示すようにした。これによって、このゲームは、エージェントが反応する対象を、純粋に「機能 function」として組み込むことができるようになった。彼らがこのように自覚的にハイデガーに接近を試みたことをドレイファスは評価する(9)。

しかしドレイファスは、アグリらに対しては批判的だ。たとえば私が部屋を出ようとしてドアに手をかけるとき、私はドアを単なるドアとして経験しているのではない。そうではなくて、私は外に出ようとする可能性に向けていわば推し動かされているのであり、私はドアを前にして（ここを通って出るようにという）一種の誘いかけ solicitation それ自体を経【56】験しているのである。アグリらの人工知能は、アフォーダンスによって行為者が引き込まれていくというこのような経験をプログラムしてはいなかった。したがってアグリらは、彼らが導入した諸機能や、エージェントにとっての状況の重要性を、ともに対象化してしまったのだとドレイファスは批判するのである。この意味で、アグリらの人工知能もまたフレーム問題を解決していない(10)。

ドレイファスは最後に、マイケル・ウィーラーの理論に言及する。ウィーラーは２００５年の著書『認知的世界を再構成する』(11)において、近年、人工知能研究に応用されてきている「身体化＝埋め込み的認知科学 embodied-embedded cognitive science」（身体化されて埋め込まれた認知についての科学）はある意味でハイデガー的であると言ってよいと書いている。しかしドレイファスは、ウィーラーもまた間違った場所を見ているのだと批判する。その要点は以下である。ウィーラーはハイデガー的と言ってはいるけれども、それはまだデカルト的なモデル、すなわち、外界のものごとが人工知能へと表象され、その表象 representation にもとづいて思考や問題解決が行なわれるというモデルに留まっている。外界のものごとが人工知能の内界に表象されて処理されるというモデルそのものが問題なのである。そこに留まっているかぎり、人間のような知能は捉えられない。ハイデガーは現存在のあり方を「世界内存在」（being-in-the-world）として捉えており、そこにはいかなる表象の介在もない。ドレイファスは言う。「・・・・世界内存在というあり方は、思考や問題解決よりもさらに基礎的である。それはまったく表象的ではない」(12)。

人間が問題解決を行なおうとしているとき、行為者とその道具のあいだの境界線は消滅している。私はすでに世界に住み込んでいるのであり、熟練した実践者にとっては、「我々は（世界から切り離された）精神なのではなく、世界とともにある存在者なのである」。もっとも基礎的な意味において、我々は「世界に染みこんだ問題解決者 absorbed copers」である。その世界に染みこんだ問題解決が私の内部で行なわれるのか、それとも世界の中で行なわれるのかをクリアー

に言うのは難しい。内部と外部の区別はそんなに簡単なことではないからである。このような、現存在が世界内存在しているというあり方に着目するのが本来のハイデガー型人工知能なのであり、「延長された精神」などという考え方は薄っぺらいものでしかないとドレイファスは言う(13)。

以上のように、ドレイファスがハイデガー型人工知能に求【57】める水準は非常に高いものである。その人工知能は現存在でなくてはならないし、世界内存在していなければならないというのがその基本線である。その段階にまで至らないうちは、人工知能を「ハイデガー的」と呼ぶのは間違っているし、その段階にまで至らない人工知能はフレーム問題をけっして解決することはできないとするのである。

人間の場合、身体化された問題解決能力によって、世界の重要な変化に対応するスキルは日々改善される。たとえば、ある部屋にいるときに、たいがいの変化を我々は無視することを学ぶが、もし部屋が暑くなりすぎたときには、窓がみずからを開けるように我々に誘いかけsolicit、その誘いかけに応じて我々が窓を開ける、というアフォーダンス的なやり方で問題解決が行なわれる。ドレイファスは人間においてフレーム問題が解決されている理由を、次のように書いている。
「一般的に言って、世界の中での我々の経験を例に取れば、現時点でのコンテクストに変化が起きたときには、過去においてその変化が実際に重大事に至ったと（思い出せる）ときに限って、我々はいつでもその変化に対処しようとする。そして変化が実に重大なものであると我々が感知するときは、我々はそれ以外のすべてを変化していないも

の（であるから無視してよい）とみなす。ただしそれには例外があって、我々の世界に対する親密さのアンテナがはたらいて、（上記の重大な変化以外にも）何かしら世界に（重要な）変化があったような気がするからその変化をチェックしないといけないというもの（については、それが本当に重要な変化なのかどうかを例外的にチェックするので）ある。このような仕組みで、局所的なフレーム問題が登場しないようになっているのである」(14)。 ここで「世界に対する親密さのアンテナ」と意訳したものは「our familiarity with the world」である。人間においては、このような世界との親密さの感覚がつねに意識のバックグラウンドで暗黙知的にはたらいていて、重要なものと無視していいものとの弁別が絶えず行なわれており、それがフレーム問題の登場を抑止しているというのである。

しかしまた、我々がコンテクストを（みずから）変えなければならない場合には、フレーム問題がふたたび立ち上がることになる。では、我々の問題解決にとってそれまで周縁的であったことが、コンテクストの変化によっていつ重要なものになったのかを我々はどうやって知覚するのだろうか。ドレイファスは、メルロポンティに言及しながら、そのような知覚はアフォーダンスの側からの「召喚 summons」によっ【58】てなされると言うのである(15)。要するに、我々にとって重要な変化が起きたときには、世界に住まう我々に対して世界のほうから誘いかけや召喚がなされて我々はそれに気づくことができる、というようなモデルを取らないかぎり、フレーム問題の解決は不可能だというのである。そして人工知能にそれができるためには、「欲求、欲望、快楽、苦痛、動き方、文化的背景などを伴った、我々の身

体に非常によく似た身体のモデルを、プログラムに書き込む必要があるだろう」とドレイファスは結論づける(16)。それができないかぎり、ハイデガー型人工知能は無理だろうとするのである(17)。

このように、ドレイファスの言うハイデガー型人工知能は、人間のような「身体」を持ち、その身体を現に内側から生きている必要があるということになりそうである。しかしこのような要求を、現在のシリコンチップと金属・樹脂等でできた人工知能型ロボットに課すことができるのだろうか。次節では、これとは異なった、生物学からのアプローチを見てみたい。

３　代謝型人工知能

人工知能がフレーム問題を解決するためには、それが身体を持つこと以前に、それが一種の生命体であることが必要なのではないかと考える論者たちがいる。生命体は、何かの困難に直面したときに、あらゆる方法を駆使してその困難を乗り越え、サバイバルしようとする。生命体にはそのような力が内発的に備わっている。生命体のサバイバルしようとする内発的な力こそが、フレーム問題の解決にとって根本的だというのである。

このアプローチに根源的な影響を与えたのが、ハイデガーの弟子でもあったハンス・ヨーナス（ヨナスとも表記する）の「代謝(18)」概念に基づく生物哲学である。彼は１９６６年に『生命という現象』を英語で刊行し、独自の生命哲学を打ち立てた。その後、内容を増補したドイツ語版『有機体と自由：哲学的生命論への手がかり』が１９７３

年に刊行された。現在、『生命の哲学』の表題で日本語訳が刊行されているが、これはドイツ語版からの訳出である。英語圏では依然として６６年の英語版が読み継がれている。【59】

ヨーナスは、原始地球において細胞膜を持った細菌が誕生したときに、そこに「自由」が誕生したと考える。その細胞は、細胞膜をとおして、その外側から内側へと栄養分を取り入れ、そして不要になった微少物質をその内側から外側へと排出する。このような細胞膜を介した微少物質の連続的な出し入れによって、細胞は生命を維持することができるのである。時間が経過すれば、細胞を形づくるところの物質はすべて入れ替わってしまう。しかしながら、細胞という生命体は、物質の入れ替わりとは別次元において、生命としての同一性を保持し続ける。ヨーナスはここに、生命の、物質次元からの解放を見る。生命という形式は、微少物質の入れ替わりという物質次元のできごとからいわば超越した次元で出現しており、この意味で物質次元から解放されているのである。この解放こそが、生命が手にした「自由」であるとヨーナスは考える。

しかしその他方において、生命は細胞膜を介した微少物質の入れ替わり・循環によって束縛されている。もし物質循環が止まってしまったなら、それを前提として成立していた生命もまた消滅してしまうだろう。生命はこの意味で物質循環に依存していると言える。ヨーナスは生命が持つこのような自由のことを「依存的自由 bedurftige Freiheit」と呼んでいる(19)。物質循環が断たれれば生命は終わりである。生命はつねに潜在的なリスクによってその生存を脅かされてい

るのである。生命は、物質循環を絶えず行ないながら自己を存続させることを宿命づけられている。ほんの少しでもその努力を怠れば、死に直面してしまう。生命は、つねに努力して自己存続を図らなければ死んでしまうはかない存在である。

ヨーナスがこのような思索を行なったとき、彼はけっして人工知能のことを考えていたわけではなかった。生命と自由をめぐるヨーナスの思索は、生物の哲学の領域で耳目を集めたにすぎず、そのサークルの外に影響を与えることもさほどなかった。ところがヨーナスの死後、人工知能研究の領域において、彼の哲学に対する注目が起きてくるのである。

『Artificial Intelligence』誌に２００８年に掲載されたトム・フレーゼとトム・ツィムカの重厚な論文「行為的産出型人工知能：生命と精神のシステム的組織化を探究する」は、フランシスコ・ヴァレラのエナクティブ主義（行為的産出主義）に立ちながら、ヨーナスの代謝の概念の方向に人工知能の将来を見据えようとするものである(20)。以下、彼らの議論を吟味する。【60】

まずフレーゼらは、フレーム問題を次のように理解する。すなわち、「世界の重要な諸特徴が、人間のデザイナーや観察者の視野の中でのみ重要なものとして立ち上がるのではなくて、当のシステムそれ自体の視野から見て重要なものとして実際に立ち上がるようにするには、人工システムをどうデザインしたらいいのか」という問題である。そしてドレイファスの論文を引きながら、フレーム問題はまだ解決され

ていないとする。そして現象学的哲学や理論生物学からの貢献が必要だと言う(21)。

近年、認知科学に身体性的転回 embodied turn が起きたが、しかしそれでもなお、人工知能に、自分自身にとっての重要な問題を自律的な方法で in an autonomous manner 認知させるやり方はまだ分かっていない。そこを突破しないかぎり、フレーム問題の根本解決はない(22)。センサーモーターの身体内在化だけでは不十分なのである(23)。そこで彼らが目を付けたのがヨーナスの生物哲学である。彼らはヨーナスの議論に言及しながら述べる。「外部の観察者の目から見て「目標指向的」ふるまいとして記述できるであろうようなものが存在していたとしても、そこから、研究対象のシステムそれ自体がそれらの目標を持っているという帰結はかならずしも導けない。なぜならそれらの目標は内部から生まれてきた内在的な intrinsic ものというよりも、外部から押しつけられた外在的な extrinsic ものだろうからである」(24)。もし人工知能ロボットがそれ自身の「目標」を持つとしたら、その「目標」は人工知能ロボットの内部から、自発的に spontaneously 生み出されたものでなくてはならない。それができるためには人工知能ロボットはどういった「身体」を持たねばならないか、というのが問われるべき問いであると彼らは言う(25)。

フレーゼらが依拠しているヴァレラのエナクティブ主義においては、認識はつねに行為とのカップリングにおいて成立する。この点をさらに突き詰めたのがヨーナスの代謝生命論である。ヨーナスの洞察を借りれば、次のことが言えるとフレーゼらは述べる。人工知能と生命体

とはどこが違うのだろうか。人工知能は「存在による存在 being by being」である。すなわち、人工知能は行動を行なうことはできるけれども、その行動は必ずしもみずからを存在させるために行なわれるのではない。これに対して、生命体は「行動による存在 being by doing」である。すなわち、生命体が存在するためには、生命体は細胞膜を介して微少物質を出し入れするという自己構成的な self-constituting 行動を絶えず行なわなくてはなら【61】ない。もし生命体がその行動をやめたとたんに、生命体は死んでしまう。この世から存在しなくなってしまうのである。生命体の存続には行動が必須であるが、人工知能の場合はそうではない。これが生命体と人工知能の決定的な違いであるとフレーゼらは述べる。まさにこの点こそが、ヨーナスが「依存的自由」という概念で言い表わそうとしたことであった。ヨーナスは１９６０年代までのサイバネティクスや一般システム論を念頭に置いた議論を行なっていたが、それをフレーゼらは現代に再生させたのである。

もちろん、代謝する人工知能を作るのは困難である。しかし人工知能がフレーム問題を解決できない根本問題は、人工知能が「行動による存在」という生命体的存在様式を持たない点にあるとフレーゼらは言うのである。たとえば、人工知能のスイッチを切って、その後にまたスイッチを入れたとしても、人工知能はなんの変わりもなく動き続けるだろう。しかしながら、生命体の場合は、いったん死んでしまえば、もう二度と動き続けることはない(26)。この、死んだらそれで終わりという切迫感こそが生命体の存在を特徴づけているものである。そして生命体がこのような切羽詰まった状況のうえではじめて成立す

るという点に、フレーム問題の解決の糸口があるはずだと彼らは言うのである。

言い換えれば、生命体が「危うく不安定 precarious」な条件下で自己を能動的に創出し維持するという点にこそ注目すべきだというのである(27)。フレーゼらはこのような存在様態をヴァレラにならって「構成的自律 constitutive autonomy」と呼ぶ(28)。構成的に自律したシステムは、みずからのアイデンティティと相互作用の領域を自己構成的に創出し、行動にともなう個別のアフォーダンスに従いながら、「それ自身にとっての「解くべき問題群 their own ‘problems to be solved’」」を構成する(29)。彼らは、構成的自律をなし得る人工知能について理論的考察を重ねる。その結果見えてくるのは、人工生命と人工知能との接合の可能性である。というのも、外界に対して行動し続けないとみずからの存在を維持できないという切迫感をもって生命体は存在しているのだが、現行の人工知能はそのような切迫感を持つことができそうにないのだから、人工知能にそのような切迫感を持たせるためには、人工知能と生命体を融合させることが必要だろうからである。

彼らはまず「細菌－ロボット共生体 microbe-robot ‘symbiosis’」の可能性を指摘する。たとえば、ロボットに埋め込んだ細菌の状態をロボットのコントローラーに反映できるよ【62】うにすれば、細菌の持つ自発的な動きをロボットの知能に内在させることができるかもしれない(30)。あるいはセルオートマトンの原理をロボットに組み込むことによって、〈産出原理それ自体は知性的ではないのだけれども帰結

を外から見たら知性的に見える〉というその特徴を利用することはできないかと言う(31)。これまでのところ、そのようなシステムはまだ誰も作っていないとしたうえで、フレーゼらは次のような見通しを語る。「志向性を持った行為者 intentional agency の生物学的ルーツについてのより良いセオリーを作り上げるためには、まず第一に、バクテリアレベルの知性についてより良く理解することが必要だ。生命の始まりそれ自体に立ちもどることによってのみ、志向性を持った行為者と認知についてのしっかりとしたセオリーを確立するチャンスが巡ってくるのである」(32)。

フレーム問題を解決し、内発性を持った人工知能を開発するためには、まずはバクテリアまで戻らなければならないというフレーゼらの提唱は刺激的であり、また理にかなっているように思われる。

マーガレット・ボーデンもまたヨーナスを引きながら代謝の重要性を指摘し、代謝は「コンピュータによってモデル化できるが、コンピュータによって実行することはできない」(33)ので、もし代謝が精神の必要条件であるならば、強いＡＩは不可能になるとしている。ヨーナスの代謝モデルは人工知能の最深の鍵かもしれない。

４　粘菌とバイオコンピュータ

ところで、バクテリアレベルにまで戻って知能について探究する試みは、すでに行なわれている。とくに、中垣俊之・小林亮らによって研究されてきた粘菌コンピュータは注目に値する。中垣らは、迷路上に飢餓状態の粘菌を膜のように広げたのち、任意の二点に餌を置くと、

粘菌はその二点を最短距離で結ぶ経路にぴったりと合わさるように自分自身の身体を変形させていくことを２０００年に発見した(34)。餌に届かない行き止まりの道に置かれた粘菌たちは、その行き止まり道から退いて、餌に続く本体の経路へと合流し、経路を一本化し、経路の断面を分厚くしていくのである。このようにして、粘菌は、自分自身である種の計算を行ない、実際に二点間の最短経路を割り出して、もっとも効率的な形へとみずからの姿を変えていくのである。飢餓状態に置かれた粘菌が行なう【63】このような行動にこそ、「危うく不安定」な状況でみずからの存在を維持しようとする生命体の根源的な姿が現われている。

中垣と小林が２０１１年に刊行した論文「原生生物粘菌による組合せ最適化法」において、粘菌のこのような生存行動は、粘菌自身による「計算」によってなされたものと考えられている(35)。すなわち、粘菌においては、生存のための行動が内発的かつ自発的に起動し、最適解を求める計算が行なわれ、実際に最適解に沿うように自分自身を変形させたのである。これはすでに生物計算機と呼ばれてしかるべきものであるし、フレーム問題はおそらく解決されているのではないだろうかと思われる。というのも、これらの粘菌にさらに新しい課題を与えて追い詰めたとしたら、彼らはそれに対応するべく戦略を練り直し、新たな最適解へとみずからを変形させていくに違いないからである。未知の環境変化に対して、みずからを変容させて創発的にどこまでも対応していくという能力が、粘菌には備わっているように思われる。

中垣らは粘菌のこのような動きを数学的にシミュレートするモデルを作成し（フィザルムソルバ）、その挙動を調べた。その結果、粘菌が行なう計算は完璧な回答を緻密に導くようなものではなく、「大ざっぱであるが素早く答えを導く」ような性質のものであることを見出した。このように、素早く近似解を出す計算機能は、「生物的な解法の注目すべき特徴である」と中垣らは書いている(36)。彼らはとくに触れていないが、素早く大ざっぱな計算でよしとするというこの点こそが、〈新しい環境に出会ったときに、何が自分にとって重要なかを素早く判断し、その他を強引に無視して行動できる〉という人間の知能を生み出しているものではないかとも考えられる。そしてこのような知能がフレーム問題の解決には必要なのであった。

小林もまた２０１５年に刊行された論文「生物に学ぶ自律分散制御」において、ロボットのほとんどは前もって想定された環境でのみきちんと動くが、動物は未知の環境との遭遇にも対応してタフに動くことができると指摘し、動物がフレーム問題を解決できていることを示唆している(37)。そしてこの動物は哺乳類や昆虫などにとどまらず、粘菌までをもふくむのである。小林によれば、昆虫や粘菌においては、身体の自律分散的な制御が行なわれている。すなわち、フレーム問題の解決のためには、中枢神経系に似た中心制御システムを開発するのではなく、むしろ身体全体に自律分散したシステ【64】ムを開発するほうがよいかもしれない。そして蛇のような移動ロボットにおいてその萌芽が見られるとし、「手応え制御」「陰陽制御」「階層的制御」の三つを合わせた「環境を友とする制御法の創生」を提言している(38)。

これらの研究は、人間の内発的で自発的なフレーム問題の解決が、人間の中枢神経系においてではなく、中枢神経系の支配を離れた身体各部の分散制御的なネットワークの次元において自動的になされている可能性を示唆している。しかし、ブルックスのサブサンプション・アーキテクチャではフレーム問題には届かないのであった。

フレーゼらの言う「細菌－ロボット共生体」はひとつの方向性を示している。彼らは、ロボットの体内に細菌のコロニーを埋め込むという発想であったが、その逆を考えてみるとどうなるだろうか。すなわち、細菌の細胞の内部に人工知能やロボットなどの人工物を埋め込むのである。それは極微ロボットや極微プロセッサーを埋め込む形でも可能であるし、人工的に構造化されたＤＮＡやＲＮＡの断片を埋め込む形でも可能である。あるいは、それらの極微人工物の群れと細菌を共生させるようなシステムを作り上げることによっても達成される。

粘菌を例に取ってみよう。極微ロボットや極微プロセッサーや人工核酸様物質などの人工物によって計算能力を増強させられた粘菌を作成することができたとする。このような人工増強粘菌は、みずからの内発性と自発性によって最短経路問題を解いて行動するだけではなく、さらに難しい課題を計算によって解決して行動することになるだろう。このとき、この人工増強粘菌は、自分自身にとっての課題を発見し、それを自律的に計算して解決するという能力を獲得しているはずである。そもそも粘菌はフレーム問題を解決しているのであるから、人工的に計算能力を増強した粘菌もまたフレーム問題を解決しているはずである。人工増強粘菌は一種のバイオコンピュータと言える。コ

ンピュータの領域に限定して言えば、バイオコンピュータこそがフレーム問題解決の鍵である。これが本論文の暫定的な結論である。

以上から見えてくるひとつの哲学的問いは、もし生命体の自律分散的な制御によってフレーム問題が解決されるのだとしたら、ドレイファスの言うような世界内存在を基盤とするハイデガー型人工知能が実現しなくてもフレーム問題は解決すると考えてよいのではないかという問題である。フレーム問題はシンボル操作を行なう中枢神経系の問題ではなく、代【65】謝型かつ自律分散的な制御系の問題なのかもしれない。ハイデガー型人工知能を提唱したドレイファスは、中枢神経系的な制御システムを前提していたであろうから、ドレイファスの見込みが間違っていた可能性もある。また昨今のディープラーニングの発展によってフレーム問題は解決するのではないかとの期待も出てきている。ただしディープラーニングの可能性の輪郭はまだクリアーではない。もちろんフレーム問題をどのような問題として捉えるかにも関わっている。これは哲学者が正面から取り組むべき問いである。

これまで見たように、フレーム問題、その解決のために構想されたハイデガー型人工知能、代謝型人工知能、そして粘菌による自律分散制御システム、その先にあるであろうバイオコンピュータによるフレーム問題の解決、これらを一本の糸で結んで見取り図を与える試みを本論文では行なった。現代哲学にとって刺激的な問題が多数内在していると考えられる。哲学研究者の読者にとっては、人工知能の限界を突破しようとするキーワードとして、ハイデガーとその弟子であったヨーナスの名前が登場したことに、大きな興味を惹かれるであろう。私

の専門ではない各科学分野における用語法等で不適切な記述があるかもしれず、その際はご指摘いただければ幸いである。

ところで、人工増強粘菌を作成する研究には大きなリスクがある。粘菌に高度な計算能力を与える可能性のある研究であるため、人工増強粘菌が暴走するのをあらかじめ防止する必要がある。もし環境中に放出されれば、甚大な被害を人間や生態系に及ぼす危険性があるから、カルタヘナ法にもとづいた物理的封じ込めの可能な施設で高いバイオセイフティーレベルで研究を行なわなければならない。そもそも計算能力を増強された粘菌がどのようなふるまいをするのか、まったく想定できない。高度な知能を獲得した人工増強粘菌が大量に増殖し、獲物を求めて地球上を覆い尽くすといったＳＦのようなことが起きないともかぎらない。また粘菌ではなく、毒性のある細菌に高度な計算能力を与えるような研究はそもそも許可されるべきではないとも考えられるだろう。

また、この種の研究は、細菌をターゲットとした人工物によるエンハンスメント研究と見ることもできる。したがって、これはエンハンスメントをめぐる生命倫理学と接合することになるだろう。

人工知能はすでにバイオテクノロジー研究を様々にサポートしているが、将来はこれとはまったく異なった形で人工知【66】能研究がバイオテクノロジーによる生命操作と直接的に結びつく可能性がある。それに備えた議論をいまのうちに行なっておく必要がある。いずれにせよ、人工知能研究、生物学、哲学の垣根は一段と引き下げられたと考えてよい。

参考文献

Boden, Margaret A. (2016). *AI: Its Nature and Future*. Oxford University Press.

Dreyfus, Hubert L. (2007). “Why Heideggerian AI Failed and How Fixing It Would Require Making It More Heideggerian.” *Philosophical Psychology*20(2):247-268.

Froese, Tom and Ziemke, Tom (2009). “Enactive artificial intelligence: Investigating the systemic organization of life and mind.” *Artificial Intelligence*173:466-500.

Heidegger, Martin (2006). *Sein und Zeit*. Max Niemeyer Verlag.

Jonas, Hans (1973, 1977). *Das Prinzip Leben*. Suhrkamp. (1973年の原題は*Organismus und Freiheit: Ansatze zu einer philosophischen Biologie*)。

Kiverstein, J. and Wheeler, M. (eds.) (2012). *Heidegger and Cognitive Science*. Palgrave Macmillan.

小林亮(2015)「生物に学ぶ自律分散制御：粘菌からロボットへ」『計測と制御』54(4):236-241.

松原仁(1990)「一般化フレーム問題の提唱」J・マッカーシー、P・J・ヘイズ『人工知能になぜ哲学が必要か』哲学書房, pp.175-245.

Nakagaki, T., Yamada, H., and Toth, A. (2000). “Path finding by tube morphogenesis in an amoeboid organism, *Nature* 407:470.

中垣俊之・小林亮(2011)「原生生物粘菌による組合せ最適化法：物理現象として見た行動知」『人工知能学会誌』26(5):482-493.

Searle, John R. (1980). “Minds, brains, and programs.” *The Behavioral and Brain Sciences* 3:417-457.

下西風澄(2015)「生命と意識の行為論：フランシスコ・ヴァレラのエナクティブ主義と現象学」『情報学研究』89:83-98.

Weber, A. and Varela, F. J. (2002). “Life After Kant: Natural Purposes and the Autopoietic Foundations of Biological Individuality.” *Phenomenology and the Cognitive Sciences* 1:97-125.

Wheeler, Michael (2005). *Reconstructing the Cognitive World: The Next Step*. The MIT Press.

Wheeler, Michael (2008). “Cognition in Context: Phenomenology, Situated Robotics and the Frame Problem.” *International Journal of Philosophical Studies* 16(3):323-349.

1) Boden, Margaret A. (2016), p.22.

2) Searl, John R. (1980), p.417.

3) Boden, Margaret A. (2016), p.55.

4) 松原仁(1990), p.179. 松原は、一般化フレーム問題は人工知能だけでなく人間においても解決されていないとする注目すべき【67】立場を取っている。この論文は必読文献である。

5) そもそもは Wheeler, Michael (2005)についてのレビュー論文であった。

6) 『存在と時間』第一八節など。

7) Dreyfus, Hubert L. (2007), pp.248, 251.

8) PP.249-250.

9) PP.251-252.

10) P.253.

11) Wheeler, Michael (2005).　なおウィーラーはドレイファスの批判に、Wheeler (2008)で反論している。これらのやりとりは Kiverstein and Wheeler (eds.) (2012)にも再録されていて重要であるが、紙幅の都合上、本論文では扱わない。

12) P.254.

13) P.255.

14) P.263. 原文は非常に簡潔なので、森岡が括弧内に言葉を補った。

15) P.264.

16) P.265.

17) ドレイファスは、ハイデガーとメルロポンティを参照した W・J・フリーマンの試みを評価しているが、しかしそれでもフレーム問題は解決していないと考えている。

18) ヨーナスは Metabolismus ではなく Stoffwechsel を使用している。S.17 など。

19) Jonas, Hans (1973), S.150.

20) ヴァレラのエナクティブ主義については本論文では扱わないが、検討されるべき重要先行研究である。当面は Weber and Varela (2002)、下西風澄(2015)などを参照のこと。

21) Froese, Tom, and Ziemke, Tom (2008), p.467.

22) P.470.

23) P.471.

24) P.472.

25) P.472.

26) P.485.

27) P.480.

28) P.480.

29) P.481.

30) P.492.

31) P.494.

32) P.495.

33) Boden (2016), pp.144-145.

34) Nakagaki, T. et al. (2000)

35) 中垣俊之・小林亮(2011), p.483.

36) P.491.

37) 小林(2015), p.236.

38) P.241.

【68】

* 科学研究費・森岡正博代表「「尊厳」と「意味」を二本柱とした生命の哲学・倫理学の基盤的研究」研究課題番号：17K02185 および蔵

田伸雄代表「「人生の意味」に関する分析実存主義的研究と応用倫理学への実装」研究課題番号：16H03337 の成果である。

雑誌『哲学』版にある誤植の訂正

p.51 「将来現われるであろう」→削除　（英語の読み間違い）

p.55 論文タイトル「なぜハイデガー型人工知能は失敗したのか、そしてその失敗を修復しようとすることが人工知能をどのようにしてさらにハイデガー的にしてしまうのか」→「なぜハイデガー型人工知能は失敗したのか、そしてどのようにその失敗を修復すれば人工知能はもっとハイデガー的になっていたのか」　（英語の読み間違い）

p.56, p.65 「サブサンクション」→「サブサンプション」　（タイプミス）

ウェブ版の追記　（２０１９年４月３０日）

本論文では字数制限のため、フランシスコ・ヴァレラの論文（Weber, A. and Varela, F. J. (2002). “Life After Kant: Natural Purposes and the Autopoietic Foundations of Biological Individuality.”）を検討することができなかったが、これは実に感動的な論文であるのでぜひ読んでみてほしい。ヴァレラの死後に発表されたもので、ヴァレラのオートポイエーシスとヨーナスの代謝型有機

体の概念は、同じものを別の角度から解明したものだとしている。生き物の生き物らしさはこの二つで記述されるシステムにあるのであり、認識側にあるのではないとする。生命の哲学への影響力は甚大であろう。