

主体と環世界

Embodiment of an Agent as the Umwelt

中島 秀之*¹

Hideyuki Nakashima

*¹公立はこだて未来大学

Future University Hakodate

Intelligence is a phenomena emerging on top of the interaction between the agent and its environment. In other words, intelligence is embodied in the interaction. The interaction creates an “Umwelt” (enclosing world) of the agent. In this sense, it is not entirely correct to say that an agent is not embodied in the environment, but rather, we say that the agent’s interaction with the environment *IS* the Umwelt.

1. はじめに

知は環境との相互作用の上に成立しており、その相互作用こそが身体性だ、いや、そもそもこの主体と環境という分離がない。「境」などどこにも存在していないのだから。実は、この主体と環境が切り離せないという考え方は古来より様々な分野において異なる言葉で表現され、繰り返して主張されて来た。本発表ではそれらを概観し、知の本質に近づきたい。

知能の身体性を考えるときに、私には二つの疑問がある。

第一の疑問。これは他でもよく議論されることだが、主観的にどこまでを自分の身体だと感じるかという疑問。特に車を運転していたりすると車全体が身体の一部とを感じるようなことがある。聞くとところによればベテラン戦闘機パイロットは自分が機体になったように感じるそうである。そこまで極端ではなくとも手に持った箸や足に履いた靴などはかなり身体と一体化しているのではなかろうか。また、背中に背負った荷物の膨らみは忘れがちになることも皆さん経験済みであろう。

もう一つの疑問。身体の状態を意識的に変えられるという点。たとえば緊張すると反応が早くなる。個人的な例としては飛行機の操縦を習っていたときに、着陸時の機体の姿勢の乱れ(主として風が原因)に対する修正が遅れがちだった。しかし意図的に早く反応するモードに切り替えることを学んだ。しかし、これは制御系のゲインを上げるのとは違う気がする。応答が良くなるのではなく早くなる。どうすればそういったことが可能なのか？

第一の疑問は空間的、第二の疑問は時間的なものである。もちろん、これら二つは互いに関連しているのだが、本稿ではとりあえず第一の疑問を掘り下げてみたい。解はまだ無いのだが。

まず2.節でAI研究における身体性の扱いの歴史を概観する。続いて3.節で主体(システム)と環境との関係についての、AI以外の議論の歴史を概観する。最後に4.節で環世界という概念について紹介した後、身体性の議論に入る。主張の要点は身体性を主体の身体に限定して捉えてはならないということである。身体性は主体と環境との相互作用から生まれ、その相互作用が主体を決定するという円環関係になる。これをUexküllに倣って環世界と呼んでみる。

2. AIにおける身体性

2.1 古き良きAI

知能とは記号の形式的操作である(物理記号仮説 [NS72])というのがAIの最初の重要な主張であった。従って、初期のAI研究においては知識の表現とその操作(推論)が重要視されていた。これは知的主体の存在する環境を主体内部に表現し、その表現を操作することによってその環境における適切な動作を推論し、その結果に従って行動を起こせば知的に振る舞えるという考え方であった。この考え方は現在でも生きており、たとえばユーザーインタフェースの設計などでは暗黙裏に使われていることが多い。すなわち、ユーザの状態や要求をシステム内に表現し、それに応じてシステムがユーザに応答しようとするものである。

2.2 フレーム問題

すべての情報を計算機内に取り込んでから処理をするシステムでは、問題が複雑になり知識が増えるに従い、内部表現の両が増え、とその操作が組み合わせ爆発的に複雑になる。計算が実用的な時間では終らなくなってしまふ。

フレーム問題とはMcCarthyとHayesによって“発見”^{*1}された問題である。計算機に知識を持たせようとする、我々が常識として、通常は忘れていたような細かい規則を与えなければならない。しかも、その大部分は x は y に影響しないという否定的な規則である。この記述量が膨大になると同時に推論が遅くなるというのが問題である。世界はその一部しか変化しないにもかかわらず、大量の不変の枠組(フレーム)を記述しなければならないのでこの名がある。

その後の研究により、そもそも単一の規則すら、例外が多くて書き切れないという一般フレーム問題に発展した。一般フレーム問題の完全な解決とは、行為の前提条件や帰結の記述の量、および行為の影響範囲の推論の量をともにある一定の範囲におさえ込みながら、なおかつ、いかなる場合にも行為の影響に関する完全な推論を行なうことを意味する。しかし、情報と処理の不完全性(2.4節)という観点からすると、これらは解決不能の問題である。実際、人間にも解決できていない。人間の場合はむしろ逆に、多くの潜在的な影響を排し、ある行為に関する本質的な前提条件や影響だけを囲い込むことにより推論の量を減らしていると考えられる。つまり、人間が日常生活に

*1 一般フレーム問題は発見以前から存在していたのにも関わらず、哲学が見落として来た問題として一時期哲学界(の一部)を騒がせた。

においてフレーム問題に悩まされていないように見えるのは、日常生活に支障をきたさないような範囲に推論を限定しているからである。本稿の文脈で言えば、人間は環境のすべてを表現・考慮することなく、環境をそのまま利用している。日常とはかけはなれたパズルの状況を人為的に作り出してしまうと、人間もフレーム問題を解決していないことが明らかになる。

2.3 状況推論

そのようなわけで、知識表現とそれに基づくプランニングという旧来の AI モデルに対する反省から、外界との相互作用が強調されるようになってきた。この極致には、従来の意味での知識表現は無用であるとの Brooks の主張がある [Bro91]:

非常に単純なレベルの知能を調べてみると、世界の明示的な表象やモデルは全く邪魔であることが分かる。つまり、世界をそれ自身のモデルとして用いる方がよい、ということが判明する。

つまり、複雑さは世界にあるのであって、計算機内の表現が複雑である必要はないという考え方である。すなわち、認識や推論は内部表象を通さずに直接環境と相互作用を行うことになる(図 1)。Brooks らはこの考え方で NASA のマーズローバーが従来方式より敏速に行動できることを示した。フレーム問題に悩まされないロボットの出現である。^{*2}

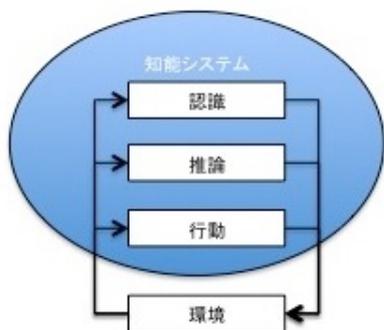


図 1: 服属アーキテクチャ

Brooks ほど極端ではなくとも、様々な手段により状況を参照しながら推論を行なうことが考えられている。その一つが状況理論 [BP83] をもとにした状況推論の考え方である。これは、例えば囲碁を打つ場合に、盤面の内部表現を持たなくても盤面を参照しながら手を考えられるようなものである。一部を内部表現に持ち、他を外部情報に頼るのである。このような推論においては、その推論の正しさは状況の方が保証することになる。本稿の文脈で言えば盤面までを身体の一部と考えるのである。

Perry が好んで用いる例に、ねずみとりがうまく働くのはねずみが特定の大きさや重さであるという状況をうまく利用しているからであるというのがある [BP83]。ねずみとりは、ねずみの大きさを測定してそれに合わせた位置にバーを落しているわけではないし、ましてやねずみの大きさ等に関する内部表現

を持っているわけではない。ねずみが小さければバーは空ぶりするし、ねずみがチーズが好きでなければそもそもバネが落ちない。この例の方が身体性らしいかもしれない。

状況推論的考え方では環境をシステムの一部と考えることになる。図 2 のように、主体だけではなく環境を含む全体が知能システムということになる。これが新しい AI の考え方である [NA96]。3. 節ではこれと相似の、環境を知能の一部とみる、AI 以外の様々な考え方を概観する。

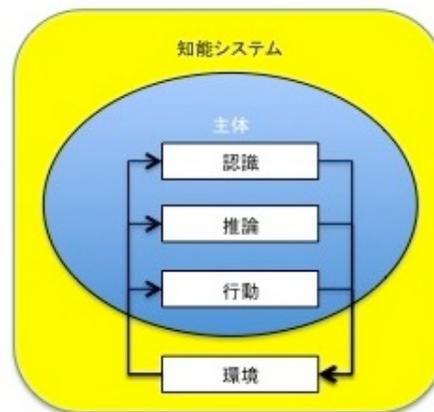


図 2: 最近の AI の知能観

2.4 情報の部分性と複雑系

外界との相互作用を重視し、外界に関する完全な情報を持ってないこと(情報の部分性)、それを処理するのに無限の時間を持ってないこと(処理の部分性)を前提にした知能の設計手法も盛んに研究されている [Has94, Mat94]。つまり、知識(情報)や推論時間が限定された状況でいかに最大限うまくやっけて行くかという行為主体(エージェントと呼ぶこともある)の視点に立脚した新しい考え方が求められている。完全処理のアルゴリズムではなく、適切な部分処理を行なうヒューリスティクスが重要である。物理学を中心として「複雑系」世界観が台頭してきたが、これも、完全予測を前提としないという意味で上記の知能観と通じる観点である。

3. 環境との相互作用

3.1 動物のしている世界

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[Uex05] 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 [Hid07]。

Uexküll はこの環世界の概念を説明するのにマダニの例を用いている。マダニの栄養源は哺乳類の血液である。交尾を終えたマダニの雌は木に登り、その下を哺乳類が通りかかるのを待つ。

この目のない動物は、表皮全体に分布する光覚を使ってその見張りやぐらへの道を見つける。この盲目で耳の聞こえない追いはぎは、嗅覚によって獲物の接近を知る。哺乳類の皮膚腺から漂い出る酪酸の匂いが、このダニにとっては見張り場から離れてそちらへ身を投げろという信号(signal)として働く。そこ

*2 筆者の主張としては、Brooks は極端に振れすぎていて、彼らの方式では昆虫の知能は作れてもそれ以上にはならないと考えている。やはり初期の AI が目指した記号の扱いも知能にとっては本質的である。思考における言語の使用が良い例である。しかし、本稿では身体性に注目するため、記号処理や言語の問題には深入りしないことにする。

でダニは、鋭敏な温度感覚が教えてくれるなにか温かいものの上に落ちる。するとそこは獲物である温血動物の上で、あとは触覚によってなるべく毛のない場所を見つけ、獲物の皮膚組織に頭から食い込めばいい。([Uex05] p. 12)

酪酸の知覚標識が働いた後になにか冷たいものの上に落ちてしまえば失敗である。ダニは最初からもう一度同じプロセスを繰り返す。

Uexküll はダニが受動的に動く機械かどうかを議論している。機械(モノ)の反応は物理法則に従う。ニュートン力学の $F = ma$ を例にとれば、力が倍になれば作用(加速度)も倍になる。またその方向は入力方向に等しい。しかし、筋肉の動きは刺激に比例するわけではないし、そして何より刺激の方向によらず筋肉の動作は収縮(あるいは弛緩)と決まっている。ダニは受容器に入る情報から知覚記号を選別し客体(獲物となる哺乳類)に一定の作用を及ぼす行動主体なのである。機械ではなく機械操作系である。そして図3に示すような機能環が成立する。ダニが主体、哺乳動物が客体である。矢印の始点と終点が主体ではなく客体側にある点に注目されたい。これがダニの環世界を形成している。

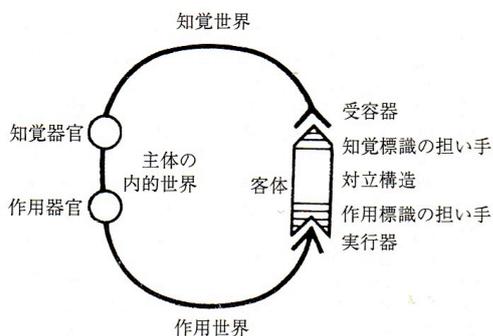


図3: ダニの機能環

ちなみに、本稿の主題とは若干ずれるが、濱野はこの Uexküll の主張に言及し、Google の検索エンジンが見ている Web 上の情報もリンクという環世界を創り出しているのだとしている [Ham08](pp. 47-48)。

3.2 アフォーダンス

Gibson の提案したアフォーダンスという概念 [Jame85] は、前述の Uexküll の環世界と説明事例は異なるものの、本質的には同じ主張であると私には見える。しかしながら、Gibson の著書にはその言及は見当たらない。

アフォーダンスは、やはり従来の主体が環境を観測するという考え方に反対し、主体は環境の提供するアフォーダンスを「ピックアップ」しているに過ぎないとする考え方である。^{*3}

3.3 オートポイエシス

「オート」は「自己」、ポイエシスは「産出」であるから自己産出システムと訳されることが多い。あるシステムが動作することによって、そのシステムを創り出している様をいう。Maturana の神経生理学的知見と Valera の位相空間論が

*3 これも振子が触れ過ぎの表現に思える。環境が一方的にアフォードし主体がピックアップするだけでは、主体の変化に応じた相互作用の変化が記述できない(フレーム問題になる)。相互作用とするのが適切だと考える。

一体となって生まれたのがオートポイエーシスの概念である [MV80]。

ここでは「システムには入力も出力もない」という一見とんでもないことが主張される。視覚系を考えると、眼の水晶体の表面で外界と区切られた視覚システムを考えるのではなく、光の発生源から水晶体、網膜、神経系、大脳、身体などが一体となって存在する刺激(情報)のシステムを考えることによって上記の言明の意味が初めて明らかになる。つまり、ある特定の視覚イメージはこれらの全システムが一体となって作り出しているものであり、視覚系の外側にあるイメージが内部に再現されるわけではないのである。このような、主体と環境を一体とした捉えかたは最近の大きな流れ:ゲシュタルト心理学、アフォーダンス(3.2節)、アクチュアリティ [Kim94] など、システムの認識動作を環境の中に位置づけようとしているもの、と同じである。

システムはその動作によって境界を形成する。しかし、これは物理的境界ではない。オートポイエーシス・システムを直接物理空間内に表象してはいけぬのである。機能的に設定される空間、つまり位相空間を考えなければ、オートポイエーシス・システムをうまく表象することができない。

3.4 ゲシュタルトクライス

システム論は、生命の本質を探ろうとする歴史として考えることができる [Hid94]。第一世代のシステム論では動的平衡が主題となったが、第二世代においては自己組織化が焦点となる。自己、境界、階層などの問題が扱われる。特に階層の成立が問題となり、円環形成 [Wei75] による階層の形成が考察される。そこには以下に示すように全体と部分の相互作用(ハイパーサイクル)が存在する。

構成要素のシステムが作動しなければ、全システムは作動しないが、にもかかわらず全システムからみれば、構成要素のシステムは、他のシステムに代替可能な要素システムの一つにすぎなくなる ([Hid94] p. 136)。

多田富雄が免疫に関してスーパーシステムと呼んだ [Tad93] のもこのようなハイパーサイクルの一種である。

このような様々なサイクル(円環)の考え方は本稿の主題である身体性と(ここで比較考察するスペースが無いのだが)様々な意味で関連している。明らかな対応としては本稿の様々な図に示しているサイクル(図2, 3)と密接に関係する。もう一つには心と身体間の円環を考えることにより DesCartes が心身問題と呼んだものの解となり得る [Nak11b]。

4. 環世界

4.1 内部視点

オートポイエーシス・システムは動作することにより構成要素を産出し現実の構造を構成する。一断面の静的構造を解析することは意味がない。サイクルが重要である。経済現象もそのようなシステムの一例である。なぜ貨幣が通用するのか?それは皆が貨幣が通用すると認めているからである。つまり、貨幣はそれが流通することによってのみ貨幣でありえるのである [Iwa93]。経済などの社会現象を扱う場合や人間の認知を扱う場合、観測者である人間が対象の系に含まれてしまう。観測者にとっての視点(境界、目的)とシステムにとっての視点(産出連関)の区別が重要である。これは、観測者の視点しか持たない自然科学の方法論ではとらえきれないものであることに注意する必要がある [Nak01, Nak06]。複雑系の議論でも内部観

測 [Gun97] が重要視されている。身体性の扱いは本質的にこのような領域に属する。

以下の議論はこの内部視点 [NSF06] を採ることによって初めて可能になる。詳細を述べるスペースが無いので、「環境」は外部視点からのもの、「環世界」は主体の内部視点からのものであることだけを指摘しておく。

4.2 身体性と環境

これまで見て来たように、知能は主体と環境との相互作用として捉えるべきである。Uexküll の主張にあるように、主体が違えば環境も異なるし相互作用も異なる。つまり、環境も相互作用も主体が造り出しているとも言えるし、逆にそれらの相互作用の下に主体が存在しているとも言える。

主体の身体性というのはそういうことであろう。即ち身体性とは：

- 環境から情報を得る (知覚) 手段
- 環境への働きかけ (作用) 手段

のことであり、それが、それら (相互作用) から産出される環世界を決めることになる。従って身体性とは環世界のことでありという言い方も成立する。人間が自立している場合と、車を運転している場合とでは身体性が異なり、環世界も異なるのである。

このように考えたとき、Brooks の服属アーキテクチャ (2.3 節) はまだ入力から出力への主体内の情報の流れだけを問題にするという従来の考え方に捕われていたことが分かる。

ここで忘れてはならないのは、主体は環世界以前に存在しているのではないという点である。主体が環世界を形成し、環世界が主体を形成するというメタレベルの円環が存在する。

マダニを Brooks の服属アーキテクチャで実現することは出来よう。しかし、それは結果だけを真似るシミュレーションに過ぎない。環境の中の何を見、何に働きかけるのかという「デザイン」が欠如している。ただし、この「デザイン」の方法論は進化論的なものしかないと言筆者は考えている [Nak11b, Nak11a]。マダニや人間等の、すべてのメタレベルの円環はこのような進化の結果として現存している。

5. まとめ

知能は主体と環境との相互作用にある。この相互作用が環世界を形成する。これが主体の身体性である。そして、この身体性が主体の (知覚や動作を含む) デザインを決める。こういった大きな円環の中で知能とその身体性を捉えるべきである。

参考文献

[BP83] Jon Barwise and John Perry. *Situations and Attitudes*. MIT Press, Cambridge, Massachusetts, 1983.

[Bro91] Rodney A. Brooks. Intelligence without representation. *Artificial Intelligence*, Vol. 47, pp. 139–160, 1991. (柴田 正良 訳. 表象なしの知能, 現代思想, 18 (3), 85–105, 1990).

[Gun97] 郡司ベギオ幸夫, 松野孝一郎, オットー・レスラー. 内部観測. 青土社, 1997.

[Ham08] 濱野智史. アーキテクチャの生態系. NTT 出版, 2008.

[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.

[Hid94] 河本英夫. オートポイエシス・システム. 岩波講座現代思想 12, 生命とシステムの思想, pp. 123–157. 岩波書店, 1994.

[Hid07] 日高敏隆. 動物と人間の世界認識—イリュージョンなしに世界は見えない. ちくま学芸文庫. 筑摩書房, 2007.

[Iwa93] 岩井克人. 貨幣論. 筑摩書房, 1993.

[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.

[Kim94] 木村敏. 心の病理を考える. 岩波新書, 1994.

[Mat94] 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.

[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: オートポイエシス, 国文社, 1991.

[NA96] 中島秀之, 有馬淳, 佐藤理史, 諏訪正樹, 橋田浩一, 浅田稔. 新しい AI 研究を目指して. 人工知能学会誌, Vol. 11, No. 5, pp. 37–48, 1996.

[Nak01] 中島秀之. 科学・工学・知能・複雑系—日本の科学をめざして. 科学, Vol. 71, No. 4/5, pp. 620–622, 2001.

[Nak06] 中島秀之. 構成的情報学と AI. 人工知能学会論文誌, Vol. 21, No. 6, pp. 502–513, 2006.

[Nak11a] 中島秀之. デザインするということ—進化に学ぶデザインの方法論. 日本デザイン学会誌, Vol. 18, No. 1, pp. 46–49, 2011. 特集「デザイン学: メタデザインへの挑戦」.

[Nak11b] Hideyuki Nakashima. Evolution-theoretic approach to synthetic study of intelligence. *Viva Origino*, Vol. 39, No. 1, pp. 1–5, 2011.

[NS72] Allen Newell and Herbert A. Simon. *Human Problem Solving*. Prentice Hall Inc., 1972.

[NSF06] Hideyuki Nakashima, Masaki Suwa, and Haruyuki Fujii. Endo-system view as a method for constructive science. In *Proc. 5th International Conference of the Cognitive Science (ICCS 2006)*, pp. 63–71, 2006.

[Tad93] 多田富雄. 免疫の意味論. 青土社, 1993.

[Uex05] 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.

[Wei75] ヴァイツェッカー V. ゲシュタルトクライス. みすず書房, 1975. 木村敏, 浜中淑彦 (訳).