

新進研究者 Research Note
コネクショニスト認識論の批判的検討
A Critical Study on Connectionist Epistemology

藤原諒祐

Abstract

In this paper, I briefly review Paul M. Churchland's new approach to epistemology, and give methodological consideration to naturalized epistemology. He criticizes what he calls "ideal sentential automaton (ISA) approach", and proposes new approach to epistemology which uses connectionist networks as models of cognition. I show that this new approach is not definitely superior to ISA approach, and suggest that methodological pluralism is better for naturalized epistemology.

(1) 研究テーマ

ポール・M・チャーチランドのコネクショニスト認識論の批判的検討を行い、その利点と欠点を明らかにする。そして、自然化された認識論にとって適切な認知モデルの選択について考察を行う。

(2) 研究の背景・先行研究

2.1. 自然化された認識論，科学的心理学と認識論・科学哲学

知識がいかにして獲得・形成されるのか，そして知識はいかにして正当化されるのか。伝統的認識論はこれらの問題にとりくむことで自然科学の基礎付けを行うことを目指した。基礎付け主義的な伝統的認識論に対し，W・V・O・クワインはそれに代わる「自然化された認識論」を提示した(Quine, 1969)。自然化された認識論は自然科学の基礎付けを目的としない。むしろ，認識論は科学と連続的な営みとして捉えられる。自然化された認識論は，第一哲学という位置づけを捨てることで，自然科学，特に科学的心理学の知見を用いることができる。科学的心理学の発展にも促され，クワイン以降様々な論者が自然化された認識論のパラダイムに属する研究を行ってきた。

アルヴィン・I・ゴールドマンは科学的心理学が知識の獲得・形成の解明だけでなく，知識の正当化という規範的問題に対しても大きな影響を与えることを論じた(Goldman, 1986)。ゴールドマンは心理学的知見をもとに信念形成プロセスの信頼性を評価することで，正当化という規範的問題に答えよう

とした。認識論に寄与しうる科学分野は実験的な心理学に限られない。ポール・サガードは AI 研究の手法を用いた「計算論的科学哲学」を展開した。サガードは科学者の問題解決をコンピュータ上でシミュレートし、科学理論の本性や科学的説明、帰納的推論にかんする科学哲学上の諸問題へのアプローチを試みた (Thagard, 1988)。

これらの研究は心理学、認知科学、AI 研究といった分野の知見が認識論に貢献することを示すものであったが、本稿との関係で特筆すべきは、彼らの研究がいずれも認識論において命題的態度に対応する文的な構造が一定の役割をもつことを認めていることである。ゴールドマンの信頼性主義では知識は信念の一形態とされているし、サガードも、問題解決のシミュレートのために構文論的な構造を用いている。

2.2. 心の哲学における消去主義

認識において文的な構造が役割を果たすという素朴な直観の妥当性については、心の哲学の分野で議論がなされてきた。チャーチランドは、フォークサイコロジー (folk psychology, FP) は誤った理論であり、FP が措定する信念や欲求などの命題的態度は科学的心理学の発展に伴い消去されると主張した (Churchland, 1981)。チャーチランドの立場は消去主義と呼ばれる。

さらに、認知科学におけるコネクショニズムの登場により、消去主義に新たな理論的根拠が与えられることになった。ウィリアム・ラムジー、ステイヴン・スティッチ、ジョセフ・ガロンは、コネクショニスト・モデルには命題的態度に対応するような構造物は存在しないと論じ、コネクショニズムと FP が相容れないものであると主張した (Ramsey, Stich and Garon, 1991)。コネクショニスト・モデルでは相互に結合したノード (ニューロン) によって入力処理され、各ニューロンの興奮パターンの変化によって様々な命題が分散的に表象される。したがって、各命題は同じ構造物によって表象されることになり、個々のニューロンが独立に意味論的な解釈を受けることもない。一方、ラムジーらによると、命題的態度は機能的に相互に区別され独立に意味論的に解釈可能な状態であるため、命題的態度に対応する構造物はコネクショニスト・モデルには見いだせないことになる¹。

自然化された認識論は科学的心理学と連続的であるため、消去主義の主張は心身問題だけでなく自然化された認識論にも影響を与えるだろう。もし科学的心理学が命題的態度 (特に信念) の存在を措定しないならば、自然化された認識論は信念あるいはそれに類する文的な構造の運動学としての認識論という企てを捨て去り、全く新たなアプローチをとらざるを得なくなる。

2.3. ポール・M・チャーチランドのコネクショニスト認識論

チャーチランドは FP の理論としての貧困さから、FP に基づくアプローチをとる認識論を批判した。(Churchland, 1979). チャーチランドは、FP に基づく文的な認識論のアプローチを理想的文オートマトン (ideal sentential automaton, ISA) アプローチと呼ぶ。ISA アプローチでは、認知状態が文や命題の集合によって表示されることが前提となる。チャーチランドによれば、FP が認知の理論として良い理論とはいえない以上、FP を前提とする ISA アプローチは否定されねばならない。

ISA アプローチをとる認識論のかわりに、チャーチランドは認知のモデルとしてコネクショニスト・モデルを採用した認識論を展開する (Churchland, 2012). チャーチランドのこのような認識論を本稿ではコネクショニスト認識論と呼ぶ。コネクショニスト認識論において、認識のための概念枠組みや理論体系は命題の集合ではないとされる。コネクショニスト・モデルは入力層、隠れ層、出力層に分かれているが、コネクショニスト認識論では隠れ層の各ニューロンによって構築される多次元の活性ベクトル空間により概念枠組みが与えられるとされる。つまり、この多次元ベクトル空間を分割する多面体が、概念に対応するプロトタイプ領域となっている。

例として、隠れ層が三つのニューロンからなるネットワークを考えよう。ネットワークに入力が与えられると、この三つのニューロンの活性度合いの順序対 $\langle a, b, c \rangle$ によって決まる活性ベクトルが三次元空間にプロットされる。このベクトルが「男性の顔」に対応する多面体の内部に位置していれば、ネットワークはこの入力に対し「男性の顔」という出力を返す。注目すべきは、この処理過程において構文論的構造の操作はなされていないことである。この特徴のため、言語を解さない動物や乳幼児の認知状態に対してもコネクショニスト・モデルを用いて記述を与えることが可能になる。

さらに、コネクショニスト認識論はゴールドマンの信頼性主義を取り入れることができるとチャーチランドは論じる (Churchland, 2007a). ある特徴 (色, 顔) に対応する概念枠組みとしての多次元活性ベクトル空間 (色空間, 顔空間) は、対応する特徴をよく捉えているならば信頼できるといえる。そして信頼できるベクトル空間を備えたネットワークから得られる活性ベクトルが外界をうまく表象できている場合、その活性ベクトルは知識とされる。

コネクショニスト認識論はコンピュータ上でのシミュレーションとの親和性が高いことにも注目したい。コネクショニスト・モデルはコンピュータ上でシミュレート可能であり、この点でコネクショニスト認識論はサガードの

計算論的科学哲学と同じ特徴をもつと言える。一方で、サガードの PI システムが文や単語によって構成されるデータ構造の操作に基づくものであり、古典的計算主義の枠組み内の研究である点は、コネクショニスト認識論と対照的である。認知科学のパラダイムとしての古典的計算主義とコネクショニズムの優劣についての論争には決着がついていないが、コネクショニスト・モデルは実際の脳の解剖学的構造を模している点で、古典的なモデルよりも解剖学的な現実性が高いと考えることができよう。

(3) 筆者の主張

以上を念頭に置いてコネクショニスト認識論の検討にうつりたい。検討に当たっては、ゴールドマンによる第一認識論 (primary epistemology) と第二認識論 (secondary epistemology) の区別を導入したい (Goldman, 1986)。ゴールドマンは、認識論の対象として、知覚や記憶などの、認知の基礎をなす心理プロセスと、アルゴリズムや発見法、学習可能な方法論といったメソッドを区別する。そして、プロセスをあつかうものを第一認識論、メソッドをあつかうものを第二認識論と呼ぶ。

この区別の理由の一つは、第一認識論が個人の認知だけを扱うのに対し、第二認識論は社会的な側面を扱う必要があるという違いがあることだ。例えば、ガリレオの望遠鏡による観測というメソッドの正当性は、現在と当時では異なるかもしれない。というのも、今となっては時代遅れの望遠鏡でも、当時は最先端の観測器具だったことを考えると、ガリレオは最も信頼できるメソッドを用いていたといえるからだ。このように、あるメソッドの正当化を扱う際には、個人の認知メカニズム以外に、時代状況や共同体内での他メソッドの受容の有無など、歴史的・社会的要素を考慮に入れる必要がある。一方、プロセスの正当性を考慮するためには、このような歴史的・社会的要素を考慮に入れる必要はなく、そのプロセスについての認知科学的な知見をもとに信頼性を評価すればよい。

本稿では第一認識論と第二認識論を規範的問題だけでなく、記述的な問題にもとりくむものとして考える。第一認識論と第二認識論を区別することで、コネクショニスト認識論に対して異なる次元から検討を行うことができる。

3.1. 第二認識論としてのコネクショニスト認識論

歴史的・社会的要素を考慮に入れるという第二認識論の特徴を考えると、コネクショニスト認識論だけでは第二認識論としては不十分であることがわかる。コネクショニスト認識論は脳内プロセスのモデルを用いて認識論的諸

問題にアプローチするが、第二認識論では、脳の外側の事実についても考慮に入れる必要があるからだ。また、第二認識論に部分的に寄与する理論としても、コネクショニスト認識論には課題が存在する。

例えば、科学理論や実験手法といったメソッドの産出・改良を可能にするメカニズムの解明も第二認識論の目標となるだろうが、このためには科学者共同体の研究評価体制や科学教育制度などの社会的要素の考察が必要になるだろう。そしてここでは、言語や記号といった外的表象が重要な役割を果たす。したがって、コネクショニスト認識論がメソッドに対してアプローチするためには、脳内プロセスのモデルに加えて、脳内プロセスと外的表象との関わり合いをその理論に組み込む必要がある²。

また、非言語的なコネクショニスト・モデルによって動物や幼児と成人の認知に統一的な説明を与えるという点は、第二認識論としては問題がある。なぜなら、メソッドに含まれる多くの発見法や方法論、特に科学者が用いる実験や推論の手法は動物や幼児には使用不可能であるからだ。このようなメソッド使用の特徴を説明するためには、動物・幼児と成人との差異を説明できなければならない。したがって、コネクショニスト認識論が第二認識論のよいアプローチとなるためには、動物・幼児と成人の間の非連続性を説明するという課題にとりくむ必要がある。

3.2. 第一認識論としてのコネクショニスト認識論

次に、第一認識論としてのコネクショニスト認識論を検討する。チャーチランドが例に出す色知覚にかんするコネクショニスト・モデル（色知覚ネットワーク）を考えてみよう。色知覚ネットワークは人間の視覚に関係する神経構造を模倣しているため、解剖学的現実性が高いと考えられる。また、色知覚ネットワークの活性ベクトル空間（色空間）上の位置関係によってそれぞれの色の現象的な類似性や対称性を説明することができる。さらに、色知覚ネットワークから、自然に存在しない色の経験が可能であるという予測と、その経験を可能にする簡単な実験を引き出すことができる（Churchland, 2007b）。知覚プロセスに対して現実に即した説明と予測を与えられる点は、ISA アプローチに基づく認識論に対してコネクショニスト認識論がもつ大きな利点であるだろう。

しかし、より高次の認知にかかわる心理プロセスを考慮に入れると、コネクショニスト認識論の優位性は怪しくなってくる。チャーチランドは色概念のような知覚的概念と同様に、理論的な概念についても多次元活性ベクトル空間上のマッピングによって説明する。したがって、理論的概念の把握と知

覚的概念の把握は同様のプロセスに支えられていることになる³⁾。しかし、色空間を担うニューロン群とは異なり、理論的概念空間を担うニューロン群が脳内に存在するかは明らかではない。そのようなニューロン群が存在するとしても、その脳内における位置づけは不明である。特に、複数の相容れない理論に精通する人（例えば、天文学者であると同時に天文学史に精通する科学史家）の場合では、複数の理論に対応する複数のニューロン群が脳内にどのように位置するのかは判然としない。このような事情を考えると、現段階では理論的概念がマッピングされる多次元空間は、ISA アプローチが提示する構文論的構造と同様に抽象的なモデルに過ぎないといえる。したがって、理論的概念の把握といった高次の心理プロセスにおいては、コネクショニスト認識論が解剖学的現実性という点で ISA アプローチよりも優れているかどうかは定かではない。

そもそも、認識論の評価において解剖学的現実性が重視されるべき理由はない。例えば、複雑な概念の把握や記憶は認識論的に重要な心理プロセスだが、その脳神経科学的なメカニズムが十分解明されていないため抽象的なモデルを用いざるを得ない。また、仮に複雑な心理プロセスについて解剖学的現実性が極めて高いモデルを提示できたとしても、そのモデル自体が非常に複雑なものとなるために心理プロセスの理解に寄与しないという可能性がある。その場合、心理プロセスのよりよい理解のためには、抽象的なモデルを用いる方が有効であるかもしれない。したがって、少なくとも高次の心理プロセスにおいては解剖学的現実性という指標を用いて認識論の優劣を評価するのではなく、抽象的なモデルを用いることで得られる説明力という観点から評価すべきである。そして、この点でのコネクショニスト認識論と ISA アプローチの優劣は決定的とはいえない。

3.3. 方法論的多元主義の可能性

以上の考察からどのような示唆が得られるだろうか。まず、第二認識論に寄与する理論としてのコネクショニスト認識論には、外的表象をモデルにとりこむという課題と動物・幼児と成人の非連続性を説明するという課題がある。もちろん ISA アプローチをとる認識論も同じ課題にとりこむ必要があるが、言語に類似する構造を措定する ISA アプローチは文字や記号といった外的表象を説明のモデルに組み込みやすい。したがって、第二認識論としてはコネクショニスト認識論よりも ISA アプローチのほうがよい理論を提供する見込みがある。

第一認識論としてみても、コネクショニスト認識論が ISA アプローチより

も決定的に優れているとはいえない。理論的概念の把握や記憶といったプロセスにかんしてはコネクショニスト認識論も ISA アプローチも抽象的なモデルを提出することしかできないからだ。しかし、知覚プロセスにかんしていえば、現実性の高いモデルと経験的予測を与えることができるコネクショニスト認識論が ISA アプローチよりも有効であるといえそうだ。

したがって、コネクショニスト認識論の評価は、特定の領域については ISA アプローチよりも優れているが、他の領域では ISA アプローチとの優劣はそれほど明確でないか、ISA アプローチのほうが優れているというものになる。このことは、認識論における方法論的多元主義の可能性を示唆する。すなわち、コネクショニスト認識論と ISA アプローチを、それぞれの適性にあわせて使い分けるということだ。この選択により認識論における統合理論の構築という魅力的な目標からは遠ざかることになるが、それでも方法論的多元主義は自然化された認識論にとって実り多いものとなるだろう。なぜならば、多元主義をとることでコネクショニズムと古典的計算主義双方の研究から認識論が恩恵をうけることができるからだ。したがって、二つのアプローチの優劣が決定的でない現時点では方法論的多元主義をとることが最良であるだろう。

(4) 今後の展望

方法論的多元主義の具体的実践のためには、ISA アプローチとコネクショニスト認識論それぞれの長所・短所をより丁寧に検討する必要があるだろう。とくに、高次の心理プロセスの記述・評価においてコネクショニスト認識論と ISA アプローチのそれぞれの説明の妥当性を評価する必要がある。また、「拡張された認知」の議論をとり入れることで、コネクショニスト認識論の枠組みと外的表象の使用とを調和させられないか検討をしたい。さらに、方法論的多元主義をとりつつもコネクショニスト認識論と ISA アプローチの統合が可能であれば、その道も探りたい。そのためには、古典的計算主義者とコネクショニストの間の議論と、双方の統合の取り組みについて検討する必要がある。

1. 広義のコネクショニスト・モデルには、表象を担う各ニューロンがそれぞれ一つの命題に対応するモデルも含まれており、このようなローカルなコネクショニスト・モデルの検討はゴールドマンやサガードも行っている (Thagard, 1989)。命題に対応する構造が役割を果たすローカルなコネクショニスト・モデルは、ISA アプローチの一種と見なすべきであろう。

本稿ではコネクショニスト認識論と ISA アプローチの区別を明確にするために、複数の命題や概念が各ニューロンによって分散的に表象されるモデルをコネクショニスト・モデルと呼ぶ。

2. 戸田山はチャーチランドの認識論の課題として同様の指摘を行っている（戸田山，1999）。チャーチランドは，科学者間の情報共有や教育，記録のための言語の重要性を示している（Churchland, 2012）。しかし，コミュニケーションの役割の解明にコネクショニスト認識論の枠組みが役立っているかは疑わしい。
3. 科学理論自体はプロセスではなく一種のメソッドだといえる。しかし，ここで問題にしているのは，理論的概念の把握を可能にする心理プロセスであり，このプロセスの記述は第一認識論の課題となる。

(5) 参考文献

- Churchland, P.M. (1979). *Scientific Realism and the Plasticity of Mind*, Cambridge University Press.
- Churchland, P.M. (1981). "Eliminative Materialism and the Propositional Attitudes", in *The Journal of Philosophy*, 78, pp.67-90.
- Churchland, P.M. (2007a). "What Happens to Reliabilism When It Is Liberated from the Propositional Attitudes?", in *Neurophilosophy at Work*, Cambridge University Press, pp.88-112.
- Churchland, P.M. (2007b). "Chimerical Colors: Some Phenomenological Predictions from Cognitive Neuroscience", in *Neurophilosophy at Work*, Cambridge University Press, pp.161-198.
- Churchland, P.M. (2012). *Plato's Camera: How the Physical Brain Captures a Landscape of Abstract Universals*, MIT Press.
- Goldman, A.I. (1986). *Epistemology and Cognition*, Harvard University Press.
- Quine, W.V.O. (1969). "Epistemology Naturalized", in *Ontological Relativity and Other Essays*, Columbia University Press, pp.69-90.
- Ramsey, W., Stich, S., and Garon, J. (1991). "Connectionism, Eliminativism, and the Future of Folk Psychology", in *The Future of Folk Psychology: Intentionality and Cognitive Science*, John D. Greenwood (ed.), Cambridge University Press, pp.93-119.
- Thagard, P. (1988). *Computational Philosophy of Science*, MIT Press.
- Thagard, P. (1989). "Connectionism and Epistemology: Goldman on

Winner-Take-All Networks”, in *Philosophia*, 19(2-3), pp.189-196.

戸田山和久 (1999). 「科学哲学のラディカルな自然化」, 『科学哲学』, 32(1), pp.29-43.

(東京大学)