

強い AI

信原 幸弘

サールは人工知能の哲学の現代の古典と言ってよい著名な論文「心・脳・プログラム」(Searle 1980)において、強い AI の批判を印象的かつ詳細に繰り広げた。印象的だというのは、この論文とともに広く知られるようになった「中国語の部屋」の思考実験によって、この批判をじつに鮮やかに展開したからである。しかし、サールの批判には多くの反論がある。サール自身が論文「心・脳・プログラム」のなかですでにそれらの反論に応答を試みているが、それでも反論は後を絶たない。以下では、この論文においてサールが展開した批判を少し詳しく検討しながら、批判の本質を見極め、本当に強い AI が不可能なのかどうかを見極めていきたい。

1. 強い AI と弱い AI

強い AI とは何か。サールは強い AI を弱い AI との対比で導入する¹。彼によれば、弱い AI はコンピュータを心の研究において非常に強力な道具を提供するものと見る。たとえば、コンピュータは心に関する仮説を厳密に定式化し、テストする手段となる。それにたいして、強い AI はコンピュータをたんなる心の研究の道具ではなく、適切にプログラムされたコンピュータは心そのものであると見る。つまり、そのようなコンピュータは私たち人間と同じく文字通り心をもつのである。

コンピュータが文字通り心もちうると見るか、そうではなくたんなる心の研究の道具と見るかが、強い AI と弱い AI の違いである。サールは弱い AI には全面的に賛成するが、強い AI には逆に全面的に反対する。なぜ強い AI に反対するのだろうか。その理由をしばらく追ってみよう。

強い AI によれば、適切にプログラムされたコンピュータは文字通り心をもつ。したがって、たとえば、ロジャー・シャンクがストーリーの理解のために「スクリプト」を用いてプログラムしたコンピュータは、文字通りストーリーを理解する。シャンクはコンピュータにレストランでの食事についての話を理解させるために、空腹の客がレストランに入ってから料金を払って出ていくまでに起こる典型的な出来事(着席、注文、飲食、など)をまとめたレストラン・スクリプトを作成した。このスクリプトを用いてプログラムされたコンピュータは、レストランでの食事を描いたストーリーとそれについての質問が与

えられると、適切にその質問に答えることができた。とくに、ストーリーに明示的に書かれていないことがらについても、正しく答えることができた。

シャンクのコンピュータが行っていることは、ストーリーと質問をそれぞれ構成する記号系列をレストラン・スクリプトにもとづいて形式的に変形し、それによって回答を構成する記号系列を産み出すことである。強い AI は、このような形式的な記号操作によって適切な答えを産出できれば、コンピュータはストーリーを文字通り理解していると言ってよいのだと主張する。

サールが強い AI に反対するのは、まさにこの点である。彼は形式的な記号操作によって文字通りの理解が生じることを認めない。たとえそのような操作によって適切な回答を産出できたとしても、それだけでは理解が生じているとは言えないのである。その理由としてサールが考え出したのが有名な「中国語の部屋」の思考実験である。

いま、ジョンが一人で部屋に閉じ込められ、中国語の記号群を与えられたとする。ジョンは中国語をまったく理解できず、中国語の記号はたんなる意味不明の文字にすぎないが、その記号の形式（つまり字の形）は明確に識別できる。この中国語の記号群は、ストーリーに関係するスクリプトである。ジョンにはもうひとつの中国語の記号群を与えられ、それとともにこの第二の記号群を記号の形式にもとづいて最初の記号群に関係づけるひと組の規則が与えられる。この第二の記号群はストーリーであり、ひと組の規則はプログラムである。このひと組の規則は英語で書かれているため、ジョンはそれを理解でき、それゆえ中国語の記号をその形式にもとづいて他の中国語の記号に結びつけることができる。さらにジョンには第三の中国語の記号群と、英語で書かれたいくつかの指示が与えられる。第三の記号群は質問であり、それらの指示にもとづいて第三の記号群（質問）を第一の記号群（スクリプト）と第二の記号群（ストーリー）に形式的に関連づけ、ある別の中国語の記号群を産み出すことができる。この最後の記号群が回答である。ジョンはこうして中国語のストーリーにかんする中国語の質問にたいして、形式的な記号操作によって適切な中国語の回答を産出できるのである。

サールはこのように状況設定を行ったうえで、核心となる問い、すなわちここに本当の意味で理解が生じているのだろうかと問う。たしかにジョンは中国語のストーリーについての中国語の質問にたいして、中国語を理解できる人とまったく遜色のない適切な中国語の回答を行うことができる。しかし、ジョンが中国語を理解していないのは明らかである。彼が行っているのは中国語の記号をその形式にもとづいて操作することだけだ。彼にとって中国語の記号は意味不明のままであり、彼にできるのは中国語の記号をその形によって識別することだけである。サールはこのような理由により、中国語の部屋の思考実験において理解は生じていないと結論づけるのである。

しかし、本当に理解は生じていないのだろうか。中国語のストーリーにかんする中国語の質問に適切に回答できるのであれば、そこには中国語の理解が生じているのではないだろうか。次節では、この疑問を検討してみよう。

2 心と一人称性

中国語の部屋の思考実験におけるサールの結論、すなわちそこでは理解は生じていないということについては、さまざまな反論がなされており、そのいくつかにたいしてはサール自身が応答を行っている。そのような反論のなかでとくに核心をついていると思われるのは「システム説」である。

システム説によれば、ジョンはたしかに中国語のストーリーを理解していないが、彼はスクリプトやプログラムを含む全体のシステムの一部であり、そのシステム全体はストーリーを理解している。中国語の質問にたいして中国語の回答を行うのは、ジョンというよりも、ジョンをその一部として含むシステム全体である。理解はジョンではなく、彼を含むシステムに帰されるべきである。こうシステム説はサールの主張に反論する。

これにたいして、サールはシステムの要素をすべて一人の人間のうちに内面化してのように促すことで応答を試みる。すなわち、ジョンはスクリプトやプログラムをすべて記憶し、彼の頭のなかで中国語の形式的な記号操作を行って中国語の質問に中国語で回答するとしよう。しかし、そうしてみても、ジョンが中国を理解していないことは明らかである。彼は中国語の記号が何を表すのかを知らず、ただその形式にもとづいて操作を行っているにすぎない。それは、英語を理解できるジョンが英語の質問に英語で回答するときとはまったく異なる。英語のときは、彼は英語の記号が何を表しているのかを理解しており、その理解にもとづいて英語の回答を産み出すのである。

しかし、このサールの応答には、システム説から見れば、明らかな誤りがある。ジョンがシステムの諸要素を内面化して自分の頭のなかにシステムを構築したとき、中国語の記号の形式的操作を行うのはもはやジョンではなく、彼の頭のなかにあるシステムの一要素である。ジョンはシステムの諸要素を内面化するとき、システムの一要素である彼自身も内面化しなければならない。つまり、彼の頭のなかに小人ジョンとでも言うべきものを構築しなければならない。小人ジョンはたしかに中国語の記号の形式的操作を行っているだけであり、中国語を理解していない。しかし、小人ジョンを含むシステム全体（すなわちジョン）は中国語を理解しているかもしれない。私たちはこのようなジョンのあり方を自分で体験したことがないので、そのようなジョンが中国語を理解しているという感じ（クオリア）をもつのかどうかよくわからないが、それでもそのような一人称的な感じをもつ可能性はある。たとえば、私たちはコウモリであることを体験したことがないので、コウモリであることがどのようなことなのかという感じをコウモリがもつのかどうか確言できないが、それでもそのような感じをもつ可能性はあるだろうし、じっさい、私たちはおそらくコウモリがそのような感じをもつだろうと思う（cf. Nagel 1974）。小人ジョンが中国語の理解の一人称的な感じをもたないことは明らかだが、そうだとすると、ジョンがそのような感じをもたないことはけっして自明ではないのである。

理解が生じているかどうかの根拠を一人称的な理解の感じがあるかどうかにもとづいて

すれば、システム(つまりジョン)が中国語を理解しているかどうかは判然としないので、理解が生じているかどうかも判然としないということになる。サールが主張するように、理解が生じていないということには必ずしもならないが、理解が生じているということにも必ずしもならない。どちらの可能性も残されるということになる。しかし、システム説はそもそもそのような一人称的な感じに理解の存在根拠を求めようとはしない。それはシステムが記号の形式的操作を介してであれ、中国語の質問に中国語の適切な回答を行うことができるという理由で、システムに理解を認めるのである。つまり、チューリングテストに合格しさえすれば、理解は存在するとみなすのである。

サールはもちろん、チューリングテストのような行動主義的、操作主義的な考え方に反対する。質問に適切に回答できれば、たしかにチューリングテストに合格できることになるが、それだけでは理解が生じているとは言えない。理解のような心のあり方は行動のあり方だけでは決まらないのである。理解にとって重要なのは適切な回答がどのようにして産み出されるかということだ。たんなる形式的な記号操作によって適切な回答が産み出されたにすぎないのであれば、チューリングテストに合格しようとも、理解は生じていない。理解が生じるような回答生成プロセスが必要なのである。

サールがこのようにチューリングテストの考えを批判するとき、彼はもはや中国語の部屋の思考実験の場合とちがって、理解の有無を理解の一人称的な感じの有無によって判定しようとはしていない。むしろ回答を産み出すプロセスの性格を問題にする。そのプロセスがたんに記号の形式的性質にもとづくだけでなく、記号を実現する物質的素材の適切な因果的性質にもとづくかどうか重要なのである。理解は適切な因果的性質をもつ物質的素材によって回答が産み出される場合にのみ存在する。では、理解を生じさせるような物質的素材の適切な因果的性質とは何であろうか。これが結局のところ謎のままにとどまっているところが、サールの議論の最大の問題点であるが、ともかく彼の言う因果的性質について考察してみよう。

3. 心を生じさせる因果的性質

サールは私たちの脳がその神経活動によって心を生じさせることを認める。彼は形式的な記号操作によって心が生じることを認めないが、けっして心を物質から独立だとするような物心二元論者ではない。そうだとすれば、サールは、脳の神経活動をコンピュータにシミュレートさせれば、そのようなコンピュータが文字通り心をもつことを認めるのではないだろうか。人工ニューラルネットワークはそのほとんどがコンピュータで脳の神経活動をシミュレートしたタイプのものであり、現在流行のディープラーニング型のものもそうである。そうだとすれば、コンピュータがディープラーニングで心の働きを学習したとすれば、そのようなコンピュータは文字通り心の働きをもつことになるのではないだろうか。

脳シミュレータ説は、脳の神経活動をシミュレートしたコンピュータは文字通り心をもつと主張する。しかし、サールはそれをも否定する。サールに言わせれば、コンピュータで脳の神経活動をシミュレートしても、それはたんに神経活動の形式的性質をシミュレートしているにすぎず、心を生じさせるような因果的性質をシミュレートしてはいない。この点を確かめるために、サールは中国語の部屋の思考実験を少し変えたつぎのような状況を想像してほしいと言う。部屋のなかのジョンは、今度は記号を操作するのではなく、ひと組の入り組んだ送水管とそれらにつながるバルブを操作する。ジョンは中国語のストーリーと質問を受け取ると、英語で書かれたプログラムを参照してさまざまなバルブの開閉を行い、送水管のなかの水の流れを操作する。その結果、最後には送水管の末端から中国語の適切な回答が出てくる。

ジョンによる送水管とバルブの操作は脳の神経活動と形式的に同じ構造をもっている。したがって、それは脳の神経活動を形式的にシミュレートしたものである。しかし、そこに理解が生じていないことは、またしても明らかだとサールは言う。ジョンは中国語を理解していないし、ジョンと送水管やバルブを組み合わせたシステム全体も中国語を理解していない。それゆえ、中国語の理解は生じていないというわけである。しかし、このシステムが中国語を理解していないことは、けっして自明ではない。もとの中国語の部屋の思考実験においてシステムが中国語を理解していないことが自明ではないように、この思考実験においてもシステムが中国語を理解していないことは自明ではないのだ。ジョンが理解していないことが明らかでも、システムは理解している可能性がある。

しかし、サールはこの新たな思考実験において理解が生じていないことの原因として、たんに理解の欠如の自明性に訴えるだけではなく、それよりももっと重要な論点を出す。彼によれば、脳をシミュレートするコンピュータは脳の神経活動について誤ったことをシミュレートしている。それはニューロンの興奮やシナプスでの興奮伝達の形式的性質をシミュレートするだけで、理解を生じさせる因果的性質をシミュレートしていない。理解を生じさせるためには、脳の神経活動の形式的性質ではなく（あるいはそれだけではなく）、因果的性質をシミュレートしなければならないのだ。しかし、コンピュータにできるのは形式的性質をシミュレートすることだけである。記号の形式的性質であれ、神経活動の形式的性質であれ、ともかく形式的性質だけなのである。したがって、記号の形式的操作だけで理解が生じないように、神経状態の形式的操作でも理解は生じない。理解には脳の神経活動がもつ因果的性質が不可欠であり、シミュレートするなら、因果的性質を因果的性質としてシミュレートしなければならない。

しかし、理解にはなぜそのような因果的性質が不可欠なのだろうか。形式的性質だけでも、その性質にもとづく記号や神経状態の形式的操作によって、中国語の質問にたいする適切な中国語の回答を行うことができれば、中国語を理解できているのではないだろうか。なぜ適切な回答だけではなく、それを産み出す適切な因果的プロセスが必要なのだろうか。

サールはここで本物とたんなるシミュレーションの区別をもちだす。台風のコンピュータシミュレーションは本物の台風ではなく、たんなるシミュレーションにすぎない。それは本物の台風がもつ因果的性質を欠いている。台風のシミュレーションに出会ったとしても、風に吹き飛ばされるわけでも、雨にずぶぬれになるわけでもない。理解もそれと同じことだ。理解のコンピュータシミュレーションは本物の理解ではなく、理解のたんなるシミュレーションにすぎない。それは本物の理解がもつ因果的性質を欠いているのである。

しかし、本当にそうであろうか。台風のシミュレーションはたしかにたんなるシミュレーションにすぎないが、理解のシミュレーションは本物の理解だということはありえないのだろうか。つぎにこの点を考察しよう。そのために、本物とシミュレーションの区別を少し詳しく検討してみよう。

4. 本物とシミュレーション

サールは、理解は乳分泌や光合成などと同じく生物学的現象であり、生物の生化学的構造の因果的性質にもとづいて生じると言う。乳分泌や光合成のコンピュータシミュレーションを行っても、本物の乳や糖は生じない。乳や糖のたんなるシミュレーションが生じるだけである。それと同様に、脳の神経活動のコンピュータシミュレーションを行っても、本物の理解は生じない。理解のたんなるシミュレーションが生じるだけだというわけである。

しかし、なぜ本物の理解が生じないと言えるのだろうか。乳分泌や光合成のコンピュータシミュレーションの場合は、たしかに本物の乳や糖は生じないが、それは乳や糖のシミュレーションが本物の乳や糖とちがって飲んだり食したりして栄養になるようなものではないからである。乳や糖のシミュレーションは飲食や栄養を可能にする因果的性質を欠いている。それはたしかに飲食や栄養のコンピュータシミュレーションにつながり、それゆえ乳や糖と同じ形式的性質をもつかもしれない。乳や糖が乳分泌や光合成ともつ形式的関係および飲食や栄養ともつ形式的関係は、乳や糖のシミュレーションが乳分泌や光合成のシミュレーションともつ形式的関係および飲食や栄養のシミュレーションともつ形式的関係と同じである。しかし、乳や糖のシミュレーションは本物の乳や糖とちがって、本物の飲食や栄養と因果的関係をもつわけではない。もちろん、乳や糖のシミュレーションは本物の乳分泌や光合成とも因果的関係をもたないが、だからこそ本物の飲食や栄養と因果的関係をもたないのである。

しかし、理解のシミュレーションの場合は、どうであろうか。脳の神経活動のシミュレーションから産み出される中国語の回答のシミュレーションは、中国語の話者に理解可能であり、適切な回答として受けとめられる。それは中国語の話者が産み出す本物の回答と同じである。つまり、回答のシミュレーションは本物の回答と同じく、聞き手の理解・

受容と因果的関係をもつのである。それは聞き手の理解・受容とたんに形式的関係をもつだけではなく、因果的関係をもつ。そうであれば、回答のシミュレーションはたんなるシミュレーションではなく、本物の回答であると言ってよいのではないだろうか。回答のシミュレーションが本物の回答だとすれば、理解のシミュレーションも本物の理解であると言ってよいはずだ。

サールはもちろん、この結論に反対するだろう。彼に言わせれば、回答のシミュレーションが本物の回答と同じように質問者に理解・受容されるとしても、それはシミュレーションを行うものにとっては、ただの無意味な形式的記号列にすぎない。ジョンないしシステムは自分の産み出した回答を理解しておらず、その回答には構文論があるだけで、意味論がない。それは記号の形式のみによって識別され、形式的な記号操作によって産み出されただけの無意味な記号列である。与えられたストーリーや質問もそうである。それらは質問者にとっては有意義な記号列であるとしても、受け取ったジョンやシステムにとっては無意味な記号列にすぎない。ジョンやシステムは中国語の記号列を無意味な形式的記号列として受け取り、それらを形式的に操作して、無意味な形式的記号列として回答を出すにすぎないのだ。それゆえ、理解のシミュレーションはたんなるシミュレーションにすぎず、本物の理解ではないというわけである。

ここにはまたしても、ジョンやシステムが中国語の記号を理解していないという論点が登場する。しかし、すでに論じたように、ジョンが記号を理解していないことが明らかでも、システムが理解していないことはけっして明らかではない。システムが記号を理解している可能性は残されている。それゆえ、サールはシステムも記号を理解していないという論点に訴えることはできない。それに訴えるのはたんなる論点先取にすぎないのである。

理解のシミュレーションと乳分泌や光合成のシミュレーションのあいだには、決定的に重要な違いがあるように思われる。理解のシミュレーションの場合、そのシミュレーションが形式的な記号操作によるものであれ、脳の神経活動の形式的反映によるものであれ、産出される回答のシミュレーションは質問者に理解され、適切な回答として受容される。質問者にとっては、それはたんなるシミュレーションではなく、本物の回答なのである。それにたいして乳分泌や光合成のシミュレーションでは、産出される乳や糖のシミュレーションはそれを受け取る者にとって飲み食いできないし、栄養にもならない。それは明らかに本物の乳や糖ではなく、たんなるシミュレーションにすぎない。この違いは理解のシミュレーションを本物の理解とし、乳分泌や光合成のシミュレーションをたんなるシミュレーションとするのに十分ではないだろうか。

サールはおそらく、たとえ百歩譲って、回答のシミュレーションを本物の回答だと認めたとしても、理解のシミュレーションはなおたんなるシミュレーションにすぎないと主張するだろう。回答のシミュレーションはたしかに本物の回答と同じく質問者に理解・受容され、その意味でそれは本物の回答だと認めざるをえないかもしれない。しかし、理解

のシミュレーションにおけるプロセスは中国語の話者が中国語を理解するときに生じるプロセスとはまったく異なる。シミュレーションのほうはたんなる形式的なプロセスにすぎないのにはたいし、中国語の話者のほうは脳のなかで生じる神経生理学的なプロセスであり、生物学的現象である。ここには記号の形式的プロセスと物質の因果的プロセスの決定的に重要な違いがある。たとえそれらのプロセスからともに本物の回答が産み出されるとしても、形式的なプロセスは本物の理解を生じさせるものではなく、物質の因果的プロセスだけが本物の理解を生じさせるのである。

しかし、なぜ形式的なプロセスからは本物の理解が生じないのだろうか。それもまた本物の回答を産み出す以上、本物の理解を生じさせるのではないか。乳分泌や光合成のシミュレーションは乳や糖のたんなるシミュレーションしか産み出さないために、本物の乳分泌や光合成とは言えないが、もしそのシミュレーションが本物の乳や糖を産み出したとしたらどうであろうか。乳分泌や光合成のシミュレーションから乳や糖のシミュレーションが産み出され、そこからさらに3D プリンタによって本物の乳や糖が産み出されるとしよう。この3D プリンタの出力である乳や糖はもちろん飲食でき、栄養になる。そうだとすれば、結局、乳分泌や光合成のシミュレーションから本物の乳や糖が産み出されるのであるから、そのシミュレーションはたんなるシミュレーションではなく、本物の乳分泌や光合成だと言ってよいのではないだろうか。それは通常の生物学的現象としての乳分泌や光合成とは異なるが、乳分泌や光合成の新たな形態であり、それゆえあくまでも本物の乳分泌や光合成なのではないだろうか。

もちろん、この結論には、サールでなくても、多くの人が反感を覚えるであろう。乳分泌や光合成のシミュレーションが乳分泌や光合成の新たな形態だなどというのは、とんでもない話だ。それはある物質が水とまったく同じ現象的性質（氷点、密度、透明性などの私たちに観察可能な性質）を示せば、たとえ H_2O という分子構造でなくても、水だと言うのと同じようなものだ。その物質は水と表面上よく似ていても、あくまでも水とは異なる。水かどうかは分子構造によって決まるのであって、現象的性質によって決まるわけではない。これと同様に、本物の乳分泌や光合成であるかどうかはその基底にあるプロセスの性格によって決まるのであって、産出されるものの性格によって決まるわけではない。

しかし、そうであろうか。水の場合でさえ、水かどうかは分子構造によって決まるというのは、疑いの余地がある。たしかに私たちは現在、水かどうかを分子構造によって決定している。しかし、なぜ分子構造によって決めているのだろうか。分子構造によって物質を分類すべきだというのは、自明のことだろうか。そうではないように思われる。物質が分子構造によって分類されるのは、分子構造が物質の現象的性質を説明することができ、その現象的性質が物質の違いにとって決定的に重要だからである。もし分子構造と現象的性質のあいだにつながりがなく、現象的性質が同じでも分子構造は異なるとすれば、どうであろうか。どんな水も同じ現象的性質を示すが、ある水の分子構造は H_2O であり、

別の水の分子構造は XYZ であり、さらに別の水の分子構造は PQR であり、等々だったとしよう。このとき、これらの水はじつは違う物質だということになるのだろうか。そうではなく、それらは同じ物質であり、水かどうかを決めるのに分子構造は役立たないということになるのではないだろうか。つまり、同じ水でも、いろいろな分子構造があるということになるのであって、分子構造ごとにじつは異なる物質があるということにはならないのではないだろうか。現象的性質が同じでも分子構造が異なるとすれば、分子構造は物質の分類の決め手にはならないと思われるのである。

これにたいしては、パトナムの有名な双子地球の思考実験 (Putnam 1975) をもちだして、反論がなされるかもしれない。いま宇宙のかなたに地球と瓜二つの第二地球があるとしよう。この第二地球は何から何まで地球と本当にそっくりで、地球に太郎がいるとすれば、第二地球にも太郎にそっくりの第二太郎がいる。ただ第二地球は地球と一点だけ異なるところがある。それは地球の水とそっくりの物質が H_2O ではなく、XYZ という分子構造をしている点である。XYZ は H_2O と同じ現象的性質を示すが、その分子構造は異なる。ここでパトナムは XYZ と H_2O が異なる物質であることを自明のこととして思考実験を続ける。いま、近代化学が誕生する以前の地球と第二地球を考えよう。そうすると、太郎が H_2O について経験することと第二太郎が XYZ について経験することはまったく同じであり、それゆえ彼らの脳のあり方もまったく同じであるが、太郎が「水は透明だ」と言うとき、彼は H_2O について述べており、それゆえ「水」で H_2O を意味しているのにたいし、第二太郎が同じ発言をするとき、彼は XYZ について述べており、それゆえ「水」で XYZ を意味している。したがって、彼らは同じ脳のあり方をしているにもかかわらず、「水」という言葉で異なることを意味しているのだ。ここからパトナムは、意味は頭のなかにあるわけではないと結論づける。パトナムのこの思考実験について、その結論にはいろいろ異論が出されるが、 H_2O と XYZ が異なる物質だとしたことにはとくに異論が出されることはない。じっさい、それはまったく自然な想定であり、とくに問題にする必要はないようにみえる。

しかし、その想定が自然だと言えるのは、けっして H_2O と XYZ がまったく同じ現象的性質を示すのに、分子構造は異なるという理由からではない。近代化学が誕生するまえは、たしかに太郎が H_2O について経験することと第二太郎が XYZ について経験することは同じであった。つまり、彼らに観察された H_2O と XYZ の現象的性質は同じであった。しかし、それは H_2O と XYZ のすべての現象的性質が同じだということではない。彼らに観察された限りでの現象的性質が同じだというにすぎない。じっさい、近代化学が誕生した後では、太郎がたとえば H_2O の電気分解を行い、第二太郎が XYZ の電気分解を行えば、彼らはただちに異なる現象的性質を観察することになるであろう。 H_2O と XYZ が異なる物質だと想定するのが自然なのは、それらの分子構造が異なるからではなく、結局、現象的性質が異なるからである。分子構造が重要とされるのはあくまでも現象的性質の違いをもたらすからであって、そうでなければ、分子構造の違いは物質の分類にとってた

いした役割を果たさないであろう。

本物とたんなるシミュレーションの区別も同様である。そこでも、本物であることにとって決定的に重要なのは現象的性質であって、その基盤の性格ではない。基盤は現象的性質をもたらす限りにおいて重要であるにすぎない。本物であることにとって本来的に重要なのは現象的性質であり、基盤はたんに派生的に重要であるにとどまる。乳分泌や光合成のシミュレーションも、それが本当に飲食でき栄養のある乳や糖のシミュレーションを産出するという現象的性質を示すとすれば、本物の乳分泌や光合成だと言ってよいだろう。それは通常の乳分泌や光合成とはその基底のプロセスにかんして大きな違いがあるが、それでも本物の乳分泌や光合成であり、それらの新たな形態なのである。

乳分泌や光合成のシミュレーションが本物でありうるとすれば、理解のシミュレーションが本物であることには、もはや異論の余地はないであろう。それは本物の理解に必要な現象的性質（適切な回答の産出能力）を備えており、通常の生物学的現象としての理解とその基底のプロセスにかんして大きく異なるとしても、それでも本物の理解の新たな形態なのである。

5. 形式的操作で適切な回答が可能か

サールは中国語の部屋の思考実験において、形式的な記号操作によって適切な中国語の回答が可能であることを認めつつも、そこに理解が生じていることを否定する。しかし、そもそも形式的な記号操作によって本当に適切な中国語の回答を行うことが可能であろうか。サールはこの点にまったく疑問を投げかけないが、形式的な記号操作によって理解が生じることを否定したいのであれば、そもそもそのような操作によって適切な回答が可能であるということこそ疑うべきではないだろうか。それを疑うべき相応の理由があるように思われる。

私たちが自分の理解できる言語で与えられたストーリーと質問にたいして適切な回答を行うとき、私たちは言葉をその形式的性質にもとづいて操作したりはしない。むしろ、ストーリーと質問に関連するさまざまな思考やイメージ、情動が立ち現れ、それらを介しておのずと回答にたどり着く。たしかに言葉による推論を行うこともあるが、それが文の構文論的構造にのみもとづく純粋に形式的な推論であることはまれである。推論もまた、そのほとんどが思考、イメージ、情動を介して行われる。もしコンピュータがそのような思考やイメージ、情動をもたないとすれば、コンピュータがそれらなしに純粋に形式的な記号操作だけで本当に適切な回答を産み出せるのかどうかおおいに疑問である。コンピュータは言語を理解する人間が抱くような思考やイメージ、情動を抱くことができるだろうか。

思考は言葉で表現される。考えることはふつつ語ることは違ふとされるが、考えることは心のなかで語ること、つまり内語を行うことだという考えもなかなか有力そうであ

る。この考えによれば、たとえば、今年の冬は暖かいと考えることは日本語話者なら「今年の冬は暖かい」と心のなかで語ることにほかならない。そうだとすれば、思考もまた言語表現にほかならないから、記号を扱えるコンピュータは思考を抱くことができるかもしれない。その可能性は否定できないように思われる。

しかし、イメージや情動はどうであろうか。イメージは文的な表象ではなく、絵画的な表象だと考えられる。可愛いネコのイメージは「このネコは可愛い」という文によってネコを表すのと同じタイプの表象ではなく、むしろ可愛いネコの絵によってネコを表すのと同じタイプの表象であろう。つまり、イメージは文のように構文論的構造をもつものではなく、絵のように構文論的構造を欠くものだと思われるのである。そうだとすれば、構文論的構造をもつ記号しか処理できないコンピュータはイメージを抱くことができないだろう²。

また、コンピュータは情動も抱くことができないように思われる。情動も構文論的構造をもたないであろうし、さらにそれだけではなく、情動には身体が不可欠であるように思われる。情動の本性については諸説あるが、情動は状況の価値的性質を身体的に感受するものだという説が有力だと思われる³。ヘビにたいする恐怖は、ヘビへの身体的反応（心臓や胃腸などの変化）を介してヘビの危険性を脳で感受したものである。そうだとすれば、身体を欠くコンピュータは情動を抱くことができないだろう。ロボットは「身体」をもつと言えるかもしれないが、その「身体」は生きているということが問題になるような生物的な身体ではない。情動に必要なのは生物的な身体だ。それゆえ、ロボットの「身体」は情動を可能にするようなものではない。

コンピュータは思考を抱くことができるとしても、イメージや情動を抱くことはできないように思われる。そうだとすれば、コンピュータが形式的な記号操作によって適切な回答を産み出せるということは、非常に疑わしいであろう。もちろん、私たちが自分の理解できる言語で与えられた質問に適切に応答するとき、関連するイメージや情動を用いているとしても、そのことは適切な応答にはイメージや情動が絶対に必要だということを必ずしも意味しない。そのようなイメージや情動がなくても、形式的な記号操作のみによって適切な回答が産み出せる可能性は、完全には否定されない。しかし、それが非常に疑わしいということは間違いあるまい。

ところで、形式的な記号操作で適切な回答を産み出せるということがきわめて疑わしいとしても、脳シミュレータ説が想定するように、脳の神経活動をコンピュータでシミュレートする場合は、どうであろうか。この場合は、中国語を理解できる人の脳の神経活動を全面的かつ詳細にシミュレートすれば、ストーリーと質問に関連する思考やイメージ、情動をシミュレートすることもできるから、中国語の適切な回答を行うことが可能であるように思われる。情動をシミュレートするためには、脳の神経活動だけではなく、身体の生理的活動もシミュレートする必要があるというのであれば、そうすればよい。身体の生理的活動を形式的にコンピュータに反映させれば、それをシミュレートすることは可

能である。このように脳と身体の神経的・生理的活動を全面的にシミュレートすれば、適切な回答を行うことは十分可能であるように思われる。

これにたいして、ことによるとサルなら、つぎのように反論するかもしれない。脳と身体の活動をいくら詳細にシミュレートしても、それはそれらの活動の形式的性質をシミュレートしたにすぎず、脳と身体を構成する物質的素材の因果的性質をシミュレートしたわけではない。コンピュータは形式的対象を形式的性質にもとづいて操作するものであるから、シミュレートできるのは形式的なものだけである。脳と身体の素材の因果的性質をコンピュータでシミュレートすることは定義上、不可能である。たしかに因果的性質に対応する形式的性質をシミュレートすることは可能である。脳や身体の活動にたいして、その形式的モデルをコンピュータ内に構築することはできる。しかし、そのような形式的モデルは脳や身体の活動の因果的側面を完全に捉えたものではない。因果的側面は途方もなく豊かであり、それを完全に形式的に捉えきるのは不可能であろう。だが、この形式的には捉えきれない因果的性質が適切な回答には不可欠ではないだろうか。そうだとすれば、脳と身体の活動をいくら詳細にコンピュータでシミュレートしたとしても、適切な回答を産み出すことは不可能であるように思われる。

しかし、本当にそうであろうか。たしかに脳と身体の活動はその因果的側面において途方もなく豊かであろう。しかし、そのような側面を形式化して捉えるコンピュータの能力は飛躍的に向上しており、さらに飛躍的に向上し続けるであろう。カーソワイルに倣って言えば、それは線形的ではなく、指数関数的に向上するであろう (Kurzweil 2005)。したがって、途方もなく豊かなものを十分に形式化できるのはずっと先のことだと思っても、意外にそれは早く到来するかもしれない。アナログのレコードの音声はデジタルのCDによってすでに十分形式化でき、人間の耳にはその違いはもはや聞き取れない。それと同じように、脳や身体の活動もやがて、コンピュータによってもはや実質的な違いがない程度にまで形式化されるであろう。指数関数的な向上が正しいとすれば、その時は意外と早くやっばこよう。たとえ指数関数的な向上が頭打ちになり、向上が鈍化しても、十分な形式化が原理的に不可能だということにはならないだろう。完全な形式化が原理的に不可能であっても、十分な形式化は原理的に可能であろう。

結局、形式的操作で適切な回答を行うためには、記号の形式的操作ではなく、脳と身体の神経的・生理的活動の形式化という方式で形式的操作を行う必要があると思われる。人間の全面的かつ詳細な形式化を行わずに、たんに記号の形式的操作を行うだけでは、適切な回答を産出できないが、それを行えば、適切な回答を産出することは可能であろう。

6. まとめ

サルは中国語の部屋の思考実験を通じて、たとえ記号の形式的操作によって質問者にとって理解可能で適切な回答が産み出されたとしても、そこには理解は生じていない

と主張したが、そもそも記号の形式的操作によって適切な回答が産み出されるという想定はきわめて疑わしい。また、かりにそのような回答が産み出されたとすれば、そこに理解が生じていないという主張は同意しがたい。適切な回答が産み出されたのであれば、その回答はたんなるシミュレーションではなく、本物の回答であり、それゆえその回答をもたらした理解もたんなるシミュレーションではなく、本物の理解であろう。それは通常の生物的現象としての理解とは基盤的性質において大きく異なるが、それでも本物の理解の新たな形態である。しかし、そもそも形式的操作によって適切な回答を産み出すためには、記号の形式的操作ではなく、脳と身体の神経的・生理的活動の全面的かつ詳細な形式化が必要であろう。そのような形式化を十分に詳細に行えば、適切な回答の産出は少なくとも原理的には可能なはずである。この回答はもちろん本物であり、それゆえそこには本物の理解が生じているのである。

註

- ¹ 本稿では、サールへの言及はすべて論文「心・脳・プログラム」にもとづく。
- ² イメージは知覚と同じ表象性格をもつと考えられるが、知覚における構文論的構造の欠如については、信原 2002, pp.197-205 を参照。
- ³ 情動の身体的感受説については、Prinz 2004、信原 2017 を参照。ドレイファスは、コンピュータは身体をもたないから心をもたないと主張する (Dreyfus 1972)。

文献

- Dreyfus, Hubert L. 1972. *What Computers Can't Do: The Limits of Artificial Intelligence*. New York: Harper & Row, Publishers. 邦訳、H・L・ドレイファス『コンピュータには何ができないか：哲学的人工知能批判』（黒崎政男・村若修訳）産業図書、1992年
- Kurzweil, Ray. 2005. *The Singularity is Near: When Humans Transcend Biology*. New York: Viking Books. 邦訳、レイ・カーツワイル『ポスト・ヒューマン誕生：コンピュータが人類の知性を超えるとき』（井上健監訳、小野木明恵ほか訳）NHK出版、2007年
- Nagel, Thomas. 1974. What Is It Like to Be a Bat? *The Philosophical Review*, 83(4): 435-450. 邦訳、トマス・ネーゲル「コウモリであるとはどのようなことか」（永井均訳）トマス・ネーゲル『コウモリであるとはどのようなことか』勁草書房、1989年、所収
- Putnam, Hilary. 1975. The Meaning of "Meaning". In Keith Gunderson (ed.), *Language, Mind, and Knowledge (Minnesota Studies in the Philosophy of Science*,

Volume 7), pp.131-193.

Prinz, Jesse. 2004. *Gut Reactions: A Perceptual Theory of Emotion*. Oxford University Press. 邦訳、ジェシー・プリンツ『はらわたが煮えくりかえる：情動の身体知覚説』（源河亨訳）勁草書房、2016年

Searle, John. 1980. Minds, Brains, and Programs. *The Behavioral and Brain Sciences*, 3: 417-424. 邦訳、ジョン・サール「心・脳・プログラム」（守屋唱進訳）D・R・ホフスタッター&D・C・デネット編著『マインズ アイ：コンピュータ時代の「心」と「私」』（下）TBSブリタニカ、1984年、所収

信原幸弘(2002)『意識の哲学』岩波書店

信原幸弘(2017)『情動の哲学入門：価値・道徳・生きる意味』勁草書房