

身体性の無い知を考える

Disembodied Intelligence – Is It Possible?

中島 秀之

Hideyuki Nakashima

公立ほこだて未来大学

Future University Hakodate

Intelligence cannot be separated from the environment in which it operates. It is even part of the definition of intelligence that it is an ability to operate well in the environment. We argue the above proposition from the point of view of situatedness (or autopoiesis). Frame problem is a good argument to highlight the point.

1. 動機

知の身体性への投稿を呼びかけられたときに先ず考えたのは、そのような当然なことに人工知能研究の立場から何か言うべきことが残っているのだろうか？という問題である。生物学や心理学、生理学などを専門にしていれば、知の身体性とは具体的にどのようにして実現されているのかを述べることができる。しかしコンピュータプログラムを研究の場とする人工知能の立場から何が言えるのだろうか？

ここでふと気がついた。先ず知があって、それが身体性を持つのではない。先ず身体があって、それが知を持つのである。生物の進化の歴史を見てもそれは明らかである。Saganの還元主義はそのような見方の好例で、著書「エデンの恐竜」[Sag78]で彼は以下のように述べている：

脳についての私の基本的な前提は、そのはたらき—ときに私たちが「心」とよぶもの—が、脳の解剖学と生理学の結果であって、それ以上のものでないということだ。

ちなみに筆者はすべての生物には（バクテリアにも）知能があると考えている。もちろん程度問題であるが、生物の階層のどこかで知能の有無の線を引くことは無理であろう。従って知の無い身体は考えられても、身体が無い知はあり得ない。では、知は身体にどのように依存しているのだろうか？身体が異なれば知は異なるのだろうか？それとも知は別の身体に移植可能なものだろうか？「マインズ・アイ」[HD85]はその問いに肯定的に答えようとするアンソロジーである。序章で Hofstadter は以下のように述べている：

そこで、本書はあなたに問いかける。もし、あなたがあなたの身体を、他のもっと頑健な、もっと美しい、そしてもっと制御のきく身体と取り換えることができるならばどうなるだろうか。

まずあなたは言うかもしれない。それは不可能だ、と。しかし、この書物は主張する。それは完全に想像可能なことであり、したがってそれは原理的に可能なことである、と。

彼の考え方は初期の人工知能研究者に共通のものである。人間の身体をコンピュータというハードに置き換えても知能は実
連絡先: 中島秀之, 公立ほこだて未来大学, 〒041-8655 函館市
亀田中野町 112-2, 電話 0138-34-6457

現できるという信念が人工知能研究を駆動していた。同書にはチューリングテストの論文 [Tur50] も収録されているが、これも通信路をテレタイプによる文字に限定し、身体性を排除した知能テストである。

その後、AIのこの方向性（物理記号仮説 [NS72]）は行き詰まりを見せており、最近では環境との相互作用などを重視する方向に変わっている。理由は様々あるが、その一つは2.1節で述べるフレーム問題である。多くの研究者が論理の枠組みの中で解決しようとしたが成功していない。また1980年代に実用化の期待をもたれていたエキスパートシステム [HRWL85]の失敗も、今から見れば身体性の無さに起因しているように思う。その典型が MYCIN であろうか。感染症の治療を示唆するエキスパートシステムであり、若手のインターンよりは能力があると言われていたが、子供に対して注射の使用を多用するなど、患者の痛みや苦痛への配慮が無いことなどから実用化はされなかったと聞いている。身体性の無いプログラムには酷な要求であろう。もちろん注射は痛みなどの知識を逐次追加することは可能であろうが、これもやはりフレーム問題となってしまう。

このような流れからしても知の身体性や環境依存性は（少なくとも私にとっては）自明なのだが、その自明のことを証明して見せるのは難しい。

そこで逆を考えてみた。背理法の原理である。身体性を持たない知を想像するところから始めるのである。コンピュータプログラムは良い例である。生物のような意味では身体を持っていない。実際、初期の人工知能研究では身体性が無視されており、論理演算のような純粋記号操作に知能を見ようとしていた。

2. 純粋知能

身体性のない純粋知能というのは考えられるだろうか？現在のコンピュータはそれに最も近い存在であろう。多少の記号（データ）入力はあるが、それは人間の感覚入力とは比較にならない程貧弱なものである。量的に比較するのは困難であるがおそらく億とか兆とか京とか、あるいはもっと大きなオーダーで違っているのだと思う。コンピュータの入力を遮断してプログラムを走らせることを考えてみよう。

定理を発見するプログラムというのが作られたことがある。基本的な公理と推論規則を与えられた上で、後はひたすら新しい定理を求めて計算を続ける。これ等は純粋知能と言うべきものだろう。

数学のような単純化された世界でこそ成立するのだと思う。単純でない世界を構築できるか？

Deutschは仮想現実シミュレータ [Deu97] というのを考えている。人間の五感をすべてシミュレートして見せるコンピュータシステムである。この中に居る人間にとって必要な外界刺激はすべてこのシミュレータが作り出す。感覚神経に直接入力するのだ。人間とこのシミュレータとの全体の系を考えると入力が無い。出力も無い。システムで閉じており外部と情報のやりとりが無いのである。つまり身体性が無い。この系は学習するだろうか？定理発見プログラムと同じ意味であらかじめ用意された反応（の新しい組み合わせ）しか人間に返さないのではなからうか？これを考える鍵はやはりフレーム問題にある。

2.1 フレーム問題

フレーム問題 [Den87, Mat90] とは推論に必要な前提条件や、行為の帰結が記述しきれないという問題である。論理的推論の枠組みで様々な解決法が試みられたが、完全な解は得られていない [HM86]。そもそも人間にも解決不能な問題である [MH89, NMO97]。しかし、人間は一見フレーム問題が存在しないかのように振る舞うことが可能である。これは環境を含めた系をうまく利用しているからであり、人間の外にあるリソースを使うという意味で身体性を使っている。

予想通り不合理 [Ari10] という面白い題の本がある。人間はときどき不合理な決断や行動をするが、それらはランダムなのではなく、多くの場合予測できる一定の法則に従っているというのである。具体的な例をここに書く余裕は無いので本を参照いただきたいが、私はこれは生きるための戦略（知恵）がうまく働かない場面のことだと考えている。

これはフレーム問題を解く為のシステム（厳密な言い方をすると、フレーム問題が顔を出さないように避けて処理するシステム）が、状況依存性 [Nak92]。を使っており、それがうまく働かない状況が存在するということが解釈している。状況推論とは推論過程の一部を環境にゆだねるものである [Kat91]。

「サイモンの蟻」の比喩 [Sim96] を使って考えたい。地面を歩いている蟻の軌跡は非常に複雑である。しかしながら、それを生成している蟻の行動原理は非常に単純だということである。地面を歩く蟻の軌跡は複雑であるし、障害物をたくみに避け、しかも最短経路をたどっているように見える。しかし、蟻は地面の高低差や障害物を避け、単に歩きやすい方へ動いているだけなのだ。地面の構造の複雑さが軌跡に反映しているのだ。つまり主体の複雑さではなく環境の複雑さが見えているのである。いくら蟻の内部構造を調べてもこの複雑さを生み出す機構は見つからない。

生物が行動の一部をこのように環境にゆだねているという例は多く見ることができる。Pabrovが犬の条件反射という学習能力を発見したのは1902年のことである（研究成果の出版が1923年頃）。餌を見ると唾液が出るというような生得的な無条件反射の他に、餌を与える度にベルの音を鳴らし続けているとベル→餌という反射が形成される。これは環境の規則性を取り込んで予測に使っていると考えられる。

Tinbergenはイトヨの攻撃行動の観察から生得的解発機構の存在を発見した [Tin51]。イトヨのオスは繁殖期に入るとメスを迎える為の巣作りをしてそこを自らの縄張りとしている。その縄張りに他のオスが入ってくると激しい攻撃行動を示す。しかし、この行動は実際にオスのイトヨではなく、下半分を赤く塗った木片でも解発される。彼はこれ以外にもヒナが親を認識する刺激なども観察し、様々な場面で解発刺激は非常に単純な構造を持っていることを示した。オスとメスの求愛行動も互いに異性が解発刺激となった反応連鎖である。これらの発見

は、一見複雑に見える行動も実はこの刺激と反応の連鎖に過ぎない場合が多いことを示唆している。この刺激の認識が単純なので良い理由はイトヨの住む環境に赤い色が他に存在しないからである。異なる環境であれば異なる解発刺激を使うようになると考えられる。

Lorenz [Lor52] は、卵から生まれたばかりのハイイロガンが初めて見た動く物体を親と思い込んでしまうという刷り込み現象を発見した。彼自身がハイイロガンの雛に母親と間違われて追い回されたそうだ。この刷り込み機構は遺伝情報として伝わるべき情報の一部を環境の側にゆだねた例と考えることが出来る。親の特徴を遺伝子に書き込むより、最初に見たものを親と思えというコードだけを書き込む方が遥かに効率が良い。また長い間の進化や環境変化により親の形態が変わってもコードを変える必要が無い。

人間の食事に関しても「おふくろの味」がそれではなからうか。我々は味で食物の適否を見分けているわけだが、これをあらかじめ遺伝子に組み込んでおくのは適切ではない。ひとつには情報量が大きくなりすぎることがある。学習メカニズムだけを組み込んでおいて、パラメータは後天的に脳神経系が成熟した後に獲得する方が良い。もうひとつは環境の変化への適応である。遺伝的变化には少なくとも数十世代（あるいは数百世代？）程度の長い時間を要すると考えられる。しかし食物環境は、個体が移動すれば一世代のうちに変化してしまうかもしれない。このような早い変化に対応するには学習が良い。人間は最初に食べたものしか食べないというような単純な刷り込みはしないが、子供時代に食べた味がその後の嗜好を決めているのは事実である。

2.2 鬱病

鬱病というのは複雑な精神疾患であるから、一言で定義することは危険なのであるが、本稿の目的に沿った形でまとめてしまうと要するに抑鬱気分や不安・焦燥気分や不安・焦燥など精神活動の低を伴う精神疾患である。しかし、これは機能障害ではなく、精神の防衛反応であるとする本 [Nom08] もある。人間の本来持っている防衛機能が、その適用範囲を超えた状況で見せる反応だということである。環境が適切であれば鬱病ではなく、正常な反応である。

2.3 離人症

離人症 (depersonalization) とは、「自分」の存在を感じることのできなくなる症状である。様々な知識はあるし、自分の周囲を客観的に捉える機能はあるのだが、自分がそこに存在しているということが感じられなくなってしまうそうである。

ある日、彼女はこんなことを言った。「時間も同じことなんです。時計を見れば時間が過ぎてゆくことは分かるし、いま何時何分だということも分かるんだけど、時間の流れというものが感じられないんです。てんでばらばらでつながりのない無数の今が、今、今、今、と無茶苦茶に出てくるだけで、何の規則もまとまりもないんです....」 ([Kim94] p. 101)

木村敏は世界を客観的に、つまり主体と切り離された存在（もの）として見る限りそこには豊かな認識は起こり得ないと言っている。そうではなくて、主体がその認識過程の一部として存在するような認識（こと）が必要であり、離人症患者には後者：「こと」の認識が欠けているのではないかと言う。

これも身体性、あるいはクオリア [Mog97] の欠如とも言えるのではないだろうか。つまり、環境から切り離された知はうまく使えないのである。

2.4 シンボルグラウンディング問題

AIでシンボルグラウンディング問題 [Har90] が採上げられる理由は、身体性の無い知を扱おうとするからだ。記号だけを扱っていても知的にならない。だから現実世界に接地させる必要がある、という本稿の趣旨からすればごく当たり前のことを述べているに過ぎない。

3. 知の身体性

そろそろ知の身体性を定義しても良いだろう。

知能と知は少し異なる。前者は能力であり、後者はその能力が発揮された結果の機能であり、またコンテンツである。以下分けて論ずる。

3.1 知脳の定義

私は知能を以下のように定義している：

情報が不足した状況で適切に処理する能力

完全な情報が入手可能で、完全な処理が可能であれば神の視点に立てば良い。たとえば、データの集合の最大値を探したり、あるいはそれらを値の順に並べ替えたり、複数の都市を最短時間で巡回するルートを求めたり、ある自然数を素数分解する問題であれば数学的アルゴリズムが存在し、それらは常に正しい解を有限時間内に計算することを保証している。たとえば膨大な時間を要することになろうとも、それを待てるのであればこれらの問題の解決には「知能」を要しない。

しかし、すべての情報が入手できない場合や、囲碁や将棋のゲームのように、原理的にすべての情報は入手可能であるが、状況が複雑すぎてそれをすべて処理する時間やメモリ容量が不足している場合^{*1}には別のことを考える必要がある。私はこれが知能だと考えている。これらの場合には過去の経験や状況判断によって様々な方略を採用する。それらは一般的にヒューリスティクスと呼ばれるが、2.1節で述べたように、処理の一部を環境にゆだねるということである。そのためたまには失敗するが、たいていは効率よく処理できる。

最近の先導的ロボット研究者たちは、もう少し過激な異見を持っていて、知能は個体と環境の総体の中に存在すると主張している。たとえば國吉は知能を

変動する複雑な環境中で安定に目標を達成する行動を生成する能力

と定義している ([AK06] 第一章)。これはロボットの文脈で語られているため行動に特化しているが、推論や小説などの抽象的行為を含めても良いと思われる。私と彼の定義には表面的な差がある (情報の不足と環境の変動) が、よく考えれば同じ性質の問題であることがわかる。

知能の本質は環境との相互作用の中にあるということができる。結果として知的に見える振舞の半分は環境の起伏などの複雑さが担っているというのである。環境と主体のインタラクションという観点を抜きにして、いくら主体の構造や性質を調べても本質にはたどりつけない。

実は生物学の分野でも Uexküll がこの点を早くから^{*2}主張している [UK05]。彼は環境ではなく「環世界」という用語を用い、環境の見えは生物が自ら作り出しているものであり、外

*1 情報が不足しているのか処理資源が不足しているのかは原理的には識別不能である ([Has94, HM94])。これはチューリングの停止性問題と類似している。

*2 「生物から見た世界」の原著は1934年で、Lorenzらの動物行動学者に影響を与えたとされている。

部から一方的に情報が入ってくるというものではないと主張している。日本語の「環境」は、英語で「environment」(包み込むもの)、ドイツ語で「Umgebung」(周囲に与えられたもの)である。環境とは一般的には主体とは別に客観的に存在するものと考えられている。Uexküllはこれを「Umwelt」(環世界)と呼んで、主体が積極的に作り出すものだという立場をとっている。Uexküllが使う例はGibsonがアフォーダンス [Jame85]の主張に使っている例に酷似しているが、Gibsonが環境を主に捉えているのとは方向が逆である。

3.2 知脳の身体性

以上の議論により知能の身体性は明らかになったと思う。

知能はオートポイエシスである。MaturanaとVarelaはハトの視覚の研究から、脳に投影される視覚像は必ずしも外界に対応物を持たないということを発見し、オートポイエシス [MV80, MV87]という考え方を提唱した。これは、システムはそれが情報として取り込む範囲を自ら規定するという考えであり、システムと環境との物理的な境界をシステムの境界とは考えるべきでないというものである。ある意味で環境とシステムが一体となって視覚などの認知が起きている。

3.3 知の身体性

では知能のコンテンツである知の方はどうだろうか？ 刷り込みや「おふくろの味」の例はコンテンツを環境に求めた例ではなかろうか？ 知もまた状況依存でしか記述できないのである。フレーム問題の一部である記述問題 [MH89]がこれに相当する。

4. まとめ

知の身体性について状況依存性の観点から述べ、知が環境にその一部を依存しているという考え方を主張した。オートポイエシスの考え方にも通じる。フレーム問題がそれらを論じる良い鍵になっている。

結論として、身体性の無い知は考えられない。

参考文献

- [AK06] 浅田稔, 國吉康夫. ロボットインテリジェンス. 岩波書店, 2006.
- [Ari10] ダン・アリエリー (熊谷淳子訳). 予想どおりに不合理 (増補版). 早川書房, 2010.
- [Den87] Daniel Dennett. Cognitive wheels: the frame problem of ai. In Z.W. Pylyshyn, editor, *The Robot's Dilemma*. Ablex Publishing Co., 1987.
- [Deu97] David Deutsch. *The Fabric of Reality: The Science of Parallel Universes-And Its Implications*. The Penguin Press, 1997.
- [Har90] Stevan Harnad. The symbol grounding problem. *Physica D*, Vol. 42, pp. 335-346, 1990.
- [Has94] 橋田浩一. 知のエンジニアリング: 複雑性の地平. ジャストシステム, 1994.
- [HD85] Douglas R. Hofstadter and Daniel C. Dennett. *The Mind's I*. Bantam Dell Pub Group, 1985. 坂本百大監訳: マインズ・アイ. TBS ブリタニカ (1992).

- [HM86] Steve Hanks and Drew McDermott. Default reasoning, nonmonotonic logics, and the frame problem. In *Proc. of AAAI-86*, pp. 328–333, 1986.
- [HM94] 橋田浩一, 松原仁. 知能の設計原理に関する試論—部分性・散層・フレーム問題—. 日本認知科学会年報「認知科学の発展」, Vol. 7, pp. 159–201, 1994.
- [HRWL85] Frederick Hayes-Roth, Donald Waterman, and Douglas Lenat, editors. *Building Expert Systems*. AIUEO 訳: エキスパート・システム (産業図書), 1985.
- [Jame85] James J. Gibson(古崎敬, 古崎愛子, 辻敬一郎, 村瀬旻訳). 生態学的視覚論. サイエンス社, 1985.
- [Kat91] 片桐恭弘. 状況推論とその機構について. 日本認知科学会第8回大会発表論文集. 日本認知科学会, 1991.
- [Kim94] 木村敏. 心の病理を考える. 岩波新書, 1994.
- [Lor52] Konrad Z. Lorenz. *King Solomon's Ring*. Methuen and Co., Ltd., 1952. 日高敏隆 (訳): ソロモンの指環—動物行動学入門.
- [Mat90] 松原仁. フレーム問題をどうとらえるか. 認知科学の発展 vol.2, pp. 155–187. 講談社サイエンティフィック, 1990.
- [MH89] 松原仁, 橋田浩一. 情報の部分性とフレーム問題の解決不能性. 人工知能学会誌, Vol. 4, No. 6, pp. 695–703, 1989.
- [Mog97] 茂木健一郎. 脳とクオリア—なぜ脳に心が生まれるのか. 日経サイエンス社, 1997.
- [MV80] Humberto R. Maturana and Francisco J. Varela. *Autopoiesis and Cognition: the realization of the living*. D Reidel Pub Co, 1980. 河本英夫訳: オートポイエーシス, 国文社, 1991.
- [MV87] Humberto R. Maturana and Francisco J. Varela. 知恵の樹. 朝日出版社, 1987. 菅啓二郎訳.
- [Nak92] 中島秀之. 状況に依存した推論. 人工知能学会誌, Vol. 7, No. 3, pp. 392–398, 1992.
- [NMO97] Hideyuki Nakashima, Hitoshi Matsubara, and Ichiro Osawa. Causality as a key to the frame problem. *Artificial Intelligence*, Vol. 91, No. 1, pp. 33–50, 1997.
- [Nom08] 野村総一郎. うつ病の真実. 日本評論社, 2008.
- [NS72] Allen Newell and Herbert A. Simon. *Human Problem Solving*. Prentice Hall Inc., 1972.
- [Sag78] カール・セーガン (長野敬訳). エデンの恐竜—知能の源流をたずねて. 秀潤社, 1978.
- [Sim96] Herbert A. Simon. *The Sciences of the Artificial*. MIT Press, Cambridge, Massachusetts, third edition, 1996.
- [Tin51] Nikolas Tinbergen. *The Study of Instinct*. Clarendon Press, Oxford, 1951. 永野為武訳: “本能の研究”, 三共出版, 1975.
- [Tur50] Alan M. Turing. Computing machinery and intelligence. *Mind*, Vol. LIX, No. 236, 1950. 坂本百大監訳: マインズ・アイ. TBS ブリタニカ (1992) に再録.
- [UK05] ユクスキュル, クリサート (日高敏隆, 羽田節子訳). 生物から見た世界. 岩波書店, 2005.