ロボットは反表象主義の夢を見るか？

呉羽　真

0. 序論

1986 年、MIT のブルックスは、知能ロボットの設計に関して「包摂アーキテクチャ（SA: subsumption architecture）」理論を提唱し、このアプローチに基づいて、実世界において自律的に移動可能なロボットを製作した。この際、彼は、〈知能にとって、表象は不要である〉という「反表象主義」の立場を打ち出した(Brooks, 1991)。近年の心の哲学における、「力学系アプローチ（DSA: dynamical systems approach）」に基づいた「表象なき認知」を巡る議論においても、ブルックスの主張はしばしば引用される。しかし、SA 理論がロボット工学において一定の評価を得たのに比べて、DSA に基づくそれをも含め、反表象主義の主張は激しい批判に曝されている。また実際、現在でもなお、AI 研究の現場では、表象は不可欠と見なされている。

表象概念は従来の認知理論が拠って立つ基盤の位置を占めることから、反表象主義の主張は、我々の認知観の根本を問いただすことになる。本稿では、反表象主義とそれに対する批判の議論をサーベイすることを試みる。この際、ブルックスのロボット論を出発点とし、ヴァン・ゲルダーやクラークらの哲学的議論を概観した後、再びロボット製作の現場に回帰し、暫定的な結論を下すこととしたい。

1. ブルックスの非表象的ロボット

1.1 古典的制御アーキテクチャに基づくロボットとその問題

ブルックス以前に主に用いられていた、ロボットの古典的な制御アーキテクチャ（CCA: classical control architecture）は、次のようなサイクルに基づくものであった。[1]感覚。センサによって環境を知覚する。[2]モデル化。知覚されたデータに計算処理を加え、世界の正確な内部モデルを構築する。[3]計画。モデル上で統合された情報に基づいて次の行動を計画する。[4]行為。計画を行為モジュールによって実行可能な形に変換し、行為を実行する(人工知能学事典, 629 頁)。このようなアーキテクチャにおいて、システム全体は、その機能によって、入力システムとしての知覚モジュール、出力システムとしての行為モジュールと、両者の間にあって情報処理と世界の内部モデルの構築を担う中央システムに分割され、この中央システムこそが「認識ボックス」(Brooks, 2002)として知能の中核であるとさ

れた。

CCA に基づくロボットの代表例が、ニルソンに率いられてスタンフォード研究所が作ったシェイキー（Shakey）である。シェイキーは、ある部屋から別の部屋へ、障害物を取り除きながら積み木を運ぶ、といった作業を達成した。

しかし、シェイキーに代表される古典的ロボットは、以下のような問題を持っていた(cf. 人工知能学事典, 629 頁)。

[a]センサへの入力から行為モジュールへの出力までの情報変換プロセスに時間がかかるため、ロボットが示す動作は緩慢なものである。シェイキーは出発点から数メートル離れた目的地に到着し、与えられた仕事を達成するために 6〜8 時間を要した(Brooks, 2002)。

[b]世界のモデルに基づいて行為計画を立てるため、実世界において絶えず起こり続ける環境の変化に対して脆弱である。この問題に対して採られた対処法は、ロボットが稼動しやすいように環境を整えてやる、というものである。ブルックスはこれを「世界を極めて簡単なものに置き換えるという欺瞞」(Brooks, 2002, 邦訳, 44 頁)と称する。

[c]ロボットを設計するに当たって、世界のモデルを構築するために、あらかじめ設計者が環境内で起こりうる状況をシミュレートしておかなければならない。しかし、この作業は実現不可能と思えるほどに膨大なものになってしまうのであり、設計者は（広義の）「フレーム問題」に直面することになる。

このため、実際のロボット製作になかなか取り掛かれず、自分の研究内容に不満を抱いたブルックスは、古典的制御アーキテクチャとは別の道を模索する。

1.2 包摂アーキテクチャ（SA）理論

知能ロボットを製作するに当たってまず、ブルックスは「知能」とは何かを考えた。当時、AI 研究の初期にあっては、チェスや数学的な問題解決などが広くテーマとして取り上げられていた。これに対して彼は、これらの知的能力が、「生存と生殖を最低限保持するのに十分なほど周囲を知覚し、動的な環境世界を動き回ることのできる能力」(Brooks, 1991, 邦訳, 86-87 頁)を土台として、その上に築かれるべきだと考えた。つまり、初めは限られた単純な能力しか持たなかった動物が次第により高度な能力を身に付けてきたように、ロボット工学もまた、「知能の進化」の歴史を辿り直すべきだと考えたのである。

実世界で動き回ることのできるロボットを製作するためには、上で見たような問題を孕んだ CCA は役に立たない。これに代えてブルックスが提案する「行動規範型制御アーキテクチャ（BBCA: behaviour-based control architecture）」では、システム全体は、機能によってではなく、それぞれ固有の「活動」、すなわちタスク達成行動によって、幾つかのサブシ

ステムに分割される。「層」と呼ばれるこれらのサブシステムは、それぞれ独自の入力と出力を具えており、他の層とは独立に環境と相互作用することができるように作られる。それぞれの層は、高次の層が低次の層の出力を「抑止」できるように連結される（ただしこの際、低次の層は出力を出し続けるのであり、正確には高次の層がシステム全体の出力を上書きする、と言った方が良い）。これを高次の層が低次の層を「包摂する」と言い、ここから BBCA は、「包摂アーキテクチャ（SA）」とも呼ばれる。例えば、ブルックスが製作した昆虫型移動ロボットは、［1］ロボットの対象回避行動を実行する「回避」層、[2]回避層が作動していないときにロボットをランダムに歩き回らせる「回遊（wander）」層、［3］回遊層を抑止し、ロボットを離れた場所に到達させようとする「探索」層の三層から成る。

ブルックスは方法論上の指針として、各層を製作する際には、それを現実世界の中で作動しうる完全なシステムとして構築すること、を挙げている。新たな振る舞いを生み出したい場合には新しい層を積み上げる、という仕方で、ロボットは、単純な能力から出発して次第に高度な能力を身に付けていく、という進化のプロセスをなぞる。

SA においてシステムを構成する各層は、CCA の直列型情報処理と異なり、低次の層は高次の層の振る舞いに依存せずに作動し続ける、という仕方で並列的に処理を行なう。こうしたアーキテクチャのおかげで、環境の状況に応じた行動が実現されるのであり、SA に従って製作されたロボットは環境の変化に対して頑健性を示す。また、SA においては、各層は直接に環境と相互作用するのであるから、情報が一旦中央システムに集められ、処理される必要のある直列型のアーキテクチャに比べて、処理や制御がコンパクト化され、ロボットが迅速に動作することが可能になる。

SA における各サブシステムの間には中心／周縁の区別はない。知能の核心と見なされた中央集権的な認識ボックスを取り払うという決断によって、ブルックスは、大きな作業負担となっていた世界の内部モデルの構築という課題を放棄し、実際にロボット製作に着手することができた。こうして出来上がったロボットは、「従来の手法で作られたどのロボットよりも優れた動きを見せた」(Brooks, 2002, 邦訳, 72 頁)。

1.3 ブルックスの反表象主義テーゼ

SA 理論に従った一連の移動ロボット製作からブルックスが得た教訓は、次の「予期せざる結論（C）」と「ラディカルな仮説（H）」の 2 点である。(Brooks, 1991, 邦訳, 86 頁)

C：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わかる。つまり、世界をそれ自身のモデルとして用いるほうがよ

い、ということが判明する。

H：表象は、知的システムの最も扱いにくい部分を構築する際にはうまく働かない抽象化の単位である。

「抽象化」とは、入力データを、関連性を有するもののみに還元する方法である。この例が、シェイキーにおける「積木世界」であり、こうした単純化された世界の上にそのモデルが構築される。しかし、ブルックスはこうした抽象化を「危険な武器」であると断じる。抽象化は、当の知的システム自身によってではなく、研究者の手によって遂行されてきた。こうした研究者による世界の抽象化、すなわち環境内で起こりうる状況をシミュレーションする作業が膨大なものになってしまうことについては既に触れた（第 1.1 節）。

また、真に知的なシステムならば、研究者の手によって遂行された抽象化の上に与えられる世界の内部モデルなどを用いることはせず、むしろそれは、何度も繰り返し環境の知覚を遂行することを通じて、環境そのものをモデルとする、と考えるべきである。実際、世界の内部モデルを用いることは、環境の変化に対する頑健性や作業の迅速さに関して不利である（第 1.1 節）。

しかし、ブルックスの提出した仮説（H）は、より一般的な主張を行っている。ブルックスの移動ロボットには中央システムが存在しないのだから、そこに投影されるべき中心的表象もまた存在しない。さらにブルックスは、「局所的なレベルにあってさえも、伝統的な AI の表象があるわけではない」(ibid., 95 頁)、と主張する。その理由は、彼のロボットにおいて実際に行われているのは、あるプロセスから別のプロセスへの数値の受け渡しにすぎず、ブルックスらは「何らかの意味論を付与しうるようなトークンを用いているわけではない」(ibid.)からである。

また、ここには、システム全体とその状態を別の領域へと適当に写像することによって、プロセス相互間の連結とそれらの数値とが何らかの仕方でコード化している表象として定義されるような、潜在的な表象が存在するのだ、という反論をもブルックスは認めない。理由は、このように通常の表象とあまりに多くの点で相違するものを表象と呼ぶことに意味がないからである。

こうしてブルックスは、工学的な関心に導かれながら、彼の製作した移動ロボットの設計原理の記述から進んで、知的システムにいかなる意味であれ表象というものを帰属させるようないかなる解釈をも拒絶する、「反表象主義」の名に相応しい一般的な主張を提出することになるのである。

2. 力学系アプローチと反表象主義批判

2.1 ヴァン・ゲルダーの力学系アプローチ（DSA）

SA 理論に基づいて製作されたブルックスの移動ロボットが、シェイキーのような古典的制御アーキテクチャに基づく従来のロボットに比べて、現実世界における動作性能の点で優れたものであったことは確かである。また、従来のロボットにおいて、世界の内部モデルの存在が様々な問題を引き起こす足枷となっており、これを取り払ったことがブルックスの成功の要因となったことも認めてよいだろう。しかし、これらの実践的な診断と、〈知能にとって表象は不必要である〉という一般的主張の妥当性とは、差し当たり別に考えられなければならない。

さて、この一般的主張に対する批判を詳しく吟味する前に、認知システムを表象なきシステムとして説明するための概念的枠組みを与える、ヴァン・ゲルダーの認知に対する「力学系アプローチ（DSA）」について触れたい(van Gelder, 1995)。

ヴァン・ゲルダーの目的は、認知に関して、計算主義アプローチと競合しうる代案を提示することにある。ここで引き合いに出されるのが、数学の「力学系理論」に基づく、認知に対する DSA である。「力学系」とは、互いに影響を及ぼし合いながら時間と共に変化する複数の状態変数から成るシステムである。システムの状態は、「状態空間」と呼ばれる抽象的幾何学的空間を通る軌道として描き出される。認知に対する DSA は、力学系理論の数学的道具を認知システムに適用することによって、認知を解明しようとする。それによれば、認知とは、力学系を構成する諸状態の、規則に従った相互依存的な発展である。システムの発展規則は、微分方程式のような数式によって与えられる。

ヴァン・ゲルダーは、計算システムが持たなければならない性質を幾つか挙げているが、その中でも、記号が他の事態の代用をする、という「表象」こそが、最も枢要な性質と見なされる。

これに対して、ヴァン・ゲルダーが力学系の例として引き合いに出すのが、ワットの遠心調速機である。調速機とは、蒸気機関において、ボイラーからピストンへ蒸気を運ぶ管の中にある絞り弁を調節することで、ピストンへ流れ込む蒸気の量を調節し、フライホイールの速度を一定に保つ装置である。ワットの遠心調速器の特徴は、フライホイールと絞り弁とが、垂直車軸とそれに留められた二本の腕を介して連動しており、素早く円滑に蒸気機関が速度を維持できる点にある。

ヴァン・ゲルダーによれば、遠心調速機は表象を用いていない。一見したところ、腕の角度がフライホイールの速度を表象していると考える誘惑に駆られるが、これは、以下の理由によって不適切である。[1]このシステムを、表象を用いて説明することには、説明上

の利点がない。実際、遠心調速機の働きを説明する際、こうした説明の中で表象については言及されない。[2]腕の角度と蒸気機関の速度の間に単なる相関関係があるというだけでは、それが表象関係であるということにはならない。[3]腕の角度と蒸気機関の速度との間に相関関係があると語ることは不適切である。両者の相関関係が断たれている場合でさえ、調速機は依然働き続ける。[4]腕の角度と蒸気機関の速度は、常時互いの振舞いを規定し、また規定されるという複雑精妙な関係にあり、これを表象関係によって処理することはできない。

時間と共に影響を与え合う二つの項の相互関係は、力学系理論の用語で「カップリング」と呼ばれる。腕の角度と蒸気機関の速度の関係は、一方向的な表象関係ではなく、相互的なカップリング関係として捉えられるべきである。

遠心調速機は、表象を用いないのみならず、これに依存する計算システムのいずれの性質をも持っておらず、従ってこれを計算システムと見なすことはできない。これに対して、力学系理論の概念的道具は、遠心調速機の働きを記述するのに適している。ヴァン・ゲルダーの結論は、認知システムは、計算システムとは性質を異にする力学系であり、計算システムの代表例であるチューリング・マシンよりも、力学系の代表例である遠心調速機の方に似ている、というものである。

さて、ヴァン・ゲルダーが提出した表象の基準からして、ブルックスの移動ロボットにも、表象は用いられていないと見なすべきである。実際、ブルックスのロボットの認知を説明するに当たって、表象は説明上の役割を演じてはいない（論点[1]）。ブルックスのロボットは、遠心調速機に類似した、力学系の一種と捉えることが出来る。

2.2 DSA の認知観

計算主義と DSA とは、その認知観において決定的に相違する。概して言えば、計算主義によれば、認知システムは、「脳すなわち身体内部に座を占める一種のコントロール・ユニット」(van Gelder, 1995, 邦訳, 185 頁)であるのに対して、DSA によれば、これは脳 - 身体 - 環境の三者を含んだ単一の統合されたシステムであると言うことが出来る。これは、力学系理論において、複数のシステムがカップリングし合い、相互依存的に発展する場合、これらを単一のシステムと見なしてよい、という事実に基づく。

より詳しく、両者の相違を踏まえて、DSA の認知観が計算主義のそれよりも優れている点を挙げるならば、以下のようになる。

[1]認知の時間的側面。計算システムは、各動作に要する所要時間と、これらの動作が起こる順序以外の点で、各動作がいつ行われるかに関して制約を持たない。これに対して、

力学系では、すべての動作は実時間の枠内で生じ、時間的制約を持つ。すなわち、認知が実時間を通じて進行するという側面を捉える上で、計算主義では不十分であり、DSA こそが適切である。

[2]認知の創発的側面。計算システムでは、システム全体の振舞いが各モジュールの振舞いの総和に過ぎないのに対して、DSA では、単純な要素の局所的相互作用から、システム全体の組織された複合的な振舞いが「創発」してくる様を捉えることが出来る。

[3]身体／環境との相互作用。計算主義において、認知システムがやり取りするのは記号表象だけであるため、身体と環境はシステムの外部として考慮の外に置かれる。これに対して、DSA では、脳 - 身体 - 環境は、一つの統合されたシステムとして扱われ、認知が身体／環境を含めた場の中でいかに生じるかを捉えることが出来る。

2.3 クラーク＆トリビオの反表象主義批判

反表象主義に対して、代表的な批判者であるクラーク＆トリビオは、次の二点から批判を試みる(Clark & Tribio, 1994)。

[1]反表象主義者は、表象一般という概念と、そのサブクラスであるところの明示的表象という概念とを混同している。認知の説明において明示的表象が役割を演じていないということは、認知の説明に表象一般が必要ないということを帰結しない。

[2]反表象主義者は、表象の役割が相対的に小さい問題領域のみを扱っており、表象に飢えた問題領域を扱っていない。

クラークらは、まず何を持って表象と見なすかに関して、ホーグランドによる表象の定義を援用している。ホーグランドから引用しよう(Haugeland, 1991, p. 62)。

> （……）もし関連する特徴が必ずしも呈示されていなくとも、これらの特徴は、少なくともある場合には、表象されうる。すなわち、何か他のものが、これらの特徴の代わりに行動を導くことができるように、これらの特徴の代役となりうる。こうした仕方で何か他のものの代役となるものこそが、「表象」である。

このような定義に依拠するならば、計算主義の明示的表象のみならず、コネクショニズムの非明示的表象（「分散表象」とも呼ばれる）もまた、内的表象という一般的な概念の下に包摂される。ヴァン・ゲルダーが認めるように、コネクショニスト・システムは、力学系の下位カテゴリーと見なすことができる。そうすると、あるシステムが力学系であることは、本質的には、これが非表象的システムであることを含意しないと言うことができる。

ブルックスが主張するように、システムが環境に適応した振舞いを示すために、環境を明示的に表象するという戦略が有効ではないとしても、表象主義の枠内で、コネクショニズムの解決法を採用することによって、環境とのカップリングを達成することが出来る。

クラークらは、遠心調速機が表象を用いない力学系理論の枠組みによって最もうまく説明されることは認める。しかし彼らは、ブルックスやヴァン・ゲルダーの取り上げている問題領域が、「例外なく、周囲の環境に適切な刺激が存在し、それらの刺激に内的表象に変わる役割を押しつけることのできるような領域」(Clark & Tribio, 1994, 邦訳, 231 頁)であり、「表象に飢えた」問題領域ではない、と指摘する。これに対して、表象に飢えた領域には、次のいずれかもしくは両方を充たす一切の領域が含まれる(ibid.)。

[a]現前しない、または存在しない対象、あるいは反事実的な事態についての推論が必要である。

[b]周囲の物理的環境において複雑で扱いにくいかたちで（たとえば、閉じない選言で）現れるパラメータにたいして、選択的に反応する能力が主体に要求される。

こうした条件によって規定される問題領域は、認知領域の「氷山の一角」であるどころか、認知の様々な事例を含む。従って、ブルックスの移動ロボットや遠心調速機の振舞いが、こうした高次の認知領域を扱いうるものでない限り、反表象主義の主張は維持できない。

クラークらは、反表象主義者たちが明示的表象の役割に疑問を呈したことには賛同する。しかし、上記の論点を踏まえて彼らが導き出す結論は、なおも表象主義の枠内で、コネクショニズムの非明示的表象のような「控えめな表象」を用いた、「表象ありの DSA」という折衷案によってこそ、認知はより良く説明される、というものである。彼らが、「表象の概念は二分法としてではなく、連続体として再構成される」(ibid., 243 頁)と言うのは、認知の場面に応じて表象の演じる役割が程度の差を持つ、という意味である[(1)]。

2.4 反表象主義批判の検討

クラークらの批判を反駁する試みの一つは、彼らの批判において表象が不可欠と見なされた高次認知の領域に対して、表象なしの説明を与えることである。このような取り組みの一例として、手に持った杖で遠くにある対象に届くかどうか、という判断を扱ったヘイスレイガーらの研究がある(Haselager, Bongers & van Rooij, 2003)。このタスクは、未来の行為の結果について予測することを要求するため、クラークらの言う「表象に飢えた」問題領域に含まれる（第 2.3 節の論点[a]）。しかし、ヘイスレイガーらは、この判断が、アトラ

クタを二つ用いた一種の力学系モデルによって、表象を用いることなく分析可能であることを示す[(2)]。ここでは、同様の判断が表象を用いた説明によって処理可能である可能性は排除されないが、彼らによれば、DSA の説明の方が、表象の不在という点で、より経済的であり、それゆえにより魅力的である(ibid., p. 239)。

反表象主義からの再反論としては他に、「表象ありの DSA」が実際に表象を用いていることを否定する立場が考えられる。なるほど、ホーグランドの定義によれば、コネクショニズムの非明示的表象のようなものも「表象」に含まれることになる。しかし、ヴァン・ゲルダーがあるシステムを表象システムと見なすべきかどうかの規準として提出した、説明上の利点という論点（第 2.1 節の論点[1]）を考慮するとき、この点は検討の余地を残していると言える。

服部(2003)によれば、人間の言語理解のような認知活動には統語論的構造を持った表象が必要であり、こうした高次認知を説明する上でこれについて「表象」の処理、という語り方をすることは不可欠である。しかし、コネクショニズムの分散表象と呼ばれるものについて、これを「表象」と呼ぶことによって認知の説明に何かが付け加わる訳ではない。従って、分散表象は「表象」と呼ぶべきでない（ただし、服部の「表象なしのコネクショニズム」は、認知の説明において計算主義的な表象の役割を認める点で、反表象主義の立場の対極に位置すると言える）。

谷(2005)は、工学的な見地から、同様の主張を行なっている。谷が研究しているような移動ロボットのナビゲーションにおいては、ある種の環境の内部モデルが必要になると想定される。この際、従来のアプローチでは、例えば迷路環境において多様な道筋の組み合わせを心的に操作しなければならないことから、合成可能性を有した記号表象の存在が不可欠と考えられた。しかし谷は、リカレント・ニューラル・ネットワーク（RNN）を用いることで、上のような組み合わせ能力を実現できるのであり、この際表象による説明は必要ない、と言う(ibid., 95-96 頁)。

> （……）内部システム自身は、何かの表現を上から眺めて、それに従い作動しているわけではない（……）。内部システムは、その心的過程としてただ力学的作動を繰り返すのみであり、それを第三者が外から観察した場合、あたかも、記述された表現があってそれが操作されているように見えるだけなのである。認知の現場に存在するのは、力学的構造とそれがもたらす系の時間発展のみと考えるのが自然であろう。

明示的表象を用いた計算システムとは異なり、コネクショニスト・ネットワークにおい

ては、表象の処理として特定されるべき個別的なプロセスは存在しない。コネクショニスト・システムの振舞いそのものは(谷の研究におけるように、ある種の内部モデルさえも)、力学系理論によって十分に説明可能である。なるほど、この際に、各ユニットの活性化パターンから成るシステムの全体的物理状態がシステムの全体的認知状態（表象）を潜在的・傾向的に実現している、と語ることは不可能でない(cf. 美濃, 2003)。問題は、システムの振舞いを汲み尽くしうる力学的説明に対して、敢えて表象概念を用いた説明のレベルを付け加えることに意味はあるのか、ということである[(3)]。

3. 結論

本稿の出発点は、現実世界を自律的に動き回ることのできるロボットを製作する、というブルックスの企図にあった。こうした具体的な目的に沿った形で、反表象主義を巡る議論を総括し、これに一応の結論を下したい。

ブルックスの反表象主義を動機づけていたのは、上記のような目的を達成しようとすると、記号のような統語論的構造を持った明示的表象を採用する古典的な表象主義（計算主義）が限界に突き当たる、という診断であった。「表象なしの DSA」は、このようなシステムに力学系理論という説明の枠組みを与え、低次の場面から高次の場面まで、認知を統一的に説明しようとする。これは、「オッカムの剃刀」の原理に則って、認知の説明から表象概念を追放し、表象を中心概念とする現行の認知理論を根本から覆すものと見なされる。

非明示的表象の概念と「表象に飢えた」問題領域の存在に訴えてこれを批判しようとするクラークらの議論は、前節で見たように、成功しているとは言えない。というのも、ヘイスレイガーらが明らかにしたように、彼らが表象なしの説明の限界として設定したものは恣意的であり、高次認知と見なされるものについても表象なしの説明が及ぶ可能性は原理的に排除されない。また、非明示的表象を「表象」に分類することは、ホーグランドの定義のみを論拠とし、その利点を明らかにしていないがゆえに、説得力に欠ける。

しかし、クラークらの批判を斥けたことによって反表象主義の主張が擁護されたと考えるのもまた、早計である。なるほど、表象が不可欠と見なされた高次認知の領域に対して、表象なしの説明を与えようとするヘイスレイガーや谷の試みは、有益ではある。しかし、これら高次認知の説明がどこまで力学系的説明に還元できるかは経験的問題である。この際、むしろ、このような事例研究を積み重ねることによって反表象主義の一般的主張を正当化するのは、認知領域の広範さから言って、困難であろう。

また、表象レベルの説明も、ブルックスやヴァン・ゲルダーの批判によって一挙に葬り去られた訳ではない。ブルックスが言うように、知能の基礎的レベルで表象が不要として

も、「知能の進化」の過程で明示的表象が有効になる段階に達することは大いに考えられる。さらに、DSA の説明が高次認知の領域をも汲み尽くせると仮定してなお、力学系的説明と表象を用いたそれとが異なるレベルの説明として両立しうることも想定可能である。この際、「オッカムの剃刀」を振りかざす前に、表象レベルの説明の意義を再検討しなければならない。本稿ではこの点に深く踏み込むことはできないが、表象主義側の言い分を要約すれば、システムのダイナミクスに意味論を与えるために表象レベルの説明が必要になる、とされている(4)。

ヘイスレイガーらは「本当にそれが必要でない限り、説明とモデル化において表象を用いてはならない」(Haselager et al., 2003, p. 232)というモットーを提案している。上に挙げた点を考慮し、反表象主義を巡る議論に暫定的な結論を下すとすれば、Anti-representationalism ではなく、ヘイスレイガーらのモットーに端的に言い表されているような、言わば'Less-representationalism'こそ、AI を含む認知の経験的探求の現場で採用すべき妥当な態度と言えるだろう。

ブルックスが SA 理論によって試みたのは、外から環境のモデルを与える代わりに、知覚系と行為系とが一体を成した身体の動作によって、システムそれ自身に環境を獲得させることであったと言える。彼のこうした戦略は、ロボット製作において有効であっただけでなく、そこから帰結する、身体と環境の相互作用をその基礎に据えた知能／認知観が魅力的であるがゆえに、多くの論者に高く評価された(5)。しかし、力学系理論によって認知を統一的に説明し、我々の認知観を根本的に変革する、という反表象主義の夢は、幾つもの困難な課題を残しており、未だ単なる夢に留まると言わざるを得ない(6)。

註

(1) ヴァン・ゲルダー自身が、反表象主義というよりは、表象の本性についての見直しを迫る折衷論に属すと思わせるような譲歩を示している。「(……) じつは、力学系が何らかの表象形式を組み込むことを妨げるものは何もないのである。それどころか、力学系アプローチの目覚ましい特徴は、広い意味で非計算的といえる枠組みの内部でさえ、認知システムにおける表象の本性を劇的に考え直すための機会を提供してくれることである」(van Gelder, 1995, 邦訳, p. 189)。解釈に困る。
(2) 「アトラクタ」とは、力学系理論において、システムの状態がそこに収束する状態空間上の点である。
(3) このようにコネクショニズムにおいても「表象」概念を用いるがために、計算主義の側から、コネクショニズムは計算主義の実現形態の一つにすぎない、として批判される結果につながる、という(服部, 2003, 美濃, 2003)。
(4) しかし、これに対しては、システムの振舞いに意味論的解釈を加えるのは、システムそれ自身ではなく観察者であり、システムそれ自身はやはり表象を用いていない、という反論が成り立つようにも思える。
(5) ブルックスの研究は、知的振舞いには人間の身体に固有の非形式的な構造が不可欠であり、コンピュータはこうした身体を持たず、また持ち得ないがゆえに、AI 研究は失敗を運命づけられている、というドレイファスの批判(Dreyfus, 1979)を乗り越え、AI の可能性を実証したと言える。
(6) 本稿のタイトルについて。ロボットが夢を見るならば、それが表象を用いていないと言うのは難しい。

文献

Brooks, R. A. (1991). ‘Intelligence without representation’, Artificial intelligence, 47, 139-159. (1990, 柴田正良訳, 「表象なしの知能」, 『現代思想』, 1990 年 3 月号, 85-105.)

——— (2002). Flesh and machines: How robot will change us, New York: Pantheon Books. (2006, 五味隆志訳, 『ブルックスの知能ロボット論—なぜ MIT のロボットは前進し続けるのか？』, オーム社)

Clark, A. & Tribio, J. (1994). ‘Doing without representation?’, Synthese 101, 401-431. (2002, 金杉武司訳, 「表象なしでやれるのか？」, 門脇俊介・信原幸弘編, 『ハイデガーと認知科学』 (205-251 頁), 産業図書.)

Dreyfus, H. L. (1979). What computers can’t do: The limits of artificial intelligence (Revised edition), Harper & Row. (1992, 黒崎政男・村若修訳, 『コンピュータには何ができないか—哲学的人工知能批判』, 産業図書.)

Haselager, P., Bongers, R. & van Rooij, I. (2003). ‘Cognitive science, representations and dynamical systems theory’, In W. Tschacher & J.-P. Dauwalder (Eds.), The dynamical systems approach to cognition: Concepts and empirical paradigms based on self-organization, embodiment, and coordination dynamics, Studies of nonlinear phenomena in life science vol. 10 (pp. 229-241), Singapore: World Scientific.

服部裕幸 (2003). 「「分散表象」は認知の説明にはたして役立つのか？」, 戸田山和久・服部裕幸・柴田正良・美濃正編, 『心の科学と哲学—コネクショニズムの可能性』 (29-51 頁), 昭和堂.

Haugeland, J. (1991). ‘Representational genera’, In W. Ramsey, S. Stich & D. Rumelhart (Eds.), Philosophy and connectionist theory (pp. 61-90), New Jersey: Erlbaum.

人工知能学会（編） (2005). 『人工知能学事典』, 共立出版.

美濃正 (2003). 「新しい認知の理論としてのコネクショニズムの可能性」, 戸田山・服部・柴田・美濃編, 『心の科学と哲学』 (79-114 頁), 昭和堂.

中村雅之 (2004). 「表象なき認知」, 信原幸弘編, 『シリーズ心の哲学 ロボット篇』 (85-117 頁), 勁草書房.

谷淳 (2005). 「身体・行為・記号—認知ロボティクスの立場から」, 『現代思想』, 2005 年 2 月号, 87-97.

van Gelder, T. (1995). ‘What might cognition be, if not computation?’, The journal of philosophy, 91(7), 345-381. (2002, 中村雅之訳, 「認知は計算でないとすれば、何だろうか」, 門脇俊介・信原幸弘編, 『ハイデガーと認知科学』 (151-203 頁), 産業図書.)

〔京都大学大学院博士課程・哲学〕