新進研究者 Research Notes

日本科学哲学会 科学基礎論学会

第1号(2018年)

公理の有用性と加速定理 ・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・
······ 三笠 雅也 17
毒パズルと脳走査装置 · · · · · · · · · · · · 佐藤 広大 26
内的確率測度とフィルタリングアルゴリズムによる整合測度の構成・・・・・・ 上田 恭平 35
存在論の方法としての Truthmaker 理論はどのような条件を満たすべきか
······ 吉田 佑介 43
心の哲学と神経科学における心脳問題と心的因果の取り扱いに関する論考
······ 櫻井 圭介 52
理由のために行為することと実践的推論 ・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・
認知的侵入可能性と認識的影響 ・・・・・・・・・・・・・・・・・ 原田 夏樹 70
真理の表現的機能と反映原理について ・・・・・・・・・・・・・・・・・ 林 大智 78
セラーズが言う自己についての「報告」とはどのようなものか? ・・・・・・・ 過能 洋平 86
状況づけられたエージェントの推論活動 ― アブダクションと常識推論をめぐって ―
·····································
様相の形而上学における傾向性主義を退ける ・・・・・・・・・・・ 大畑 浩志 101
死の害のタイミング問題 ・・・・・・・・・・・ 佐々木 渉 110
行為の合理化の論理構造 ― 主たる理由の時間的・様相的特性を巡って ―
・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・ 笹本 もも 119
部分構造論理における埋め込み定理に対する相意味論的アプローチ
······ 田中 大海 127
Contemporary Debates on Possible Worlds · · · · · · · · · 遠藤 進平 135
グラウンディング概念の理解における実体の役割について ・・・・・・・ 後藤 真理子 144
ダメット意味理論再考 — 実践能力としての理解 — ・・・・・・・・・・・ 三上 温湯 153
ファフェ心が上間コプ 大阪間月でして、水土所 二上 価物 100

JAPANESE STUDENT RESEARCH NOTES OF PHILOSOPHY OF SCIENCE

Philosophy of Science Society, Japan Japan Association for Philosophy of Science No. 1

Contents

Takuma Imamura: Usefulness of Axioms and Speedup Theorem · · · · · · · · · · · · · · · · · · ·
Karen Yamazaki: On Autonomy of Artificial Intelligence · · · · · · · · · · · · · · · · · · ·
Masaya Mikasa: Ecological Approach in psychiatry: From Thomas Fuchs's studies17
Kodai Sato: The Toxin Puzzle and The Brain Scanner · · · · · · · · · · · · · · · · · · ·
Kyohei Ueda: Construction of Coherence Measure from Internal Probabilistic Measure
and Filtering Algorithm $\cdots 35$
Yusuke Yoshida: What Is Required for Truthmaker Theory to Be a Methodology of
Ontology?43
Keisuke Sakurai: Discussion on mind-brain problem and mental causation in philosophy
of mind and neuroscience $\cdots 52$
Anju Nakane: Acting for Reasons and Practical Reasoning61
Natsuki Harada: The Cognitive Penetrability and Epistemic Influence · · · · · · · · · 70
Daichi Hayashi: On the Expressive Function of Truth and Reflection Principles78
Kano Yohei: What kind of "report" about self argued by Sellers?86
Yuki Mori: Reasoning activities of a situated agent: Abduction and Commonsense
reasoning revisited · · · · · · · 93
Hiroshi Ohata: Dismissing the Dispositionalism in Metaphysics of Modality
Wataru Sasaki: The Timing Puzzle of Death's Badness · · · · · · · · · · · · · · · · · ·
Momo Sasamoto: How to analyze rationalizations to actions: a proposal to some
extension of modal logic · · · · · · · · · · · · · · · · · · ·
Hiromi Tanaka: A Phase Semantic Approach to Embedding Theorems in Substructural
Logic127
Shinpei Endo: Three layers of counterarguments to concrete modal realism
Mariko Goto: On the role of a substance in understanding the concept of grounding144
Onyu Mikami: Dummett's Program of Theory of meaning revisited: On Understanding
as practical ability · · · · · · · · · · · · · · · · · · ·

公理の有用性と加速定理 Usefulness of Axioms and Speedup Theorem

今村 拓万

Abstract

Usefulness is one of the most useful criteria when choosing a new axiom. This criterion is not based on mathematical platonism, so is acceptable for both mathematical formalists and anti-platonists. In this paper, I argue that this criterion is void when interpreting the term "useful" as "shortening the proofs". The Ehrenfeucht–Mycielski speedup theorem plays an essential role in my argument. I briefly discuss the implication of my argument to mathematical platonism.

1 研究テーマ

Gödel [4] は、数学の基礎理論足り得る如何なる公理系も、その内には証明も反証も出来ない文を持つことを示した。 Post はこの定理について次のように述べている (Post [11, p.295]):

The conclusion is unescapable that even for such a fixed, well defined body of mathematical propositions, mathematical thinking is, and must remain, essentially creative.

A を何らかの数学の基礎となる公理系とすれば、Gödel の定理より、A からは証明も反証も出来ない文(独立な文) φ が存在する。我々は A に新たに φ または $\neg \varphi$ を公理として付け加えることで、より強い公理系 A' を得ることができる。 A' も Gödel の定理の前提条件を満たすので、A' からは独立な文 ψ が存在する。我々は再び ψ または $\neg \psi$ の何方を公理として採用するかの選択を迫られる。この手続きはいくらでも繰り返すことが出来、決して完全な公理系に到達することはない。この公理選択という「創造的」営みは数学基礎論のひとつの役割であろう。本研究のテーマは公理選択の合理的基準である。すなわち、公理選択の営みはどのように進められるべきか、そのような指針は設定可能なのか、といった問いを扱う。

2 研究の背景・先行研究

いま A を数学の基礎となる公理系とする。Gödel の定理より,このような体系は必ず証明も反証も出来ない文 φ を持つ。A に φ または $\neg \varphi$ を付け加えることでより強い公理系 A' を作ることができる。この際,どちらの公理を付け加えるのがより良いかを判断する基準を考えたい。もし我々が A に関する数学的実在論にコミットするならば,A は数学的実体に関する真理を記述

する体系として意図されたものであるから、付け加える公理は数学的実在の世界に於いて真であるべきである。これが公理選択の指針となる。次に反実在論的な立場から公理の良し悪しを判断する基準を考えよう。既存の体系と新たな公理が無矛盾であるか、という整合主義的な判断基準は、この状況では何も教えてくれない。 A に φ と $\neg \varphi$ のどちらを付け加えても無矛盾だからである。本紙では、実在論にコミットしない公理選択の基準として、どちらの公理の追加がより有用であるか、というプラグマティックな基準を検討する。

3 筆者の主張

上の議論を完全にする為には「有用」とは何かということを明らかにしなければならない。そのひとつの基準として「既存の定理の証明がより簡単になるか」がある。ZF に於ける AC はこの基準を満たしているように思われる。例えば、ZF の定理であって、超準解析を(したがって隠伏的に AC を)用いることでより簡単に証明できるものが、多数知られている。この他にも AC が有効な場面は少なくない。本紙の目的はこの基準が実際には空虚であることを示すことである。

3.1 Ehrenfeucht—Mycielski の加速定理 計算可能性理論の基本的な概念を用いる (Griffor [5] を参照). いま計算的枚挙可能 (c.e.) な公理系 A に文 φ を追加した公理系 $A+\varphi$ を考える. ただし $A+\neg\varphi$ は決定不可能であると仮定する. 矛盾した公理系は決定可能であるから, $A+\neg\varphi$ は無矛盾である. すなわち φ は A で証明できない. したがって $A+\varphi$ は A よりも真に強力な公理系である. このとき新たに証明可能となる文はどれくらい豊富に存在するであろうか. 次の定理はそのような文の集合が計算論的に非常に複雑であることを述べる.

補題 1. 任意の c.e. 公理系 A と文 φ に対して, $A+\neg\varphi$ が決定不可能ならば,

NewThm $(A, \varphi) = \{ \psi \mid A \not\vdash \psi \land A + \varphi \vdash \psi \}$

は c.e. でない.

証明. 仮定より $A+\neg\varphi$ の定理集合 $\mathrm{Thm}\,(A+\neg\varphi)$ は co-c.e. でない. ところが

$$\begin{split} \varphi \lor \psi \in \operatorname{NewThm}\left(A,\varphi\right) &\iff A \not\vdash \varphi \lor \psi \\ &\iff A + \neg \varphi \not\vdash \psi \\ &\iff \psi \notin \operatorname{Thm}\left(A + \neg \varphi\right). \end{split}$$

したがって、もし NewThm (A,φ) が c.e. ならば、Thm $(A+\neg\varphi)$ は co-c.e. である. これは不合理. ゆえに NewThm (A,φ) は c.e. でない.

公理系 A に於いて既に証明可能な文であっても, $A+\varphi$ ではもっと簡単に証明できる, ということが有り得る. 次の定理はそのような文が必ず存在し、さらにその簡単化の度合いを計算可能関数の範囲で自由に制御できることを示す.

定理 2 (Ehrenfeucht and Mycielski [3]). 任意の計算可能公理系 A と文 φ に対して, $A+\neg\varphi$ が決定不可能ならば, 任意の計算可能関数 $r\colon \mathbb{N}\to\mathbb{N}$ に対して, 文 ψ が存在して, $A\vdash\psi$ かつ $r\|\psi\|_{A+\varphi}\leq \|\psi\|_A$ が成り立つ. ここで $\|\varphi\|_A$ は A に於ける φ の証明図の複雑さ 5 の最小値を表す.

証明. いまこの定理が成立しないと仮定する. すなわち, 任意の文 ψ に対して,

$$A \vdash \psi \implies r \|\psi\|_{A+\omega} > \|\psi\|_A$$

が成り立つ. 図形 Π に対して $|\Pi|$ で Π の複雑さを表す. すると

$$\operatorname{NewThm}\left(A,\varphi\right) = \left\{ \left. \psi \in \operatorname{Thm}\left(A+\varphi\right) \, \right| \, \begin{array}{l} \forall \Pi. \, |\Pi| < r \, \|\psi\|_{A+\varphi} \\ \to \Pi \text{ is not a proof of } \psi \text{ in } A \end{array} \right\}$$

より NewThm
$$(A, \varphi)$$
 は c.e. である. これは前補題に反する.

例として $r(x)=2^x$ の場合を考えてみよう。Ehrenfeucht-Mycielski の定理より,A で証明可能なある文 ψ が存在して, $2^{\|\psi\|_{A+\varphi}} \leq \|\psi\|_A$ が成り立つ.この最後の不等式は $\|\psi\|_{A+\varphi} \leq \log_2 \|\psi\|_A$ と同じことである.すなわち, φ を公理として A に付け加えることによって, ψ の証明の複雑さが対数を取った分だけ減少するわけである.r をより急増加な関数とすれば,それに応じて証明の複雑さは急減少する.一般に公理系 A とその拡大 A' について「A で証明できる文が A' ではもっと簡単に証明できる」という形の定理を加速定理(speedup theorem)という.Ehrenfeucht-Mycielski の定理は加速定理の典型例である.

3.2 有用性条件の空虚性 以上の準備のもとで、最初に述べた「有用性」の条件が空虚であることを示す。A を数学の基礎となる公理系とする。より正確にいえば、A は無矛盾かつ計算可能な公理系であって、R Robinson 算術を翻訳可能であると仮定する。R Rosser [12] はこのとき A が本質的決定不可能となることを示した。 φ を A に於いて証明も反証もできない文とする。 φ は A で証明されないので $A+\neg\varphi$ は無矛盾である。A は本質的決定不可能であるから $A+\neg\varphi$ は決定不可能である。ゆえに A と $A+\varphi$ に対して Ehrenfeucht-Mycielskiの定理が適用できる。すなわち、A に於いて既に証明可能でありながら、 $A+\varphi$ に於いては遥かに簡単に証明できる。というような文が存在する。すなわち φ は有用である。ところがこの議論は A と $A+\neg\varphi$ に対しても通用する。すなわち φ だけではなく $\neg\varphi$ もまた有用である。したがって「既存の定理の証明がより簡単になるか」という有用性の条件は空虚である。

Ehrenfeucht-Mycielski の定理がその存在を教えてくれる文は、数学的には 無内容であるかもしれない. そこで, 有用性の条件を「数学的な内容を持つ定 理の証明がより簡単になるか」といった風に変更すれば、上記の議論から逃 れられる. ただし「数学的な内容を持つ」とは何であるかという別の問題が 生じてくる. Arrow [1] の不可能性定理は、非独裁制などの望ましい諸条件を 満たす社会的選択ルールが存在しないことを示す、社会的選択理論の基本定 理である. これは一見すると数学的内容に乏しい定理のように思われるが, 実 際には、有限集合上には非単項超フィルターが存在しないという、集合論の定 理と本質的に同じことを述べているのである (Kirman and Sondermann [9]). この事実を鑑みれば、Arrowの定理もまた(超フィルターに関する定理と同様 に) 数学的な内容を持つものであることが分かる.6 これは次のことも示唆し ている. すなわち. ある命題が数学的な内容を持つものであるか否かは. それ を評価する主体の知識、さらには、その時代の数学の在り方や他の学問との関 係、といったものに依存する. 一見して数学的な内容を持たないように思われ る定理も、数学やその周辺分野の発展によって、あるいはパラダイムの転換に よって、数学的な内容を持つものであると再評価される可能性を秘めている、 ということである.

再び ZF と AC を例にこの状況を説明しよう. AC によって証明が簡単化される ZF の定理は, Ehrenfeucht-Mycielski の定理に訴えるまでもなく, 具体的に挙げることが出来る. それらは解析学に於ける定理であったり, 位相空間論に於ける定理であったりする. それらは数学的な内容を持つ. 他方で, Ehrenfeucht-Mycielski の定理がその存在を保証する, ¬AC によって証明が簡単化される定理は, 数学的な内容を持っている保証がないのである. したがっ

て、AC は有用であると言えるが、 $\neg AC$ は有用であるとは言えない. 7 しかしながら、このような公理選択基準は永久不変のものではなく、数学の有り様に左右されるものである.

3.3 数学的実在論への影響 前節の議論は数学的実在論に対しても含意を齎す. いま A を数学的実在の世界 W に於ける真理を記述する体系, φ を A に 於いて証明も反証も出来ない文とし, φ の W に於ける真偽を調べたい. そこで, A に対して φ を公理として付け加えた場合と, $\neg \varphi$ を公理として付け加えた場合とを比較し, より良い公理を W に於いて真であると判断する. もちろん, これはアブダクションの形式を取っているから, 可謬的であり, 一般には真理性を担保しない. さて, ここで「良い」という語を前節までの「有用」と同じ意味に解釈するならば, 前節の議論より, A に於いて φ と $\neg \varphi$ はどちらも同様に良い公理となる. したがって φ の W に於ける真偽について何も知ることが出来ないわけである.

4 今後の展望

第一に、本紙の議論を科学的実在論にまで拡張することである. 通常、検証したい自然法則が与えられたとき、それが現象と矛盾せず、かつ現象をより良く説明するならば、その物理法則は(近似的に)真であると見做される(Harmann [6]). この場合には、「より良い説明」を「より簡単な説明」と同一視したとしても、本紙の議論をそのまま適用することは出来ない. 基礎となる科学理論が加速定理の前提条件を満たすとは限らないからである. さらに、仮に加速定理の前提条件が満たされたとしても、定理によって存在が保証せられる加速可能な文が現象の記述になっている保証はない. したがって、本紙の議論を科学的実在論に適用する場合、何らかの修正を必要とする. もしそのような修正が可能ならば、これは Fraassen [14, p.146] による科学的実在論批判の議論を支持する傍証となりうる.

第二に、「有用性」のより精密な概念規定を行い、以って本紙で述べた空虚性の問題を克服することである。これにより、有用性に基づく公理選択の可能性が理論的に保証される。

注

¹ 選択公理の独立証明の詳細については Kunen [10] を参照.

²Howard and Rubin [7] には選択公理が本質的に関与する数学の諸定理が 豊富に記されている. 3 ZF に超準解析の原理を追加した公理系 IST は ZF よりも真に強くなることが知られている。より正確には ZF に BPI と呼ばれる選択公理の弱形を追加した公理系 ZF + BPI の保存的拡大となる (Hrbacek [8]).

 4 ただし、このことは、 $\neg AC$ を仮定した数学がよく研究されていないという、非本質的な理由によるものであるかもしれない。 実際、1960 年代より、AC と矛盾する (したがって $\neg AC$ を導く) 公理である決定性公理 AD が研究され、その有用性が徐々に明らかになってきている。例えば田中 [15] を参照。

⁵ ここで証明図の複雑さの尺度は static complexity measure (Blum [2]) を成すものであれば何でもよい. 最も簡単な例としては証明図のコードがある.

 6 なお、Arrow の定理は Brouwer の不動点定理とも同値であることが知られている (Tanaka [13]) が、これらの間のつながりは、超フィルターに関する定理との間のつながりほどには直接的ではない.

 7 これもまた $\neg AC$ を仮定した数学の研究が不十分であることに起因するものである可能性は否めない. 数学的に内容のある定理の中から $\neg AC$ によって簡単化されるものが見つかることは有り得る.

文献

- [1] Kenneth J. Arrow. Social Choice and Individual Values. John Wiley & Sons, 1951.
- [2] Manuel Blum. On the Size of Machines. *Information and Control*, 11(3):257–265, 1967.
- [3] Andrzej Ehrenfeucht and Jan Mycielski. Abbreviating proofs by adding new axioms. *Bulletin of the American Mathematical Society*, 77(3):366–367, 1971.
- [4] Kurt Gödel. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatshefte für Mathematik und Physik*, 38(1):173–198, 1931.
- [5] Edward R. Griffor, editor. Handbook of Computability Theory. Elsevier, 1999.
- [6] Gilbert H. Harmann. The inference to the best explanation. *The Philosophical Review*, 74(1):88–95, 1965.
- [7] Paul Howard and Jean E. Rubin. Consequences of the Axiom of Choice. American Mathematical Society, 1998.

- [8] Karel Hrbacek. Axiom of Choice in nonstandard set theory. *Journal of Logic & Analysis*, 4(8):1–9, 2012.
- [9] Alan P. Kirman and Dieter Sondermann. Arrow's theorem, many agents, and invisible dictators. *Journal of Economic Theory*, 5(2):267–277, 1972.
- [10] Kenneth Kunen. Set Theory: An Introduction to Independence Proofs. North Holland, 1983.
- [11] Emil L. Post. Recursively enumerable sets of positive integers and their decision problems. *Bulletin of the American Mathematical Society*, 50(5):284–316, 1944.
- [12] Barkley Rosser. Extensions of some theorems of Gödel and Church. The Journal of Symbolic Logic, 1(3):87–91, 1936.
- [13] Yasuhito Tanaka. On the equivalence of the Arrow impossibility theorem and the Brouwer fixed point theorem. *Applied Mathematics and Computation*, 172(2):1303–1314, 2006.
- [14] Bas C. van Fraassen. *Laws and Symmetry*. Oxford University Press, 1989.
- [15] 田中 尚夫. 決定性公理に関する最近までの諸結果について. **数学**, 29(1):53-64, 1977.

(京都大学)

新進研究者 Research Note

人工知能の自律性について

山﨑かれん

Abstract

Can artificial intelligence be autonomous like human agents? In this article, I adopt Frankfurt's and Bratman's views of "personal autonomy" as possessed by human agents and examine whether BDI agent can have it. BDI agent is an artificial agent in a field of artificial intelligence studies. The architecture of BDI agent focuses on mental states of human agents. I point out that BDI agent doesn't hold some features of personal autonomy, but it can be said that it will have "self-governing policy" which Bratman takes as one important element of personal autonomy.

(1) 研究テーマ

自らの行為を自ら決定するという行為者としての人間が持つ自律性を、人 工的なエージェントが持ちうるかどうか検討する。

(2) 研究の背景・先行研究

昨今なにかと話題にのぼる人工知能についての言説の内では、「自律性」が一つの標語となっている。人工知能が自律性を獲得することで人間による制御が効かなくなることが不安だとか¹、人工知能に汎用性を持たせるためには自律性を付与することが効果的だとか言われている(栗原ほか, 2017)。

しかし、人工知能の研究開発の現場や、人工知能開発について言及する研究を見ても、自律性という言葉の使用は多岐にわたる。そうした状況を指摘している研究として、認知科学者であるトム・フレーゼらのものがある。フレーゼらは、人工生命研究² において適切な自律性の形を提案するために、ロボティクスや人工知能研究、人工生命研究を参照して人工的な主体を構築する際の自律性を分析している。フレーゼらは、環境との間での安定した、柔軟な相互作用能力から特徴づけられる「行動の自律性」と、オートポイエーシス研究的な伝統から特徴づけられる「構成的な自律性」という二つの区

別がつけられることを指摘した。行動の自律性は、環境と相互作用しつつ、 初期の設計や行動時の持続的な介入によることなく動作する能力や、さらに その上で目標を作り達成する能力、また行動の柔軟性が高いということなど から特徴づけられる。一方で構成的な自律性は、自己創出性などから特徴づ けられる生命システムに注目したもので、本来的には生物にしか達成しえな い、非常に制約の大きい自律性概念だと指摘されている。

今回は前者のような、主体の動作に主眼を置いた自律性を主題としたい。動作する主体、あるいは行為者を人工的に構築することや、そうして構築した人工的な行為者を用いて研究を行う人工知能研究の領域を「エージェント研究」という。エージェント研究にはいくつかの異なる領域があるが、その中に、世界の中で環境に応答し、判断して行動するエージェントを作る研究がある。エージェントの例としては、現実世界で活動する知能ロボットや、仮想的な世界で行動するゲームのキャラクターが挙げられる。特に自律性を強調したものを自律エージェントとも言うが、こうしたエージェントについて言われる自律性を詳しく見てみよう。

ゲームのプログラマであるマット・バックランドは、「『自律エージェント』という用語を、ある程度に自律的な動作を遂行するようなエージェントに関して用いる。例えば、道中で壁に出くわすとかいうような予期せぬ状況が起きたとき、自律エージェントはそれに対して反応し、適切に動作を生成することができる」(傍点は筆者)と著している。

人工知能に関する哲学的論考を多く行うヴィンセント・ミュラーは、主体の行動の柔軟性に着目し、それを与えるために自律性が求められるとした。ミュラーは、自律的な主体に関して、「主体 X は、X が主体 Y からの入力なしに目的を追い求める程度において、Y から自律的である」と特徴づける。

人間の行動決定過程についてのマイケル・ブラットマンの分析をもとに、主体の心的状態に焦点を当てて自律エージェントを構築しようする研究も行われている。BDI モデルに基づく BDI エージェントである。この研究では自律エージェントとは、人間と同じように目的を持ち、それを達成する方法を考えて、そのために行動するようなエージェントだとされている。BDI モデルは、心的状態のモデルとその時間的変化を用いて、複数の目的を並行して扱うことを可能にすることで、人間の合理的な行動決定を模倣し、自律エージェントを実現しようとする。

心的状態の中でも情動が行動を動機づけることに注目し、ロボットに情動システムを与えることで自律的な行動生成を目指す研究もある(錦田, 2016など)。こうした研究では、ロボットの情動を作り出すことで、ロボットが

自ら考え、自らの意志によって行動するようになり、それこそが自律性の実 現だと考えられている。

このようにして、人工的な行為者の自律性を見ていくと、自律性についてさらに二つの観点があることが見て取られる。一つめの観点は、その主体の行動が環境に対して柔軟であり、外部からの制御を受けずにふるまうという点から自律性を言うものである。たとえば、バックランドは自律的に動作すれば、自律エージェントだとした。ミュラーも、外部からの入力なしに目的を追うというところに自律性を定めた。二つめの観点は、その主体のふるまいが、まさにその主体自身の欲求や信念、意図、目的選択といった心的状態によって生み出されているという点から自律性を言うものである。BDIエージェントは人間の心的状態に着目していた。また、情動モデルを与えることで自律的な人工物を作るという構想も、情動が行動生成のための主体の心思考や意思を生み出すことに期待している点で、情動や思考、意思といったとで自律を主体のよるまいから捉えるか、それとも主体の心的状態から捉えるかという観点の違いである。以降では、上で挙げたうちの二つめの観点を問題にする。

哲学には、行為者たる人間の重要な特性として自律性を分析した「人格的自律性 (Personal Autonomy)」という考えがある。粗く言うと、人格的自律性では「自らの行為を自らの欲求や信念などによって決定できる」ときにその行為者は自律しているとされる。つまり人格的自律性によると、行為者の心的状態のあり方がその自律性を実現するのだ。

先に挙げた二つめの観点をとる上で、人格的自律性は人間という行為者の自律性についての適切な分析だと考えられよう。人工知能の開発目標には、「人間のような」知的ふるまいをする人工物を作る、というものがある³。そのために、人間という行為者の一つの特徴として人格的自律性があるとすると、それを人工知能あるいは人工的な行為者に持たせるということが考えられてもよいだろう。先に挙げた自律エージェントの自律性についての二つめの観点は、エージェントの自律性を構築するための手段として、人格的自律性が要求するような心的状態をエージェントに持たせなければならないと考える、エージェント設計上の立場であるとも言える。では、この立場で設計されたエージェントは、人格的自律性として分析されている自律性を本当に持つことができるのだろうか。ここからは、人格的自律性を人工的な行為者が持ちうるかどうかを問うていく⁴。

(3) 筆者の主張

まずは、問題にする人格的自律性がどのようなものかを明らかにしておきたい。人格的自律性の議論は 1970 年代のハリー・フランクファートやジェラルド・ドウォーキンの論考に端を発する。ここでは、その後の議論に広範な影響を与えるフランクファートの議論を取り上げ、そこから人格的自律性の骨子を理解する。そして、フランクファートの不足を補うものとしてブラットマンの説を導入し、今回検討する人格的自律性の概念を明確にしたい。

フランクファートによると、人間はなにかをする/しないといったことを欲するような「一階の欲求」に対して、その欲求を持つことがよいのかどうかを反省的に評価する態度を持つような行為者である。この反省的な自己評価を行う心的状態は「二階の欲求」とも呼ばれる。そして、その二階の欲求のうち、ある一階の欲求が自らの意志になる(=実際に行為を引き起こす)ことを望む二階の欲求のことを「二階の意欲」と呼び、フランクファートは二階の意欲を持つものを人格として認めた。また、二階の意欲が成立している状態を、その人が一階の欲求に同化しているとも言う。この立場は高階の心的状態による自己統制の仕組みを明らかにすることで自律性の成立を説明し、階層的アプローチと呼ばれて広く影響を与えている。

しかし一方で、単に欲求の階層的な構造から自律性を説明するのでは、高階の欲求も単なる欲求にしかすぎないので、それだけでは行為者が一階の欲求を支持することを保証する権威を持ちえないだろう、という問題が起こる。フランクファートはそうした問題を避けるために、ただ二階の意欲が一階の欲求を欲するというだけではなくて、その人が二階の意欲を自らのものとして受け入れることも必要だと付け足した。これを満足と呼び、行為者が一階の欲求を欲する二階の意欲について反省を行ったあともなお、その二階の意欲を変更するという気が起こらないときに、その二階の意欲に満足していると言われる。そしてそのように二階の意欲に満足しているときに、行為者は一階の欲求に同化していると考えるのである。

しかし、ブラットマンはこのようなフランクファートの提案では十分ではないと考え、その論考をさらに発展させる形で分析を行う。ブラットマンによると、自律的な行為者の特徴として「計画すること」があり、計画によって複数の時点にまたがる行為の組織的な統一が可能になる。人間は時間的な広がりの中で活動する社会的な行為者だが、計画を持つことで行為の直前にその手段について考える必要がなくなり、また個人内部での活動の調整と他人との間の活動の調整が可能になるのだ。そしてブラットマンは、計画を構成する一つの要素として、「意図」という心的状態に、ある重要な独自の役割

を与えている。このような意図は比較的特定された状況においてある種の仕方で行為するという、行為に対してのコミットメントである。意図の中には、あるタイプの状況が生じるたびに一定の行為を行うこと(車に乗るときはシートベルトを締める、とか)にコミットする、より一般的な意図がある。これは「方針」と呼ばれる。

このような方針は、行為に関するものから一階の欲求に関するものへと拡張され、一階の欲求に関するものは特に「自己統制的方針」と呼ばれる。ブラットマンは自己統制的方針の例として、チョコレートの誘惑に負けないようにしよう、とか、陽気であろうとすることに努めよう、といったことを挙げている。自己統制的方針は、実践的推論においてある欲求を行為の理由として扱うことを可能にするものであり、このことが欲求に対する行為者の支持を意味するのである。そして、ブラットマンはここに「満足」についての彼なりの考察を組み込む。ある人が持つ自己統制的方針に、異議申し立てをする別の自己統制的方針がないということが、その人が自己統制的方針に対して満足しているという状態である。行為者が自分の満足する自己統制的方針で欲求を肯定的に評価し、その欲求を実践的推論における行為の理由として扱うことで、その自己統制的方針がその欲求に対する行為者の支持を保証する権威になると考えるのだ。

ブラットマン自身もほのめかしているが、他にも自律的な行為者の特徴づけとして適切な性質はあるだろう。しかし、少なくともここまで見てきた考察は自律的な行為者が持つ重要な特徴を捉えていると考えられよう。以下では、今まで見てきたフランクファートとブラットマンの考察から、人格的自律性を考える。つまり、高階の態度が一階の欲求を反省するという欲求の階層的な構造の上で、長期的な視野を持った自己統制的方針が一階の欲求を肯定的に評価し、それを実践的推論における行為の理由とすることで自己統制的な行為をすることのできる主体を、人格的自律性を有する主体だと言おう。

人工的な行為者はこのような人格的自律性を持つことができるのだろうか。しかし人工的な行為者一般について問うことは難しいため、ここでは具体的な例として先にも挙げた BDI エージェントを分析の対象とする。BDI エージェントは、ブラットマンの意図と計画についての論考を参考にし、人間の心的状態に焦点を当ててエージェントの行為をモデル化した BDI モデルに基づくものであり、その意味でもここでの分析対象として興味深い。

ブラットマンは、人間が合理的に行為するときには、信念(Belief)、欲求(Desire)、意図(Intention)という三つの心的状態が深く関与していると考えた。BDIモデルはその分析を下敷きにして自律エージェントを構築しよ

うとする、工学的な設計指針である。BDI エージェントは工学的なモデル化を経て実現された三つの心的状態とその時間的変化を反映させて意思決定を行う。特にBDI エージェントは自ら欲求を持たないため、BDI モデルにおける欲求は外部から与えられる目標となっている。

BDI エージェントのとる動作は、次のようにまとめられる5。

- ① 与えられた目標(欲求)を読み込む。
- ② 実践的推論によって目標を達成する手段を定めて、その手段を実行するという計画を、意図として形成し、それを保持する。
- ③ 環境についての信念を得て、実行の前提となる条件が満たされれば、そ の意図が実行しようとしている手段が遂行される。
- ④ 意図はそれらの手段を順次実行に移させる。またその手段が、BDI エージェントが直接遂行できない複雑な動作である場合、その手段を副目標として設定し、副目標を達成するための副計画を立てる。
- ⑤ 一度形成された意図はある程度の持続性をもち、これによって、目標達成に向けた動作が一貫したものとなる。

今のところ BDI エージェントが完全な形での人格的自律性を持ちえないことは明らかであるように思われるが、何が不足しているのかを具体的に挙げてみたい。まず、BDI エージェントには、欲求を自ら生み出す機構が欠けていることは明らかであろう。BDI エージェントにとって欲求とは、設計者や使用者から与えられた目標であり、BDI エージェントはそれを反省、評価せずに、その目標を達成するための行為を実行しようとする。それゆえ、BDIエージェントは欲求を反省、評価するような機構も持っていない。

ただし、BDI エージェントは与えられた目標を達成するための手段を講じて、その手段を実行する意図を形成することで、人間のような自律的な行為者の一側面を捉えるように作られている。こうした意味では、「計画すること」という自律的な行為者の一側面を掬うものになっている。

目標を実行しようという意図が、BDIエージェントが直接実行できる動作よりも複雑な手段を必要とするとき、BDIエージェントはその手段の実行自体を新たな目標(副目標)とする。そしてBDIエージェントはその副目標を達成するために、さらに手段を講じてそれを実行するという意図を形成する。

BDI エージェントがはじめに持つ目標(欲求)は、設計者や使用者から与えられる。一方、与えられた目標を達成するための副目標は、自ら立てた目標(欲求)である。副目標を立てるとき、そこではその副目標が与えられた

目標を達成するための適切な目標たりうるか、という評価や反省がなされるはずである。つまり副目標に対して、それが自分に与えられた目標を達成しようという意図に沿うかものなのかどうかという評価、反省がなされるのだ。ここにおいて、自己統制的方針を BDI エージェントが持つことになると考えてもよいのではないか。長期にわたる複雑な計画を通してはじめて達成されるような目標を BDI エージェントに与えたとき、その目標を首尾よく遂行するために一般的な意図、つまり方針を立てなければならないだろう。この方針の中には自己統制的方針も含まれよう。なぜなら、自分の立てた副目標(欲求)が、与えられた目標を実現するために適切であるかどうかを評価、反省し、かつ、その副目標を実践的推論における行為の理由とすることを可能にするような方針(すなわち自己統制的方針)も必要になると考えられるからである。このことから、BDI エージェントは今回定めた人格的自律性が必要とする行為者の性質を、部分的にではあるが備えていると言えよう。

(4) 今後の展望

BDIエージェントは人間のような形で自ら欲求を持つことがないという点において、今回定めた人格的自律性を持つことにはなっていない。この問題に対処するための研究として二つの方向を考えている。一つは人工物に自ら欲求を生み出す機構を与えるという方向、もう一つは自ら欲求を持たない人工物に自律性を認めるという今回とは異なる形の自律性概念の構築を行うという方向である。特に後者については、道徳的責任を帰せられる人工的主体の研究として倫理学の方面からも検討が試みられているようなので、そうした知見を取り入れつつ考察を深めたいと考えている。

- 1長倉(2017)などで、こうした不安を煽る言説が批判的に指摘されている。
- 2 人工生命研究は、ロボットやソフトウェア上のモデルによって、生命システムを人工的に模倣するものである。人工知能と人工生命という双方の研究領域は、どこまで共通しているかを指摘することは難しいが、人工的な行為者の自律性を考える上ではフロースらの分析が役立つと考えられる。
- 3 中島 (2013) など、このシリーズでは人間の知能の解明が人工知能研究の目的の一つだと繰り返し述べられている。そのために、「人間のような」知的ふるまいをする人工物の構築が目指されていると考えられるだろう。
- 4 この種の考察として、たとえば Schmidt et al. (2006) は以下の本文で述べるようなフランクファート的な欲求の階層構造を人工知能が持つかどうかを問うている。

5 新出 (2010) のまとめを参考にした。

(5) 主要参考文献

Bratman, Michael (1987). *Intention, Plans, and Practical Reason*, Harvard University Press. (門脇俊介&高橋久一郎 (訳) (1994). 『意図と行為 合理性、計画、実践的推論』,産業図書.)

Bratman, Michael (2000). "Reflection, Planning, and Temporally Extended Agency," *The Philosophical Review*, 109(1), 35-61.

Buckland, Mat (2004). Programming Game AI by Example, Jones & Bartlett Learning.

Frankfurt, Harry (1971). "Freedom of the Will and the Concept of Person," *The Journal of Philosophy*, 68(1), 5-20.

Froese, Tom et al. (2007). "Autonomy: a review and a reappraisal," In: e Costa, F. Almeida et al. (eds.) *Proceedings of the 9th European Conference on Artificial Life*, Springer-Verlag, 1-13.

Müller, Vincent (2012). "Autonomous cognitive systems in real-world environments: Less control, more flexibility and better interaction," Cognitive Computation, 4(3), 212-15.

Rao, Anand et al. (1995). "BDI Agents: From Theory to Practice," Proceedings of the First International Conference on Multiagent Systems, 312-319.

Schmidt, Colin et al. (2006). "Robots, Dennett and the autonomous: a terminological investigation," Minds and Machines, 16(1), 73-80.

Taylor, James (eds.) (2005). PERSONAL AUTONOMY New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy, Cambridge University Press.

栗原聡ほか (2017). 「汎用 AI 実現のための鍵となる自律性とマルチモーダル性についての考察」, 第 31 回人工知能学会大会発表原稿.

新出尚之 (2010). 「自律エージェントの論理モデル」, 『人工知能学会誌』, 25(3), 419-428.

長倉克枝(2017). 「『人工知能脅威論』が覆い隠す、本当の問題は何か?―― 日仏の研究者が議論」, ハフィントンポスト日本版, http://www.huffington post.jp/katsue-nagakura/ai-ethics_b_16904442.html, (2017/11/26 参照).

中島秀之(2013). 「レクチャーシリーズ『人工知能とは』第一回 人工知能とは(1)」, 『人工知能学会誌』, 28(1), 139-143.

新進研究者 Research Note

錦田将悟 (2016). 「自律ロボットのための情動に基づく意思決定システムに関する研究」,山口大学大学院理工学研究科博士論文.

人工知能学会(編)(2017).『人工知能学大事典』, 共立出版.

(東京大学)

精神医学における生態学的アプローチ -Thomas Fuchs の論考から-

三笠 雅也

Abstract

It is a well-known fact that each Japanese psychiatrist is trying to approach to his/her patients in various styles. What theoretical approach is the most appropriate to a person with mental illness?

The brain has been widely studied in biological psychiatry, particularly in psychopharmacology and neuroimaging. They think that mental illness can be reduced to a disease of the brain. This thinking is objected to by Thomas Fuchs' ecological concept of the brain. From the standpoint of a clinical psychiatrist, ecological approach would give full weight to a person with mental disorder. By the adoption of this approach, a patient could reinforce each narrative identity from the dialogue with his/her psychiatrist.

(1) 研究テーマ

本邦の臨床では、精神科医によって患者へのアプローチが著しく異なっているという現況がある。それでは、精神疾患をもつ患者に対して、どのような理論的アプローチが適当なのだろうか。

また、このアプローチ選択は、精神疾患をどのように捉えるかという問題にも繋がる。果たして精神疾患は単なる脳の障害と考えるべきなのだろうか、それとも脳を超える障害と捉えるべきなのだろうか。

(2) 研究の背景・先行研究

18世紀後半から 19 世紀にかけて精神医学が生まれてから、生物学的アプローチュと心理学的アプローチが互いに興隆と衰退を繰り返してきた。20 世紀中頃には精神分析を代表とした心理学的アプローチが主流の時代もあった。しかしその後、1950 年代から再び生物学的アプローチが主流になった。それは精神薬理学の興隆が大きい。例えば統合失調症に対するクロルプロマジン、うつ病に対するイミプラミンなどの有効性が実証されたからである。その後も D_2 仮説やモノアミン仮説などによって、より副作用の少ない薬物が数多く発売されてきた。また神経画像に関しては、黒木(2016)は「近年の脳画像研

究の進展は目覚しく、従来、想像もつかなかったような斬新な脳の理論を提供しつつあり、精神疾患の分類にも影響を与えつつある」(154項)として評価している。しかし、一方で黒木(2016)は「近年、真に新たな向精神薬の候補薬の開発は途絶えており、従来のモノアミン仮説を凌駕する確かな開発の里程標となるべき理論も登場していない。欧米の主要な薬理学専門誌が、しきりに精神薬理学の危機や衰退を警告しているのが現状である。……現在もまたfMRI(機能的磁気共鳴画像法)や MEG(脳磁図)など脳科学の巨大技術の精神医学への応用が精力的に進められている。しかしながら、他の医学領域と異なり、精神医学の場合、それらが新しい世界を見せてくれている興奮の期間はあまりにも短く、幻滅へと転じやすい」(154-164項)と述べているように、最近の生物学的精神医学には閉塞感が漂っていて、精神薬理学が停滞していること、神経画像がなかなか臨床に応用できていないことを示している。これは臨床現場での精神科医が実感しうるものであり、一時期の新規薬物の発売ラッシュが落ち着いてきていること、脳画像検査が補助診断に留まっていることからも推測されるだろう。

現代では患者の対人関係の問題も注目されるようになり、この社会的アプ ローチも含めた、生物・心理・社会モデルの折衷主義的アプローチに落ち着 いたかのようになっている。それに対し、アメリカの精神科医であるナシア・ ガミー Nassir Ghaemi は多元主義 2的アプローチを訴えている。ただし村井 (2016)は、これは本邦では馴染みのアプローチであり、ドイツの精神病理学 の影響を受けた本邦の精神医学は、精神疾患を「心因」「内因」「外因」に区別す る三分法で診断を行い、それによって異なる治療を選択するという過程、こ れこそ多元主義そのものである(210 項)と述べている。すなわちエビデンス に基づく医療 ³Evidence-based medicine(以下 EBM)全盛の米国精神医学で は目新しくても、本邦の臨床には昔ながらの多元主義が根付いているのであ る。しかし、これが余計に本邦の現代精神医学のなかでの乖離を顕在化させ ている。患者に対して上の世代の精神科医は昔ながらの多元主義的アプロー チを採る。一方で、下の世代の精神科医は折衷主義的アプローチを採る。す なわち若い精神科医は操作的診断基準4に従って診断を行い、治療は生物学 的アプローチ(薬物療法)や EBM に基づいた認知行動療法を採用する。現代の 精神医学では、他の医学領域(身体医学)のように EBM に従って、ある意味で 均一的かつ画一化された医療を提供する傾向が強い。例えば、精神疾患にお いてもガイドラインの方針に沿って、治療を行う精神科医も珍しくなくなっ た。また、ベンゾジアゼピン系の抗不安薬の長期投与を実際に控える風潮も 出てきている。EBMに従うと、必然的に生物学的精神医学に偏重していくこ

とになるであろう。しかし臨床での精神医学の特徴として、同じ病名であっても患者ごとに病状が異なるため、その時点での、すなわち横断面での症状だけでなく、時間の経過で病態の質の変化を診ていく縦断面が必要不可欠である。そのため EBM は臨床場面では不適当であるというのが上の世代の精神科医の見解である。その EBM の反動として、物語に基づく医療 $5Narrative-based\ medicine(以下 <math>NBM$)の重要性が唱えられたりしてきた。

(3) 筆者の主張

この現代の精神医学における乖離の原点として、1845年にグリージンガーが唱えた有名な命題である「精神病は脳の病気である Die Geisteskrankheiten sind die Gehirnkrankheiten」は重要であろう。近年神経科学の発展とともに、この命題は現代の精神科医に強く影響を与えている。特に下の世代の精神科医ほど、生物学的アプローチを患者に対して採っている場合が多いであろう。もちろん、この潮流は理解できるものである。脳の機能不全により精神疾患が発症し、病的体験などの症状として現出する。問診によって患者の症状を並べて、横断的に診断し病名をつける。そして患者に対して D_2 仮説やモノアミン仮説などを説明し、病気の原因は脳内の神経伝達物質の過剰や不足によるものであるとして薬物療法を行う。これは現在の臨床現場で最も行われているアプローチであろう。

しかし本稿では、精神疾患が脳の障害に還元しうるとする生物学的アプローチを否定的に論じる。その前段階として、まず主観(主体)は脳活動に還元されるという還元主義を否定する。その論拠としてドイツの精神病理学者トーマス・フックス Thomas Fuchs の「脳の生態学的考え方 ökologische Konzeption des Gehirns」を採用する。この考えは「脳は人と環境の相互作用に根差しており、脳はせいぜい生物学的、精神的、社会諸過程が相互に反応しあうための媒介および変換の器官とみなすべきである」(フックス,2016,265項)というものである。

フックス(2016)は「人の脳を調べてみても、そこには人の知覚も、志向も、不安も、痛みも見出せない……なぜなら脳も、個々の脳中枢も、体験の主体ではないのだから。……それは生き物であり、生きている有機体であり、ひと自身である」(266 項)という記述にあるように、現象学における主体の体験を重視しており、これは決して三人称パースペクティヴ(客観性)には還元されえないとしている。一人称パースペクティヴ(主観性もしくは自己意識)が生じるのは、生きている存在そのものに生じており、それは環境との相互作用により成り立っていると考えるのである。すなわち、精神は脳が生み出す

のではなく、有機体(生体)6全体が生み出すのであり、そしてその成立には環境との相互作用が必要不可欠であると考えている。この考えは現在では珍しいものではなく、認知科学ではエナクティヴ 7enactive な認知として知られている。フックス(2010)はこのエナクティヴ認知神経科学を「身体性や行為を強調し、前反省的な次元から意識を捉え直そうとする」(106項)と述べていて、神経科学を現象学的立場からみたものであると考えている。

フックス(2016)は生態学的考え方として、3 つの次元を提示している (267-269 項)。1 つ目は脳と身体の相互関係、2 つ目は「脳と身体」と環境の相互作用、3 つ目は人と人の相互作用である。以下に簡略に述べる。

1つ目に関しては、「植物的次元で成立している脳と有機体の統一が、より高次の脳機能も包含している……知覚、感情、思考、行為など全ての意識活動は、決して新皮質の神経過程のみに基づいているのではない。すべての意識活動は有機体全体とそのアクチュアルな状態に関係する生命維持および情動の持続的な制御過程に基づいている」(フックス,2016,267項)と述べているように、脳活動には身体の下位システム(生理学的側面)との相互関係が不可欠であり、素朴に考えるような、脳の高次機能からの指令がトップダウンに降りてくるだけではないと主張する。一方で自己意識(精神)は、その状況における有機体(生体)全体の複雑な働きであるとしている。

2つ目に関しては、「脳はまた、環境との感覚運動の相互作用にも根差しており、脳は我々の知覚と運動を媒介している。……脳は第一に、媒介の器官として、知覚と運動を結びつける器官として作用する」(フックス,2016,267-268項)と記述するように、有機体(生体)の知覚と運動は表裏一体であるが、これを媒介しているのが脳だとしている。そして「意識された状態とは常に、環境の中で活動する有機体全体の状態である。主観性(主体性)は、身体に具現化され、環境に根差しており、常に『生態学的な主観性(主体性)』である。……我々は本来、常に既に、我々の身体を超え出ているのであり、我々の自己体験もまた『生態学的』なのである」(フックス,2016,268項)とも述べるように、自己意識(精神)は身体に具現化された単なる有機体(生体)ではなく、環境と相互作用している有機体(生体)全体こそが自己意識(精神)であると考える。すなわち自己意識(精神)は、有機体(生体)と環境のどちらにも根差しているのである。身体を超越して環境側に重心を移すときに、これを強調して「生態学的」と名付けると考えてもよいかもしれない。

3つ目に関しては、「脳は、対人的および生活史的に形成され、影響される器官である……脳が精神を生産するのではない。生きている、身体に具現化された精神が、脳を創り出す……脳は、我々の経験すべてを、永続的な準備

性や素質や能力に転換するための基質として働く」(フックス,2016,268-269項)の記述にみられるように、自己意識(精神)が脳を創り出すという一見すると逆説的な考えを提示しているが、脳は可塑性を持つ器官であり、環境からの影響を強く受けることを考えれば納得のいくものである。また、脳は媒介の器官としてだけでなく、対人関係のなかで沈殿(蓄積)していく経験(体験)を、未来における投企の準備として変形しうる、変換の器官でもあることを示している。

上記のような精神(自己意識)と脳の関係性における、生態学的考え方によ る臨床的妥当性を鑑みると、精神疾患をもつ患者に対しては、生態学的アプ ローチ8が最も適当ではないかと考えている。「精神疾患は脳の疾患を超える ものである……生物学的、心理学的、家族的、社会的影響が互いに作用しあ い、抑制しあったり、強化しあったりするような円環的出来事として捉える」 (フックス,2016,269-270項)と記述するように、それぞれの関係性はつねに円 環的(循環的)な相互作用があり、フィードバックの作用を互いに持つと考え るべきである。フックス(2012)は精神疾患の病理として、垂直な円環的因果 性 vertical circular causality と水平な円環的因果性 horizontal circular causality の 2 つに大別し、前者は有機体(生体)の水準、すなわち身体などの 内部の円環に機能不全があるとするのに対し、後者は対人関係の水準、すな わち環境などの外部との円環に機能不全があるとする、一方で脳はそれぞれ において媒介器官もしくは変換器官として働くとする(p334-338)。よって精 神疾患はあくまでも、脳単独の機能不全ではなく、有機体(脳を含んだ身体) と環境(他者を含む)との相互作用の機能不全によって生じるものである。す なわち、精神疾患は脳を超える障害と捉えるべきであろう%。

精神疾患において、患者の世界内存在を考慮に入れることが不可欠であり、 患者は話を聞いてもらうことで、物語的自己同一性 narrative identity を高 めることができる。対話 dialogue に触発されて、患者が自らの物語を形づく ることで、世界での在り方に強く異他性を感じている患者であったとしても、 世界での在り方に対して親和性を抱く機会が増えるであろう。

(4) 今後の展望

臨床的に生物学的アプローチから距離を置くことで、精神疾患は脳の障害から出現するという因果性からは離れることができる。精神医学における生態学的アプローチの最大の効果・効用は、グリージンガーの命題からの離脱であろう。特に下の世代の精神科医による生物学的精神医学への偏重が是正されうる。すなわち本邦の精神科医が患者に対して、今までよりも全体論の

立場を取ることによって、患者自身の世界での在り方を含めた精神疾患という捉え方が生まれるであろう。例えば、「内因」の精神疾患(統合失調症や躁うつ病など)であっても、問診からすぐに診断、治療(薬物療法、認知行動療法)の流れだけではなく、診察場面での対話という地平が開かれるのではないだろうか。

(注)

- 1. 生物学的アプローチとは、本稿では精神疾患を脳の病気だと考えて研究 を進めること、もしくは患者に接することとする。現代では精神医学研 究の主流はこの生物学的なものであり、遺伝子研究や精神薬理学、神経 画像の研究が進んでいる。
- 2. この多元主義 Pluralism とは、「特定の状態や状況に対して、何らかの方法は他の方法に比べてより適正であるという見方。この主義によれば、少なければ少ないほどよい。それぞれの方法は、それぞれ純粋なかたちで使う。複数の方法をひとつひとつ順次組み合わせるのはよいが、同時にいくつもの方法を組み合わせてはいけない。適正な理論や適正な方法はただ1つであるが、精神医学のさまざまな局面においてそれがいつも同一のものだというわけではない。方法の選択は、もしそれが利用可能であれば経験的データに基づいてなされる。私たちは自分の方法選択が概念的または経験的に正当だということを示さなければならない。ただの好みで選択したということであってはならない」(ガミー,2010,251 項)と記述している。
- 3. EBM は「疫学的手法を用いる臨床疫学の研究結果を重視し、非系統的な臨床観察や生理学的研究を、エビデンスのヒエラルキーの下位に位置づけている。……EBM は、エビデンスの検討を経ずに、個人的経験や医学界の権威の意見のみに基づいて臨床判断を行うことを批判する。さらにEBM は生理学的理論のみに基づいて臨床判断を行うことに対しても批判的である」(信原,2016,270-273項)というものである。
- 4. 操作的診断基準とは、「米国精神医学会が編纂した精神疾患の診断と統計マニュアル第 3 版(DSM-Ⅲ)が採用したもので、一定の手順に従えば誰もが精神疾患の診断を行えるようになったことで臨床疫学的研究が加速し、精神医学における EBM 発展の基礎となった」(信原,2016,273 項)というものである。1980 年に出版された DSM-Ⅲは、精神障害の症状やその持続期間などの記述的な特徴を重視し、原因に関する考察を原則として排除するも

のであった、そしてこの背景には精神障害に関する診断が診断者や地域によって大きく異なっていると言う問題意識のため、原因に関する探究を棚上げして診断の一致率(信頼性)が高い診断基準を作成することが目指されたことがあったのである(信原、2016、267・268項)。

- 5. NBM は「EBM の隆盛を追いかける形で、EBM とともに医療の両輪をなすものとして強調されるようになった……患者が主観的に体験する物語を全面的に尊重し、医療者と患者との対話を通じて、新しい物語を共同構成していくことを重視する医療のこと」(信原,2017,275-276項)である。フィンランドで効果を挙げて近年注目されている、オープンダイアローグ(統合失調症患者などの急性期に「開かれた対話」で介入する手法)もこの流れを汲んでいるだろう。
- 6. 有機体(生体)は、本稿ではいわゆる「人間」に相当するものと考えてよい。
- 7. エナクティヴとは、フランシスコ・ヴァレラ Francisco Varela の提唱した概念であり、「認知が所与の心による所与の世界の表象ではなく、むしろ世界の存在体が演じる様々な行為の歴史に基づいて世界と心を行為から産出すること」(ヴァレラ,1991,31 項)と述べている。これは反表象主義の考え方であり、身体に具現化された行為として認知を捉えるものである。この概念は本人も認めているように、メルロ=ポンティ Merleau-Pontyの影響を強く受けている。
- 8. 生態学的アプローチとは、「アメリカの知覚心理学者ギブソンは、知覚と行動への自身のアプローチをこう呼び、その検討の中からアフォーダンスという概念を練り上げた」(信原,2017,220項)と記載がある。アフォーダンスは環境が有する性質のことであり、有機体(生体)ではなく環境側に注目した概念である。環境と有機体(生体)の関係を重視する点で、生態学的といえるであろう。また、「生態学的アプローチでは、……環境の性質を特定する(教える)生態学的情報が生物の外部に存在し、この情報を発見し利用して自己の動作を制御する機構(知覚システム)を仮定する。……生物は、刻々と変化する情報の拾い上げ(知覚)と、拾われた情報を使って身体各部の運動を制御すること(行動)とをコインの裏表のように同時に行っている。知覚と行動は『見るために動き、動くために見る』活動として切り離せない」(信原,2017,222項)と記述のあるように、現在の認知科学でも認められている、知覚と行為の分節できない一体性を示している。そこにはやはり有機体(生体)の身体と環境の相互作用という生態学的考え方がみられる。
- 9. ただし、ここで注意したいのは、生物学的研究が精神疾患の解明において役に立たないということを述べているわけではない。あくまでも臨床に

おいて、患者に対する生物学的アプローチには疑義があり、生態学的アプローチが適当であることを主張している。

京都大学大学院 人間·環境学研究科 共生人間学専攻

(5) 参考文献

ヴァレラ(1991): Varela,F., Thompson,E.and Rosch,E.,1991 "The Embodied Mind: Cognitive Science and Human Experience" *MIT Press*. [田中靖夫訳,2001,『身体化された心―仏教思想からのエナクティヴ・アプローチ』,工作舎]

ガミー(2010): Nassir Ghaemi,S.,2010 "The Rise and Fall of the Biopsychosocial Model - Reconciling Art and Science in Psychiatry" *The Johns Hopkins University Press*. [山岸洋ほか訳, 2012,『現代精神医学のゆくえーバイオサイコソーシャル折衷主義からの脱却』,みすず書房〕

黒木(2016): 黒木俊秀, 2016,「ポストモノアミン時代の精神薬理学―シニシズムを超えて」(石原孝二ほか編, 2016,『精神医学の科学と哲学』,東京大学出版社, 152-171項)

信原(2017):信原幸弘編,2017,『心の哲学―新時代の心の哲学をめぐる哲学の問い』,新曜社

フックス(2010): Fuchs,T.,2010, "Phenomenology and Psychopathology" (In Schmicking,D& Gallagher,S.,eds. "Handbook of Phenomenology and Cognitive Science") Springer,547-573 〔田中彰吾訳,『現象学と精神病理学』〕(石原孝二ほか編,2016,『精神医学の科学と哲学』,東京大学出版社,81-112項)

フックス(2012): Fuchs,T.,2012, "Are mental illness diseases of the brain?" (In Choudhury,S.,Nagel,S.K. & Slaby,J.,eds. " Crinical Neuruscience: Linking Neuroscience and Social through Critical Practice") Wiley-Brackwell,London,331-344

フックス(2016): Fuchs,T., 2016, "Verkörperung und Psychopathologie: Eine ökologische Sicht von Gehirn, Organismus und psychischer Krankheit" Jpn.J.Psychopathol.37:254-263 〔和田信訳, 2016, 『身体 への具現化と精神病理学―生態学の視点から脳、有機体、精神疾患を捉える―』, 臨床精神病理第 37 巻 3 号, 264-273 項〕

村井(2016): 村井俊哉, 2016,「生物・心理・社会モデルの折衷主義を超えて一ガミーの多元主義とヤスパースの方法論的自覚」(石原孝二ほか編, 2016,『精神医学の科学と哲学』, 東京大学出版社, 198-219項)

新進研究者 Research Note 毒パズルと脳走査装置 The Toxin Puzzle and The Brain Scanner

佐藤 広大

Abstract

The toxin puzzle is a famous thought experiment. In my view, this puzzle consists of several questions. In this research, I deal with the one question: is the toxin puzzle possible in principle? In this question, there are two questions: is the 'mind-reading' brain scanner which will correctly detect the presence or absence of the relevant intention possible in principle; if this scanner is not possible in principle, then is the toxin puzzle not worthy of consideration? I answer, to these two questions, that the brain scanner probably is not possible in principle, and that even if so, the toxin puzzle is still worthy of consideration.

(1) 研究テーマ

本研究のテーマは、G・カフカによって提出された「毒パズル(the toxin puzzle)」である(Kavka 1983)。毒パズルは次のような思考実験である。あなたが大富豪にある取引を持ちかけられるとする。すなわち、もし毒を明日の午後に飲むことを今夜の午前 0 時に意図できたら、それだけで大富豪があなたに 100 万ドルを明日の朝に支払うという取引である。意図できたかどうかは脳走査装置で確認されることになっている。その毒を飲むとまる一日ひどく苦しむことになるが、死んだり、後遺症が残ったりはしない。

あなたは 100 万ドルを獲得できるだろうか。 100 万ドルを獲得することなど簡単だと最初は思われるかもしれない。なぜなら、毒を実際に飲まなくても、毒を飲むことを意図するだけで 100 万ドルが獲得できるという取引だからである。しかし、考えてみてほしい。「毒を実際に飲まなくてもいい」と思いながら、毒を飲むことを本気で意図することなどできるのだろうか。 そのようなことは非常に困難に感じられる。「毒を実際に飲まなくてもいい」と思うことと、毒を飲むことを本気で意図することとが両立しないことに注目すれば、「毒を実際に飲まなければならない」と思う戦略が有望に思われてくるかもしれない。だが、「毒を実際に飲まなければならない」と思うことなどで

きるのだろうか。もし毒を飲むことを意図する前に、「毒を 100 万ドル獲得後に実際に飲む理由がない」と思っていたとしたら、「毒を実際に飲まなければならない」と思うことは難しいだろう。すると、毒を飲むことを意図することも難しいということになり、100 万ドルを獲得することも実は難しいということになる。毒パズルがパズルたるゆえんは、簡単だと最初は思われる100 万ドルの獲得が、よく考えてみると実は難しいという点にある。

以上の説明を読んでいて、腑に落ちていないところがあるとすれば、それは、毒パズルではそもそも何が問題になっているかが見通しにくいからかもしれない。筆者の見立てでは、毒パズルは実は複数の問題(たとえば、「100万ドルを獲得できるか」という問題や、「毒を飲むことは合理的か」という問題)から構成されている。毒パズルが複数の問題から構成されているということを自覚しさえすれば、毒パズルは行為論における主要な問題、すなわち「意図と何か」という問題や「合理性とは何か」という問題それぞれについて考える機会を与えてくれるだけでなく、意図や合理性といった基本的な概念同士の関係について考える契機にもなるような豊かなパズルなのである。

本研究は、毒パズルを構成するそれら複数の問題のなかでも、毒パズルが原理的に可能かという問題の一側面を扱う。

(2) 研究の背景・先行研究

本節では、毒を飲む意図の有無を読み取る脳走査装置が原理的に可能かという問いと、そのような脳走査装置が原理的に不可能だとしたら毒パズルは 論じるに値しないものになるのかという問いの二つを確認する。

第一の問いについて、毒パズルという文脈のなかでは、毒を飲む意図の有無を脳走査装置で読み取ることなど不可能だと主張する人々がいる。たとえば、D・ゴティエは「そのような機械は不可能だと思う」と述べている(Gauthier 1998, p.47)。

毒パズルの文脈を離れてみると、そもそも脳の状態と心的状態はどのような関係にあるのかという問題や、脳の状態から心的状態を読み取ることができるのかという問題は、心の哲学のなかではおなじみの問題である。

そのなかから、比較的新しい古田徹也の議論を取り上げてみよう。古田は「心の働きは脳内の物理的過程に付随する」という命題を否定している(古田 2013, p.130)ので、毒を飲む意図の有無を脳走査装置で読み取ることができるというカフカの考えも否定するだろう。

古田は、意図や信念の特徴をいくつか挙げた(*ibid*, pp.65-76) うえで、意図や信念が脳内の物理的過程に付随していながら、そうした特徴を持つと考

えると無理が生じると述べている (*ibid*, pp.76-79)。 たとえば、古田は、長時間持続しうるという意図の特徴を取り上げたうえで、もし意図することが脳の特定の活動だとすると、意図が持続している間、その意図に対応する脳の特定の活動もずっと持続しているということになってしまい無理が生じると主張する。

他方で、近年、脳の信号を読み取って機械を操作することや、機械から脳に刺激を与えることを可能にする機器 Brain-Machine Interface (BMI) が、マインドリーディングの文脈で盛んに研究されている。意図が脳の特定の活動だということを前提にしている BMI は実用化もされていて、意図の有無を読み取るような脳走査装置の実現可能性が高まっているようにも見える。

BMI は、手術をして体の中に埋め込むのか(「侵襲的」)それとも手術せずに体の外から測定するのか(「非侵襲的」)という軸(「体に対する影響という軸」)と、機械から脳に情報が伝達されるのか(「入力型」)それとも逆に脳から機械に情報が伝達されるのか(「出力型」)という軸(「情報の伝達の向きという軸」)の二つの軸で分類される(長谷川 2008, pp.1066-1069; 吉峰他2016, pp.964-966)。

ここでは、毒パズルに関係しそうな実例として、侵襲的な出力型の BMI を利用した、ラットが脳波でロボットアームを動かすという実験を取り上げてみよう(Chapin et al. 1999)。まず、ラットが前足でレバーを押すとロボットアームが動き、水を 1 滴飲めるようにする。ラットの脳の運動野などにあらかじめ電極を刺しておき、ラットがレバーを押すたびにニューロン集団の活動を検出する。次に、ラットがレバーを押す直前に発生するニューロン集団の活動を検出するとロボットアームが動くようにし、レバーを押してもロボットアームは動かないようにする。つまり、ラットがレバーを押さなくても、レバーを押す直前にラットの脳で発生するニューロン集団の活動を検出するだけで動くように設定されたロボットアームという BMI を用意したことになる。すると、ラットはレバーを押さずにロボットアームを動かすように発生するニューロン集団の活動(つまり、レバーを押す「意図」)を発生させることができたように見える。

第二の問い、もし脳走査装置が原理的に不可能だとしたら、毒パズルは論 じるに値しないものになるのかという問いについてはどうだろうか。

脳走査装置が原理的に不可能なら毒パズルは論じるに値しないと諦めてしまうのは早い。原理的に不可能なおそれがあるのは、毒パズル全体ではなく、毒を飲む意図の有無を脳走査装置で読み取るという部分だけである。その部

分さえ取り除いてしまえば、毒パズルを論じる価値が残るかもしれない。

たとえば、M・ブラットマンは、意図の有無を脳走査装置で読み取るとい う部分が取り除かれているのに毒パズルと似た、より日常的な次のようなケ ース(「恩返し事例」)を挙げている(Bratman 1998, pp.63f)。なお、以下 の事例は、分かりやすさのために、オリジナルの事例の細部を変更したもの である。飛行機のなかで、私が頭の上の棚にスーツケースを入れていた。そ こにあなたが来て、あなたもその棚にスーツケースを入れた。飛行機が目的 地に到着して、スーツケースを降ろすことになった。スーツケースを各自で 降ろすよりも、協力して一緒に降ろす方がよいとする。ただし、私とあなた の座席の位置の関係から、まず私があなたのスーツケースを降ろすのに協力 し、次に、あなたが私のスーツケースを降ろすのに協力することになる。私 があなたのスーツケースを降ろすのに協力した時点で、自分のスーツケース を降ろすというあなたの目的は達成されてしまうということに私とあなたは お互いに気づいている。「あなたが私に恩返しするだろう」と、つまり、「あ なたのスーツケースを降ろす際に私は協力してあげたのだから、私のスーツ ケースを降ろす際にもあなたは協力してくれるだろう」と私があなたのスー ツケースを降ろす際に確信しているときにだけ、私はあなたのスーツケース を降ろすのに協力するだろう。

毒パズルと恩返し事例の共通点を強調するブラットマンに対して、ゴティエは相違点を強調する(Gauthier, op. cit, pp.50f)。その相違点とは意図が果たす重要性の違いである。毒パズルのなかで、大富豪にとって重要なのは、あなたが毒を飲むことを意図するかどうかであって、毒を実際に飲むかどうかではない。一方、恩返し事例のなかで、私にとって重要なのは、あなたが私のスーツケースを降ろすのに実際に協力するかどうかであって、協力することを意図するかどうかではない。ゴティエが相違点をこのように強調している理由は、毒パズルと恩返し事例では、あなたが相手(毒パズルで高豪、恩返し事例では私)に利益を与えるのが相手よりも先なのか後なのからである。毒パズルでは、あなたが先に大富豪に利益という違いがあるからである。毒パズルでは、あなたが先に大富豪に利益となり、その見返りとして大富豪は100万ドルをあなたに支払う。一方、恩返し事例ではあなたは後に私に利益を与える。つまり、私があなたのスーツケースを降ろすのに協力する(ibid, p.53)。

(3) 筆者の主張

前節では、脳走査装置は原理的に可能かという問いと、もし脳走査装置が原理的に不可能だとしたら毒パズルは論じるに値しないものになるのかという問いの二つを確認した。一つ目の問いについては、脳走査装置は原理的に不可能だと主張する議論として古田の議論を取り上げた。一方で、脳走査装置の成立可能性を高めているように思われるものとして BMI を取り上げた。二つ目の問いについては、もし脳走査装置が原理的に不可能だったとしても、毒を飲む意図を脳走査装置で読み取るという部分を取り除き、毒パズルを恩返し事例のようなものとして扱えば、毒パズルは依然として論じるに値すると主張する議論としてブラットマンの議論を取り上げた。一方で、意図の果たす重要性という点で毒パズルは恩返し事例とは異なるので、毒パズルを恩返し事例のようなものとして扱うことはできないと主張する議論としてゴティエの議論を取り上げた。

これら二つの問いそれぞれにどのように答えるべきだろうか。

一つ目の問いに対して、筆者は、脳走査装置が原理的に可能な見込みは現時点では高まっていないと答えたいが、その根拠は古田とは異なるものである。脳走査装置が原理的に不可能だと考える古田の根拠は、意図が脳内の物理的過程に付随するにもかかわらず、概念的な特徴(たとえば、長時間持続しうるという特徴)を持っていると考えると、無理が生じるというものだった。しかし、この根拠は、長時間持続しうる意図が脳内の物理的過程(ニューンの発火)にずっと付随していると考えてしまうと無理が生じるということを示しているだけであって、長時間持続しうる意図の形成がその時点の脳内の物理的過程に付随すると考えることを、つまりは、毒を飲む意図の有無を脳走査装置で読み取ることができると考えることを否定するものではない。したがって、古田の議論は、脳の物理的状態から心的状態を読み取る可能性を論駁するものではない。

しかし、古田の議論が上手くいっていなかったとしても、そのような読み取りが不可能だということは依然としてありうるように思われる。なぜなら、筆者の考えでは、現在の BMI は、脳状態から心的状態を読み取る可能性を少しも高めてはいないからである。そのように考える根拠は二つある。第一に、筆者の知る限り、BMI 研究では、行為の直前の意図の読み取りが中心に研究されていて、毒を明日の午後に飲むことを目指す今夜の午前 0 時の意図といった、遠い未来の行為を目指す意図を読み取ることに関する研究がほとんど存在しておらず、遠い未来の行為を目指す意図が脳のどのような状態と関係しているかが明らかではないことである。第二に、行為の直前の意図の読み取りだけを考えてみても、BMI によって意図の内容まで読み取ることができ

ているようには見えないことである。先程挙げた J・K・チェーピンらの実験で、レバーを押すという内容を持ったラットの「意図」を読み取ることができているように見えたのは、実験環境が極めて限定されているからではないだろうか。この実験では、ラットは水を飲むためにレバーを押すように訓練されていて、チェーピンらもレバーを押すという意図を検出することに専念している。このように環境が限定されているからこそ、検出されたニューロンの発火に、レバーを押すという内容の意図を帰属させられるのではないだろうか。さらに、筆者には、毒を明日の午後に飲むという複雑な内容を持った意図を読み取るためにどのように実験環境を設定すればよいのか見当さえつかない。

二つ目の問いに対して、筆者は、ゴティエの議論に同意し、意図の果たす 重要性という点で毒パズルは恩返し事例とは異なるので、毒パズルを恩返し 事例のようなものとして扱うことはできないと答える。意図の有無に主眼を 置いた毒パズルという特殊な事例のなかには、日常的で理解しやすい恩返し 事例と置き換えてしまうと、失われてしまうものがある。

一つ目の問いと二つ目の問いに対する筆者の答えをまとめると次のようなものになる。脳走査装置が原理的に可能な見込みは現時点では高まっておらず、かつ、もし脳走査装置が原理的に不可能な場合、毒パズルを恩返し事例のようなものとして扱うと失われてしまうものがある。毒パズルが原理的に可能かという問いに対しては、毒パズルにおいて脳走査装置が欠かせないならば、毒パズルが原理的に可能な見込みも現時点では高まっていないと答えることになる。

(4) 今後の展望

まず、二つ目の問い(もし脳走査装置が原理的に不可能だとしたら毒パズルは論じるに値しないのかという問い)について今後の展望を述べていくiii。二つ目の問いに対して、筆者は、もし脳走査装置が原理的に不可能だった場合、毒パズルを恩返し事例のようなものとして扱うと失われてしまうものがあると答えたが、だからといって、脳走査装置が原理的に不可能だった場合、毒パズルを救い出すことがまったくできないと主張しているわけではない。恩返し事例のようものとして扱っていては、毒パズルを救い出すことができないと主張しているだけである。それでは、脳走査装置を使わずに、かと言って、恩返し事例に同化させてしまうことなく毒パズルを救い出すためにはどのような方法があるだろうか。たとえば、毒を飲む意図の有無を読み取る脳走査装置が不可能だと考えていたゴティエは、意図の有無を読み取る

脳走査装置の部分を、あなたが本当に意図しているかを鋭く見抜く人に置き換えて考えている(Gauthier, op. cit, p.47)。毒パズルを成り立たせるためには、「見抜く人」が全知の神のようにあなたの意図を完璧に読み取る必要はなく、あなたが「見抜く人」をだますよりも本当に意図した方が楽そうだと(つまり、意図したふりをしてだますという戦略は見込みがないと)考える程度に正確に読み取るだけで十分である。このような「見抜く人」を導入することによって毒パズルを救い出すことができるかもしれない。このとき、「見抜く人」の代わりに全知の神を導入するとどうなるだろうか。哲学の議論では、全知の神のような存在が当然のように不可欠な前提とされていることがあるが、そもそも神のような存在を導入することはいかにして正当化されるのだろうか。こうしたより一般的な観点から、神という存在の導入について検討することになる。あるいは、「見抜く人」よりも現実的な嘘発見器のようなものを導入したほうがよいだろうか。毒パズルを恩返し事例と同化させてしまうことなく救い出す方法を今後このように考えていくことができる。

それでは、本研究全体の今後の展望はどのようなものになるだろうか。本稿の冒頭で、筆者は、毒パズルが実は複数の問題から構成されていると述べた。毒パズルが原理的に成立可能かという問題についての本研究の考察が、毒パズルを構成する他の問題とどのような関係にあるのかということも考えていく必要がある。そのように考えていくことによって、意図や合理性に関する概念的な制約と毒パズルの成立可能性がどのような関係にあるのかということも明らかにしていきたい。そのためにも、毒パズルに対する変更を最小限にとどめ、できるかぎりオリジナルに近い形のまま保っておくことによって、そのパズルとしての豊かさを守っていく必要がある。

i 脳の物理的な状態と心的状態が同一性を満たしていたとしても、その同一性がトークン同一性の場合、脳の状態から心的状態を読み取ることができるとはかぎらない (「非還元的な物理主義」)。また、心的状態がなんらかの物理的状態に付随していたとしても、その物理的状態が脳を含んだ周りの環境全体であることも可能であり、その場合、脳の状態だけから心的状態を読み取ることはできない (「外在主義」)。

ii さらに、古田のように、意図の持続が脳の物理的過程に付随していると考える必要はない。たとえば、意図の持続は脳の物理的な状態(意図の形成によって変化した脳の静的な状態)に付随していると考えることもできるだろう。

iii 一つ目の問い(脳走査装置は原理的に可能かという問い)に対して、筆者は、脳走査装置が原理的に可能な見込みは現時点では高まっていないと答えたが、その一つ目の根拠は、現時点のBMI研究では、遠い未来の行為を目

指す意図を読み取ることに関する研究がほとんど存在していないことだった。しかし、遠い未来の行為を目指す意図を読み取る研究のための足掛かりが全く存在しないというわけではない。遠い未来の行為を目指す意図は「展望的記憶」と深く関わっていると考えられる。「展望的記憶」とは、「意図した行為をタイミングよく自発的に想起すること(梅田 2010, p.2)」である。たとえば、毒を明日の午後に飲むことを今夜の午前 0 時に意図した場合、そのように意図したことを明日の午後に思い出せなければならない。未来の行為を目指す意図と展望的記憶の関係についての研究に注目することによって、未来の行為を目指す意図の読み取りについて研究するための糸口を見つけ出すことができるかもしれない。

(慶應義塾大学)

(5) 参考文献

- Bratman, M, 1998, "Toxin, Temptation, and Stability of Intention.", in Coleman & Morris. eds, 59-83.
- Chapin, J. K. et al, 1999, "Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex.", *Nature Neuroscience* 2, 664-679.
- Coleman, J. L. & Morris, C. W. eds, 1998, Rational Commitment and Social Justice: Essays for Gregory Kavka, Cambridge University Press.
- Gauthier, D, 1998, "Rethinking the Toxin Puzzle.", in Coleman & Morris. eds, 47-58.
- Kavka, G, 1983, "The Toxin Puzzle.", Analysis 43, 33-36.
- 梅田聡, 2010,「し忘れはなぜ起こるのか:認知神経心理学から見た展望的記憶研究」,『認知リハビリテーション』, 1-10.
- 櫻井芳雄, 2013, 『脳と機械をつないでみたら-BMI から見えてきた』, 岩波書店.
- 戸田山和久他編, 2003, 『心の科学と哲学―コネクショニズムの可能性』, 昭和堂.
- 信原幸弘編,2004,『シリーズ心の哲学 I 一人間篇』,勁草書房.
- 長谷川良平,2008,「ブレイン・マシン インタフェースの現状と将来」,『電子情報通信学会誌』91,電子情報通信学会,1066-1075.

新進研究者 Research Note

古田徹也,2013,『それは私がしたことなのか一行為の哲学入門』,新曜社.

吉峰俊樹他, 2016,「ブレイン・マシン・インターフェイス (BMI) が切り開く新しいニューロテクノロジー」,『脳神経外科ジャーナル』 25, 964-972.

内的確率測度とフィルタリングアルゴリズムによる整合測度の構成 Construction of Coherence Measure from Internal Probabilistic Measure and Filtering Algorithm

上田恭平

Abstract

Following Siebel[12], we introduce coherence measures in the mathematical study of coherentism. In this paper, firstly, we construct a new coherence measure through considering Siebel's criticism for coherence measures. Using a simple algorithm, provisionally called filtering algorithm, almost counterexamples by Siebel can be eliminated. Our coherence measure is based on Olsson's measure. Moreover, we will extend the probabilistic measure into the internal hyperfinite probabilistic measure by employing nonstandard analysis and show that hyperfinite measure is more intuitive than standard measures. Finally, we also introduce fuzzy set theory to eliminate Siebel's last counterexample.

1 研究テーマ

本稿では、Siebel の批判に基づいて整合主義者らによる整合測度の構成を概観し、その検討を通して新たな整合測度を構成する。また、採用されるべき整合測度がなんであれ、超準解析の手法がその正当化に貢献することを示し、整合測度の構成に内的超有限確率測度を用いることを提案する。加えて、現在取り組んでいる課題としてファジィ集合論の適用についても触れる。

2 研究の背景・先行研究

「整合性」を確率論の言葉で定式化しようという試みは何度か繰り返されてきた。その中でも Shogenji [11] をはじめとする整合測度(coherence measure)の構成,すなわち入力された信念集合 1 の列 $S=(A_i)_{i\in I}$ に応じてそれらがどれだけ整合的であるかを示す実数を返すような函数 C を考える,という方針は明快である。しかし,Siebel [12] が指摘するように,我々の直観に沿った整合測度の構成は困難である。Siebel は整合測度が直観に反するような値を返すおそれのある 5 つの場合を提示した。

1. 一方の信念が他方の信念の否定を含む場合。例えば、Research Notes の ある投稿者が日本科学哲学会に所属していることを $\varphi(A)$ 、科学基礎論学 会に所属していることを $\varphi(B)$ で表すことにすると、A は B^c を含む 2 多くの整合測度は A と B を両方とも含む信念体系を「不整合」である と判定してしまう。

- 2. 信念体系に含まれる信念の論理積あるいは論理和を信念体系に加えた場合.
- 3. 信念体系から論理的に導き出される信念を信念体系に加えた場合.
- 4. 信念体系に必然的真理 NT を加えた場合.
- 5. 排反であるが近い事象を含む場合. 例えば、ある乾電池の電圧が $V_1:1V$ であること、 $V_2:2V$ であること、 $V_{50}:50V$ であることはそれぞれ互いに両立し得ないが、 V_1 と V_2 は V_1 と V_{50} よりもなんらかの意味で整合的な組である、ということを表現したい. しかし、この問題はいずれの整合測度もクリアできていない 3 .

以下,代表的な整合測度の定義を眺めてみる. \mathcal{B} を事象(信念集合)の集合, $P:\mathcal{B}\to [0,1]$ を確率測度とする.

Definition 1 (Shogenji, 1999)

$$C_S(A_1, A_2, \dots, A_n) \stackrel{\text{def}}{=} \frac{P(\bigcap_{j=1}^n A_j)}{\prod_{i=1}^n P(A_i)}$$

Shogenji 測度は非負値を取る. $C_S(S)$ が 1 未満であれば S は不整合,1 を越えれば S は整合的,と解釈される 4 . $S_1=(A,A)$ と $S_2=(A,A,A)$ という同じ信念体系を表す列の整合性を比較すればわかるように 5 ,この整合測度は好きなだけ増大させることができてしまう.また,Shogenji 測度には Siebel の反例のうち 4 番目以外 6 が当てはまる.第 1 反例について見てみよう. A_2 が A_1^c を含むような (A_1,A_2) を考えると, $C_S(A_1,A_2) \le 1$ となる 7 . $P(A_2) < 1$ のとき, $C_S(A_1,A_2) < 1$ となるから,必ず不整合であると判定されてしまう.

Definition 2 (Olsson, 2002)

$$C_O(A_1, A_2, \dots, A_n) \stackrel{\text{def}}{=} \frac{P(\bigcap_{j=1}^n A_j)}{P(\bigcup_{i=1}^n A_i)}$$

Olsson 測度は 0 以上 1 以下の値を取る. $C_O(S)$ が 0.5 未満であれば S は不整合,0.5 を越えれば S は整合的,というわけである.Olsson 測度には第 1,第 2 以外の反例が当てはまる.4 番目について見てみよう. $P(A_1 \cup A_2) < 1$ なる信念 A_1, A_2 と P(NT) = 1 なる必然的真理 NT を考えると,

$$C_O(A_1, A_2, NT) = \frac{P(A_1 \cap A_2 \cap NT)}{P(A_1 \cup A_2 \cup NT)} = P(A_1 \cap A_2)$$

となるが、これは $C_O(A_1,A_2)=P(A_1\cap A_2)/P(A_1\cup A_2)$ より真に小さい、必然的真理 NT の追加によって信念体系の整合性が小さくなるなどということは不自然である.

Fitelson [3] の整合測度は構成が少しややこしい.

Definition 3 (Fitelson, 2003) ふたつの信念集合を引数に取る Fitelson 函数 $F(A_i,A_j)$ を, A_i が A_j を含むときは 1, A_i^c が A_j を含むときは -1,それ以外のときは

$$F(A_i, A_j) \stackrel{\text{def}}{=} \frac{P(A_j|A_i) - P(A_j|A_i^c)}{P(A_j|A_i) + P(A_j|A_i^c)}$$

で定義する.

 $S = (A_1, A_2, \dots, A_n)$ の組み合わせ集合 Com_S を

$$\bigcup_{2 \leq m_1 + m_2 \leq n} \left\{ \left(\bigcap_{k=1}^{m_1} A_{i_k}, \bigcap_{l=1}^{m_2} A_{j_l} \right) \middle| \{i_1, \cdots, i_{m_1}\} \sqcup \{j_1, \cdots, j_{m_2}\} \subseteq [1:n] \right\}$$

とおく. ただし, [a:b] で a 以上 b 以下の整数全体の集合を表す. 整合測度 C_F は以下で定義される.

$$C_F(\mathcal{S}) \stackrel{\text{def}}{=} \frac{\sum_{O \in \text{Com}_{\mathcal{S}}} F(O)}{|\text{Com}_{\mathcal{S}}|}$$

すなわち、考えられる組み合わせすべてを Fitelson 函数に適用した値の平均を返すようなものを整合測度とする.

Fitelson 測度には Siebel のすべての反例が当てはまる.2 番目について見てみよう. A_1 と A_2 が交わっているとする. $C_F(A_1,A_2,A_1\cap A_2)$ と $C_F(A_1,A_2,A_1\cup A_2)$ はいずれも $C_F(A_1,A_2)$ より真に大きい.しかし,これらは同じ状況を表している.

ふたつの信念集合を引数に取る函数 f を用いて

$$C_f(\mathcal{S}) \stackrel{\text{def}}{=} \frac{\sum_{O \in \text{Com}_{\mathcal{S}}} f(O)}{|\text{Com}_{\mathcal{S}}|}$$

とすれば、これは Fitelson 測度の一般化になっている。f に Fitelson 函数でないものを採用した整合測度の構成は Douven & Meijs [2] などで試みられているが、こちらもやはり Siebel の反例すべてが当てはまってしまう。Siebel は Bovens & Hartman [1] による整合測度も紹介している。しかし、これもうまい構成とは言い難く、第 1 反例しか克服できていない。

Herzberg は [5] で整合主義への超準的アプローチを提示したが、整合測度への適用については触れていない。Robinson [10] によって広く知られることとなった超準解析(nonstandard analysis)は、Leibniz 流微積分学のモデル理論的手法による正当化である。これにより、イプシロン・デルタによる解析学の整備に伴って排除された「無限小」概念(そして「無限大」概念)を用いた数学が可能となる。構成は今村 [6] および竹内 [13] に従う。

推移的な宇宙(universe) \mathbb{U} が ZFC の公理を必要なだけ満たしているとする。内的宇宙(internal universe) \mathbb{U} を考えるにあたり,以下の 4 つの公理を導入する。 \mathcal{L}_X を集合 X の元を定数とした一階述語論理の言語(language)とする。

Axiom 1 (自然延長) $\mathcal{L}_{\mathbb{U}}$ -項 t および $\mathcal{L}_{\mathbb{U}}$ -論理式 φ に対し, $\mathcal{L}_{*\mathbb{U}}$ -項 *t および $\mathcal{L}_{*\mathbb{U}}$ -論理式 $*\varphi$ が対応.これらを自然延長(natural extension)とよぶ.

Axiom 2 (移行原理) 任意の $\mathcal{L}_{\mathbb{U}}$ -閉論理式 ψ に対し、 $\mathbb{U} \models \psi \Leftrightarrow *\mathbb{U} \models *\psi$.

Axiom 3 (共起性原理) \mathbb{U} の部分集合族 $\{A_i\}_{i\in I}$ が有限交叉性を持つ,すなわち任意有限個の A_{i_1},\cdots,A_{i_n} がいつでも $\bigcap_{k=1}^n A_{i_k} \neq \emptyset$ を満たすならば, $\bigcap_{i\in I}^* A_i \neq \emptyset$ が成立.

Axiom 4 (飽和原理) $*U | U |^+$ -飽和である.

移行原理(transfer principle)より標準数学の議論と超準数学の議論を行き来して考えることが許され、共起性原理(concurrence principle)より無限小・無限大といった理想元の存在が保証される. $b \in *\mathbb{U}$ は b = *a なる $a \in \mathbb{U}$ が存在するとき標準的(standard)であるといい、そうでないとき超準的(nonstandard)であるという。 $*\mathbb{U}$ の部分集合で標準集合の元であるようなものは内的(internal)であるといい、そうでないものは外的(external)であるという。以後、我々は $*\mathbb{U}$ 上で数学を展開する。特に $*\mathbb{R}$ を考え、その元を超実数(hyperreal number)と呼ぶ。

一様空間 (uniform space) Xの点x,yに対し,Xの任意の近縁(entourage) U が $(x,y) \in {}^*U$ を満たすとき,x と y は無限に近い(infinitly close)といい, $x \approx y$ と書く。 $x \in X$ のモナド(monad)を $\mu(x) \stackrel{\mathrm{def}}{=} \{y \in {}^*X | x \approx y\}$ で定義する。X がノルム空間で 0_X がその零元のとき, $\mu(0_X)$ の元を無限小(infinitesimal)と呼ぶ。

3 筆者の主張

本稿における筆者の主張はふたつである. ひとつは,あるアルゴリズムに従って与えられた順に信念体系を「整頓」することで Siebel の第 2,第 3,第 4 反例は解消される,というもの. これを仮にフィルタリングアルゴリズム (filtering algorithm) と呼ぶことにする. もうひとつは,超準確率論を用いることで整合測度をより直観にフィットさせることができる,というものだ.

まず,ひとつめの主張について.フィルタリングアルゴリズムは「無駄な」信念を信念体系から抜き取り,あり得ない信念,すなわち P(A)=0 なる A がひとつでも含まれていた場合, $P(\alpha)P(\beta)\neq 0$ かつ $P(\alpha\cap\beta)=0$ なる α と β を信念体系に加える. α と β は条件を満たすような任意の信念であるが,最終的に我々が得たいのは整合測度であるため,ここで α と β の任意性によってアルゴリズムの出力が一意的でなくなることは問題ではない.この操作により,あり得ない信念を含む信念体系は必ず整合測度 0 を返すようになる.次に,信念体系があり得ない信念を含まない場合を考える.我々がいま整頓したい「無駄な」信念は,論理積・論理和・論理的帰結・必然的真理である.

 $S = (A_1, A_2, A_3)$ とする. 信念 A_3 に対するフィルタリングを

$$\lceil P(A_1 \cap A_2) = P(A_1 \cap A_2 \cap A_3)$$
 であれば S から A_3 を除く」

とする. $A_3 = A_1 \cap A_2$ すなわち $\varphi(A_3) = \varphi(A_1 \cap A_2) = \varphi(A_1) \wedge \varphi(A_2)$ のとき, $P(A_1 \cap A_2 \cap A_3) = P(A_1 \cap A_2)$ となり, A_3 は S から除かれる.明らかに,論理和・論理的帰結・必然的真理もこのフィルタリングによって除かれる.一般の長さの S について,信念 A_i に対するフィルタリングを

「
$$P(\bigcap_{k=1}^{i-1} A_k) = P(\bigcap_{k=1}^{i} A_k)$$
 であれば S から A_i を除く」

とすれば、Siebel の反例のうちの3つが解消される。このアルゴリズムでは第1反例を解消することはできないが、フィルタリングされた信念体系をOlsson測度で判定すれば形式的には第5反例以外のすべての批判をかわすことができる。フィルタリングアルゴリズムが出力する信念体系はもとの信念体系が含む信念の順序に依存するが、この性質を修正すべきかどうかは要検討である。

ふたつめの主張に移ろう.標準世界の確率測度に対応する内的確率測度,特に有限確率測度に対応する内的超有限確率測度を考える.標準的な(有限または無限の)確率測度の代わりにこれを使おう,というわけである.

内的超有限確率測度 $P: \mathcal{B} \to *[0,1]$ を用いる利点は少なくともふたつある. (1) 無限和や積分などといった極限操作を経ることなく,無限個(超有限個)の事象に対応する確率が定まる. n は無限大超自然数,すなわち $n \in *\mathbb{N} \setminus \mathbb{N}$

であるとして,n 個の互いに共通部分を持たない事象 $\{A_1,A_2,\cdots,A_n\}$ を考える.これは無限集合である.このとき,少なくともひとつの事象が起こる確率は $\sum_{i=1}^n P(A_i)$ となる.(2) 標準的な測度論であれば,確率が 1 であるときに,その事象が「絶対に成り立つ」ことなのか「ほとんど確実に成り立つ」ことなのかを確率のみからは判別できない.しかし,このことは内的超有限確率測度が 1 であるか,ある非負の無限小超実数 ε を用いて $1-\varepsilon$ と表されるかによって区別することができる.同様に,「絶対に成り立たない」ことと「ほとんど確実に成り立たない」ことは測度が 0 であるか非負の無限小であるかによって区別される.これらの違いをもたらす内的超有限確率測度は,標準的な測度に比べ,より直観にフィットしていると言えよう.

まとめよう. 信念の超有限列全体の集合を \mathcal{B}^{ω} とおく. 内的整合測度 (internal coherence measure) を次のように構成する.

Definition 4 (Internal Coherence Measure) P を内的超有限確率測度 とする. フィルタリング (filtering map) $\mathcal{F}: \mathcal{B}^{\omega} \to \mathcal{B}^{\omega}$ はアルゴリズム

「入力 $S = (A_1, A_2, \dots, A_n)$ に対して:

- 1. $S' = \emptyset$ とする.
- 2. i=2からnに対し、
- 3. $P(\bigcap_{k=1}^{i-1} A_k) > P(\bigcap_{k=1}^{i} A_k)$ かつ $P(A_i) \neq 0$ なら
- A_i を S' に加える.
- 5. $P(A_i) = 0$ なら
- 7. それ以外の場合は何もしない.
- 8. S'を出力する.」

に従うものとする. 整合測度 C_I を次で定義する:

$$C_I(A_1, A_2, \dots, A_n) \stackrel{\text{def}}{=} {}^*C_O \circ \mathcal{F}(A_1, A_2, \dots, A_n).$$

4 今後の展望

Siebel の反例のうち 5 番目は、その性質上整合測度の定義をどのように工夫しても必ず発生する。筆者は現在、この問題は信念を通常のクリスプ集合 (crisp set) ではなくファジィ集合 (fuzzy set) で表現することによって解決されると考え、その定式化に取り組んでいる。通常の集合 (クリスプ集合) はある x がその集合に属すか属さないかがハッキリしており、各クリスプ集合 A は特性函数 (characteristic function)

$$m_A(x) = \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases}$$

で特徴づけられる。ファジィ集合は特性関数の一般化である帰属度函数(membership function)によって特徴づけられる。ファジィ集合 A の帰属度函数 m_A は 0 以上 1 以下の値を取る。例えば,次のようなファジィ集合 V を考えてみる。x をある乾電池の電圧とする。

$$m_V(x) = \begin{cases} 1, & x = 1.5\\ 1/2, & x \in [1.3, 1.5) \cup (1.5, 1.7]\\ 1/4, & x \in (1, 1.3) \cup (1.7, 2)\\ 0, & \text{otherwise} \end{cases}$$

購入した 1.5V の乾電池が電圧測定によって実は 1.6V であると知ったとしても,我々の「1.5V の乾電池を購入した」という信念 $\varphi(V)$ が揺らぐことはない.実際,乾電池の初期電圧は多くの場合 1.6V であり,中には初期電圧が1.7V のものも存在する.それでも「1.5V の乾電池を購入した」という信念は損なわれない 8. 他方,もし 100V であれば確実にこの信念は損なわれる. m_V はこの違いを表すことができている.このように,信念の曖昧さを認めれば Siebel の第 5 反例を解消することは可能であるように思われる.

注

 1 集合を命題に写す全単射 φ によって各信念集合はある信念を表現しているとみなす。和集合は論理和,共通部分は論理積,補集合は否定,部分集合であることは含意に対応する。

² Research Notes に投稿するためにはふたつの学会のうち少なくとも一方に所属していなければならないため, $B^c = (A \cup B) \setminus B \subseteq A$ となっている.

 3 V_2 が V_{50} よりも V_1 に「近い」という感覚が確率の概念とはまったく関係しないことからわかるように、確率測度のみに頼る構成でこの問題を解消することは不可能である.

⁴整合的であることと不整合であることのしきい値は Koscholke [8] に基づく. Siebel [12] は整合測度に必ずしもこのようなしきい値があるわけではないという立場に立つ.

 5 実際,P(A) < 1 であれば $C_S(A,A) = P(A)^{-1} < P(A)^{-2} = C_S(A,A,A)$ となる.適用する列の長さを大きくしていけば, C_S は(まったく同じ信念体系の整合性を検査しているにもかかわらず)際限なく増大していく.この問題は本質的には Siebel の第 2 および第 3 反例の特別な場合に過ぎない.

 $^{6}C_{S}(A_{1},A_{2},NT)=C_{S}(A_{1},A_{2})$ となり、必然的真理 NT の追加は整合性

に影響しない.

 ${}^{7} P(A_1) = 1 - P(A_1^c) = 1 - P(A_1^c \cap A_2) \ge 1 - P(A_1^c \cap A_2)/P(A_2) = 1 - P(A_1^c | A_2) = P(A_1 | A_2) = P(A_1 \cap A_2)/P(A_2)$

 8 もちろん,この事実を知らなければその限りではないだろう.文字の上では同一の信念が実際にどのようなファジィ集合に対応しているかは,あくまでも主観的にしか決定されない.

文献

- [1] L. Bovens and S. Hartmann, Bayesian Epistemology, Oxford University Press, 2003.
- [2] I. Douven and W. Meijs, Measuring Coherence, Synthese, 156(3):405-425, 2007.
- [3] B. Fitelson, A Probabilistic Theory of Coherence, Analysis, **63**(3):194-199, 2003.
- [4] F. S. Herzberg, A Graded Bayesian Coherence Notion, Erkenntnis, **79**(4):843-869, 2013.
- [5] F. S. Herzberg, The Dialectics of Infinitism and Coherentism: Inferential Justification versus Holism and Coherence, Synthese, 191(4):701-723, 2014.
- [6] T. Imamura, Nonstandard Homology Theory for Uniform Spaces, Topology Appl. 209:22-29, 2016.
- [7] 釜江哲朗, 超準的手法にもとづく確率解析入門, 朝倉書店, 1990.
- [8] J. Koscholke, Evaluating Test Cases for Probabilistic Measure of Coherence, Erkenntnis, 81(1):155-181, 2016.
- [9] 西田俊夫, 竹田英二. ファジィ集合とその応用, 森北出版, 1978.
- [10] A. Robinson, Non-standard Analysis, North-Holland, 1966.
- [11] T. Shogenji, Is Coherence Truth Conductive?, Analysis, **59**(4):338-345, 1999.
- [12] M. Siebel, Against Probabilistic Measures of Coherence, Erkenntnis, **63**(3):335-360, 2005.
- [13] 竹内外史, 無限小解析と物理学, 遊星社, 1985.

(京都大学理学研究科)

存在論の方法としての Truthmaker 理論はどのような条件を満たすべきか What Is Required for Truthmaker Theory to Be a Methodology of Ontology?

吉田佑介

Abstract

The aim of this paper is to clarify how it should be for things to make a proposition true from a metaontological point of view. Recently, some philosophers have shown that a theory of truthmaking can play an important role in ontology as a methodology for revealing what there is.

But, the traditional understanding of truthmaking relation is not suitable for general inquiry into ontology in that it unjustifiably skews our intuition concerning truths and reality and thus narrows the scope of the theory. I will try to make clear the problem and propose an appropriate revision of the formulation of truthmaking relation.

(1) 研究テーマ

本稿のテーマは、存在論の方法として見た Truthmaker 理論(以下、「truthmaker」および「truthmaking」は、「TM」と略記する)が、存在論の方法であるために、どのような仕方で理論化されるべきかというところにある。すなわち、本稿の内容は、TM 理論という存在論における一つの問題領域と、メタ存在論というより大きな問題領域の双方に関わる。

(2) 研究の背景・先行研究

まずは、本稿で扱う TM 理論がどのようなものであるかを一般的な仕方で説明しておきたい。TM 理論とは、「ある種の真理は、何らかの仕方で実在の特徴に基礎づけられている」という直観(以下、これを「実在論的直観」と呼ぶ)に基づいて、ここで成り立つ、実在的対象が命題を真にするという関係(以下、これを「TM 関係」と呼ぶ)がどのようなものであるかを探究するものである。具体的には、TM 理論が扱う問題は、大きく分けて、「TM 関係はどのようなときに成り立つか」「どのような存在者が TM の役割を担うか」「どのような真理が TM をもつか」という三つの問いに分けられる。

本稿の内容は、特に、一つ目の問いと二つ目の問いに関わるので、これらの問いに対する応答を、具体例をあげて説明しておこう。多くの哲学者は、上記の一つ目の問いに対する応答として、従来、必然化 necessitation という関係にうったえた様相的特徴づけ(以下、これを「必然化条件」と呼ぶ)を採用してきた。必然化条件とは、次のようなものである。

(NR) i 存在者 x が命題 P の TM であるならば、必然的に (x が存在 するならば P は真である) ii

この条件は、TM が真理の実在的根拠であるためには、その存在が、当該の命題の真理性を保証し、それが偽である可能性を排除するものでなければならないということを明示化したものであるとされる。

二つ目の問いに移ろう。多くの TM 論者は、真理の TM となる存在者として、事態 state of affairs の存在を支持するiii。この背景には、上記の (NR)が、 TM として、命題的存在者とも呼べる存在者を要請するという事実がある。命題的存在者とは、命題を表す文を構成する名前や述語に対応する存在者について、それらが適切な仕方で関連づけられた全体に対応する(以下、このことを「命題にまるごと対応する」と表現する)存在者のことである。

例えば、あるヒイロフキヤガエル(「ヒロシ」と呼ぶ)がおり、「ヒロシは 黄色い」という文が表す命題(以下、「Sという文が表す命題」を単に「Sという命題」と表記する)が真であるとしよう。この命題の TM は何であろうか。普遍者 universal の存在を認める哲学者によれば、「ヒロシは黄色い」という命題の真理性は、ヒロシが黄色さの普遍者を例化している instantiate ということによって説明される。しかし、ヒロシと黄色さの存在だけでは、先の(NR)は満たされない。ヒロシも黄色さも存在しながら、ヒロシが黄色さの普遍者を例化しないことは可能だからである。ここで、(NR) は、TM として、その存在だけで命題の真理を保証する存在者を要求しているのである。こうして、多くの哲学者は、先の真理の TM として、個体や普遍者とは区別された、ヒロシが黄色いという事態という命題的存在者を受け入れる。

以上のような TM 理論は、それ自体、一つの形而上学的問題領域として、様々な形而上学的問題と関連づけられて、広く論じられてきた。一方で、近年では、「真理と存在者の関係を扱う」というこの理論の本性に着目し、TM理論を、種々の形而上学的問題を統一的に扱う、存在論の方法としてとらえなおそうとする哲学者も現れているiv。

彼らの主張は、しばしば、伝統的な存在論の方法である、クワインの存在

論的コミットメント ontological commitment の基準とそれに基づいた言語分析的方法論(以下、「クワイン的方法論」)と対比的な仕方で導入される。クワインは、「何があるのか」という問いを存在論的問いとして定式化し、その答えを与えるものとして、存在論的コミットメントという概念を提示したv。この考え方のもとでは、何が存在するかについての答えは、我々が最終的に採用する理論の量化子の意味によって決定される。したがって、このような問題設定のもとでは、存在論における対立は、言語上の問題をめぐる対立として理解されることになる。これに対し、存在論の方法としてのTM理論を支持する論者は、上記の問題設定を疑い、「何があるのか」という問いの意義や、それに対するクワイン的な言語分析的アプローチを否定する。そして、代わりに、真なる命題のTMを与えることが存在論の本分であると主張する。

(3) 筆者の主張

ここでは、存在論の方法としての TM 理論の具体化の指針を示した上で、(2)で導入された従来の必然化条件 (NR)が抱える問題を指摘し、北村(2014)を手掛かりに、これを解消する修正案を提示する。先に、TM 理論とその理論的動機としての実在論的直観の関係を確認することで、存在論の方法としての TM 理論のもとで、TM 関係がどのように特徴づけられるべきかについての指針を示しておこう。

第一に、先にも述べた通り、実在論的直観とは、TM 理論の理論的動機であり、それゆえ、その明確化が TM 理論の最大の目的となるところのものである。この意味において、TM 関係は、実在論的直観の内実を正確に定式化するものでなければならない。

一方で、存在論の方法として見た TM 理論は、実在論的直観の明確化を踏まえ、真理の実在的根拠を明らかにする営みから、世界に存在する事物について、一定の示唆を得ようとするものである。ここで注意しなければならないのは、真理の実在的根拠である TM が、どのような種類の存在者であるかについて、実在論的直観自体は、何も含意しないということである。この意味において、メタ存在論における一つの立場としての TM 理論は、それ自体では、真理が何らかの実在的根拠をもつという、真理と実在の関係についてのミニマルな主張のみを含むと考えるべきである。

以上の点を踏まえ、(NR)の問題を見ていこう。先に指摘したいのは、(NR)が命題的存在者の存在を要請する一方で、(NR)が依拠する実在論的直観は、必ずしも命題的存在者の存在を要請しないということである。

(2)で見た、事態の存在論による TM 理論的説明を思い出してみよう。そ

こでは、「ヒロシは黄色い」という命題の真理性は、ヒロシが黄色さの普遍者を例化しているというアイディアに始まり、(NR)を満たすために、命題にまるごと対応するような存在者として、事態を措定することで説明されたのであった。一方で、単に実在論的直観に説明を与えるだけであれば、事態の存在を持ち出さずとも、それ以前に提示された素朴な説明で事足りるように思われる。すなわち、「真理が何らかの仕方で実在の特徴に基礎づけられている」ということの説明としては、普遍者と個体の間の例化関係を明示すれば、十分にも思われる。もちろん、実際に、命題的存在者を措定せずに、TM理論の対象となる全ての真理に説明が与えられるかは明らかではない。だが、この点についての結論は、実在論的直観から直接導かれるものではないであろう。だとすれば、従来採用されてきた(NR)は、実在論的直観から、過剰な要求を読み込む、厳しすぎる条件であるように思われるvi。

(NR)のもう一つの問題点を見てみよう。それは、(NR)は、実在が、真なる命題を表す文の言語的構造を反映することを要求することで、形而上学的問題を不用意に招くというものである。

先にも見たように、(NR) は、「ヒロシは黄色い」のような述定的真理を説明するために、事態のような命題的存在者の存在を要請するのであった。しかし、(NR) が要請する命題的存在者は、単純な主語述語文に対応するようなものにとどまらない。例えば、「 $Fa \land Fb$ 」という連言的真理を考えてみよう。(NR) のもとで、これをその存在だけで真にする存在者とはどのような存在者であろうか。もちろん、「Fa」や「Fb」に対応するような事態ではない。一方の存在だけでは、(NR) を満足することはできない。だが、(NR) は、二つの存在者が一つの命題の TM になることも許容しない。ここでは、当該の真理の TM として、「Ta」「Tb」に対応する事態が合わさった何か、すなわち、連言的な存在者が要請されるように見える。しかし、多くの哲学者が、このような存在者を受け入れることについては、懐疑的であり、例えば、「連言的真理の TM はその連言肢の TM によって与えられる」といった新たな原理を持ち出すことで、連言的存在者から逃れようとするvii。

だが、そのような原理の導入が正当化できたとしても、上記と同様の問いは、連言的真理以外のあらゆる真理について立てることができる。例えば、全称的真理や否定的真理、あるいは、様相的真理の TM はどのような存在者だろうか。それぞれの問いについて、(NR) は、問題の真理をそれだけで真にするような命題的存在者を要請することで、実在のうちに、ある種の言語的構造が反映されることをも要求する。ここで、その都度、前段落のような新たな原理が、アドホックではない仕方で導入できるかは定かではない。

もちろん、これらの真理が実在のうちに根拠をもち、TM を要求するものであるかは自明ではない。だが、この点についての制限は、実在論的直観の明確化としての TM 関係が担うべきではない。真理が実在のうちに根拠をもつということは帰結しないからである。したがって、ここでの TM 関係は、どの種類の真理が TMをもつのかについて、さらなる議論に開かれていなければならない。以上の点に鑑みると、(NR) が、真理を表す文の言語的構造をも実在のうちに要請するという事実は、先に述べた真理の実在的根拠についてのミニマルな主張としての TM 理論のあり方と整合しない。

以上の二つの問題点を整理しよう。一つ目の問題点からわかるように、従来採用されてきた TM 関係の特徴づけ (NR) は、実在論的直観の意味するところを超えて、命題的存在者の存在を不可避的に要求する。このことの問題は、それが、TM 関係を、実在論的直観の正確な定式化から遠ざけてしまうだけでなく、説明される真理を表す文の言語的構造に依存して、新たな種類の存在者の存在をも要請することにある。ここで要請される存在者は、それぞれに、固有の形而上学的問題を含んだ、議論の余地のある存在者である。したがって、少なくとも、議論の余地なくそのような存在者の存在を要請してしまう (NR) は、そうではない形式にゆるめられるべきである。

それでは、上記の問題を受けて、(NR) はどのように修正されるべきだろうか。先の二つの問題は、共に、(NR) が、それだけで命題を真にするような、命題的存在者の存在を要請することに起因するのであった。この事実が示唆するのは、次の二つのことである。第一に、ある種の真理の説明責任を全面的に実在の側に押しやるべきではない。すなわち、必然化条件が要求する真理の根拠は、部分的に、説明に必要なだけの存在者を引き合いに出すものである。このためには、その存在だけで命題の真理性を説明することはできないが、一定の条件のもとで、真理の説明に重要な仕方で寄与する TM のあり方を許容する必要がある。第二に、TM は、一個の対象である必要はない。TM がそれだけで命題を真にするものであるという要請は、複数の存在者が合わさって命題を真にする可能性を排除する。このことは、連言的存在者の事例のように、形而上学的に問題含みな対象の存在を導く可能性がある。以上の二点を踏まえた(NR)の修正案の一例は、北村(2014)に見出だすことができる。本稿でも、北村による修正案を手掛かりにしたい。北村は、本稿と部分的に類似した論点から、以下のような修正案を提案している。

(NR') 存在者 xx が命題 P の TM であるならば、必然的に、(xx が存

在し、かつ、xx についてΦ が成り立つならば、P は真である)

ここで、 $\lceil xx \rceil$ は、複数の対象が合わさって文を真にする可能性を許容する複数変項である。なお、複数変項は、TMが一個の対象である可能性を排除しない。また、 $\lceil \Phi \rceil$ にはそれが xxについて成り立つことで、Pの真理性を保証するような条件を表す記述が入る。つまり、(NR')を満たすような TMとは、特定の記述的条件 Φ のもとで命題を真にする存在者であると言える。

ところで、北村 (2014)においては、存在者とともに命題を真にする記述 Φ の具体的な特徴づけが与えられていない。この点が明らかにされなければ、 (NR') が、実在論的直観の明確化になっているとは言い難いであろう。

この点の明確化に際して、直ちに問題になるのは、「xx について ϕ が成り立つ」という記述が、複数の対象についての述定であるというところにある。ここでいう複数の対象が実在的である限りにおいて、この述定の正しさもまた、ある種の実在的根拠を要請するように思われる。だとすれば、(NR) と変わらず、(NR') も真理の説明責任を実在に全面的に押しつけていることになりかねない。この点が示唆するのは、 ϕ は、(NR') が、真理の説明に際して、実在的根拠と非実在的根拠を選別的に引き合いに出すことを可能にするものでなければならないということである。

そこで、Φに入る記述について、「「xx」に入る存在者に加えて新たな存在者を要請しないような記述」という特徴づけをしておきたい。例えば、「その項となる存在者が存在するとき、それらの間に必然的に成り立つ関係」として理解された内的関係についての記述は、その候補となるであろう。

以上のように (NR) を修正することで、本節の前半で示した問題を回避することができる。すなわち、真理の説明において、TM を部分的に引き合いに出す (NR') は、その存在だけで真理を保証するような命題的存在者の存在を必ずしも要請しない。また、それゆえ、(NR') は、真理を表す文の言語的構造を実在のうちに読み込むことも強制しない。以上をもって、(NR') により、実在論的直観のより正確な定式化を得ることができたと言えよう。

(4) 今後の展望

本稿で提示された (NR') とそれに基づく存在論の方法としての TM 理論は、従来の存在論的議論に、新たな土俵を与える可能性がある。(NR') によって得られる実在的根拠についてのミニマルな主張のみを含んだ方法論は、それ自体では (命題的存在者のような) 特定の種類の存在者の存在を前提しないという意味で、存在論的にニュートラルでもある。これにより、様々な

存在論的立場を、統一的な観点から比較することが可能になる。

一方で、具体的な存在論的議論に先立って、(NR')に基づいた TM 理論が、存在論的方法論として、従来の方法論とどのように区別されるのかは、明らかにされなければならない。(2) でも述べたように、存在論の方法論としての TM 理論を支持する哲学者は、これをクワイン的方法論と対比的に導入する。だとすれば、クワイン的方法論との相違点が明らかにされて初めて、TM 理論をメタ存在論における実質的な主張として理解することができよう。

また、(NR') は、(4) の前半で示された (NR) の問題点を解消するために導入された。しかし、注 vi でも述べたように、「実在論的直観は命題的存在者を要請しない」という点は、TM 理論それ自体に反対する論者が、TM 理論への批判の材料とする論点でもある。命題的存在者への要請をはずした TM 理論が、なおも(従来の方法論とは区別された)理論的意義をもつということが示されなければ、上記の批判的論者と TM 論者との対立は、「何を TM と呼ぶか」というターミノロジーにおける対立に過ぎないということになる。

ここで、二つの方法論を区別する方法としては、次の二つが考えられる。 すなわち、一つは、TM 理論とクワイン的方法論では、同じ問題に対して異なる答えを与えるということを示す道であり、もう一つは、TM 理論の扱う問題がクワイン的方法論の扱う問題と異なるということを示す道である。

前者をとる場合、まずは、「何が存在するのか」というクワインの提示した問いについて、TM 理論が十全な答えを与えうるものであるということを示す必要があるだろう。その上で、この問いに対するアプローチとしての TM 理論の優位性を示さなければならない。もう一つの選択肢をとる場合、存在するもののリストを作ることと、真理の実在的根拠を明らかにすることが別の理論的営みであることを示す必要がある。この場合、クワイン的方法論とTM 理論が相容れない理論的営みなのかという点も、重要になるだろう。

いずれの道をとるにせよ、その方法論的独自性、有用性が明らかにされて初めて、TM 理論は、メタ存在論における一立場として正当化される。そして、その正当性が示されて初めて、(2) で見た「どのような存在者が TM の役割を担うか」「どのような真理が TM をもつか」という二つの問いへの答えも、存在論における有意味な結論として評価できるのではないだろうか。

^{*}本稿の内容は、JSPS 科研費 (特別研究員奨励費) (JP17J06386) の助成を受けた研究の成果の一部である。

i 'Necessitation'と'Relation'からイニシャルをとった。

ii この特徴づけからただちに見て取れる通り、多くの哲学者は、必然化条

件を TM 関係の成立の必要条件としてのみ認める。十分条件としての必然化を認めると、全ての存在者が全ての必然的真理の TM であるという帰結が得られてしまうからである。この点については、Cameron (2008: 263-4), Lewis (2001: 604), Restall (1996: 332-3)などを参照。

iii TM としての事態の存在を擁護する代表的な TM 論者としては、Armstrong (1997; 2004)などを参照。また、ここでは取り上げないが、TM としてトロープ trope という存在者を擁護する論者もいる。こちらについては、秋葉 (2014)や Mulligan et al. (1984)などを参照。

iv 例えば、Dyke (2007), Heil (2003), 北村 (2014)などを参照。

v Quine (1948)を参照。

vi 以上の点をもって、TM 理論は無用の長物であるとする批判もある。例えば、Dodd (2002), Hornsby (2005)などを参照。

vii 例えば、Rodriguez-Pereyra (2002, 38-40)を参照。

東京大学

(5) 参考文献

Armstrong, D. M, 1997, A World of States of Affairs, Cambridge University Press.

——— 2004, Truth and Truthmakers, Cambridge University Press.

Cameron, R. P. 2008, "Truthmakers and Modality," Synthese, 164, 261-280.

Dodd, J, 2002, "Is Truth Supervenient on Being?" *Proceedings of the Aristotelian Society*, 102, 69-86.

Dyke, H, 2007, Metaphysics and the Representational Fallacy, Routledge. Heil, J, 2003, From an Ontological Point of View, Oxford University Press. Hornsby, J, 2005, "Truth without Truthmaking Entities," in H. Beebee & J. Dodd (eds.), Truthmakers: The Contemporary Debate, Clarendon Press, 33-47.

Lewis, D, 2001, "Truthmaking and difference-making," Noûs, 35, 602-15. Mulligan, K, Simons, P. and Smith, B, 1984, "Truth-Makers," *Philosophy and Phenomenological Research*, 44, 287-321.

Quine, W. V. O, 1948, "On What There Is," Review of Metaphysics, 2, 21-

36.

Restall, G, 1996, "Truthmakers, entailment and necessity," Australasian Journal of Philosophy, 74, 331-340.

Rodriguez-Pereyra, G, 2002, Resemblance Nominalism: A Solution to the Problem of Universals, Oxford University Press.

秋葉剛史, 2014, 『真理から存在へ 〈真にするもの〉の形而上学』,春秋社. 北村直彰, 2014,「存在論の方法としての Truthmaker 理論」,『科学哲学』, 47, 1-17. 心の哲学と神経科学における心脳問題と心的因果の取り扱いに関する論考

櫻井圭介

Abstract

In the philosophy of mind, mental causation problem and mind-brain problem are being discussed most actively. In the process of solving these problems, various opinions about the ontological position of qualia have been advocated. Neuroscientific findings are also important clues to think about the problem of consciousness because it is thought that the specific activity of the brain is creating consciousness. Therefore, referring to neuroscientific findings, I examined which position to take in order to solve mental causation problem and mind-brain problem. As a result, I judged that Epiphenomenalism is the strongest candidate.

(1) 研究テーマ

物質である脳の情報処理になぜ、また、どのようにしてクオリアという心的現象が生まれるのかという意識のハードプロブレムがデイヴィッド・チャーマーズにより、提起されてから、心の哲学の分野において、心脳問題や心的因果についての議論が盛んに行われるようになった。また、神経科学の分野で研究されている問題はほとんどが意識のイージープロブレムに分類される問題ではあるものの、脳の特定の活動が意識を作り出していると考えられるため、意識の問題を考える上では、神経科学的な知見も重要な手がかりとなる。そこで、この論文では、心の哲学者であるデイヴィッド・チャーマーズの精神物理法則や、神経科学的な知見を参考にして、心脳問題と心的因果の問題についての論考を進めることを目的とする。

(2) 研究の背景・先行研究

心脳問題や心的因果の問題を考える際には、心的現象であるクオリアをどのようなものとして扱うかによって、様々な立場があり、それぞれの立場での利点や問題点が存在する.よって、ここでは、クオリアの取り扱いに対す

る代表的な立場の哲学者の主張に対する利点と問題点を確認し、心脳問題や 心的因果の問題を解決するためには、どの立場を取るのが最も適当であるか を考える.

古来より、様々な宗教においても、魂と身体の関係についての議論が交わ されてきたが、デカルトが実体二元論を主張したことにより、心脳問題が哲 学の最も重要なテーマの一つとして、議論されるようになった. 実体二元論 では、この世界には魂や自我、意識などと呼ばれる心的実体が脳や身体など の物理的実体と独立して存在するという立場を取る、この立場では、脳が情 報処理を行い、情報を魂に伝え、魂から脳に指令を出すという脳と魂が相互 作用をするモデルを提唱する.様々な宗教が魂と身体を別々の独立したもの として考えていたように、実体二元論は人間のありふれた世界観であるもの の,以下の理由により,現在ではこの立場を支持する哲学者は少ない.まず, 第一にデカルトが考えたように、魂のような心的なものと脳のような物理的 なものが相互作用すると考えると現在の科学的な知見との矛盾が生じるから である.神経科学の発展により,脳においても,他の物理的な現象と同じよ うに、それぞれの神経細胞を構成する分子や原子の振る舞いの結果として、 作動しており,人間の感情や行動は物理法則に従って,制御されていると考 えられているため,そこに魂のような心的なものとの因果関係を考える必要 がないようになってきた、このように、現代科学では物理的領域での因果的 閉包性が前提とされており、その前提を覆し、心的なものと物理的なものと の相互作用を認めると不都合が生じる、例えば、魂などの心的なものの働き により、脳の状態が変化して、人間の感情や行動が変化するならば、物理的 な力なしに脳を構成する分子や原子の運動が変化することとなり、エネルギ 一保存則との矛盾が生じる.1また,実体二元論では,魂が脳に作用して,意 図的な行動を引き起こすと考えられているが、神経科学的な見地からのその 説を覆すような指摘がなされている.これは,ベンジャミン・リベットが行 った随意運動に先行して準備電位という脳の活動が起きるが、それが行動を 意図した瞬間と一致するかという実験であり、結果は意思が意識される 400 ミリ秒ほど早く準備電位は発生していた.つまり,神経活動を記録すれば, 被験者が指を動かそうと自覚するより早く,その行動を予測することが可能 になるため、結果の解釈の仕方によれば、意識は行動を追認しているだけで あり、自由意志の否定となる.2このような問題点がありながら、実体二元論 を支持する哲学者が存在するのは、この立場を取ることで、自我に関する説 明がしやすくなるという利点が存在するからである.つまり,なぜ様々なク オリアが一つの私という体験として感じられるのかという問題に関して,魂

のようなものが存在し、それがクオリアを知覚しているという説明ができるということである。しかし、この利点においてでも、以下のような問題点が存在する。もし、脳内のどこかに魂が存在し、それがクオリアを体験しているならば、その魂がクオリアを体験するための魂が必要になるという無限後退に陥る。3また、神経科学的な知見からも、分離脳の実験があり、一つの自我が存在するというわけではないようである。4以上の点から、私は意識の説明をする際に、心的な実在と物理的な実在が存在し、特に、それらが相互作用するという立場を退けることにする。

実体二元論の批判から、心的な実在は存在せず、この世には物理的な実在だけが存在し、心的とされているものはすべて物理的なものの運動に還元可能であるという物理主義が盛んに主張されるようになった。物理主義は一元論であり、心的なものは存在しないとされるため、心脳問題や心的因果の問題を回避できるという利点をもつが、以下の問題点を抱えるため、私はこの立場も正しくないと考える。その問題点とは物理的なものと心的なものは存在論的な区分が異なるため、どうやっても心的なものを物理的なものに還元することができないということである。ソール・クリプキの必然性と偶然性の概念がその問題点の妥当性を支持する。心的なものを物理的なものに還元できるということは、全ての可能世界で、あるクオリアはある神経の発火であるということは偶然的なものであるため、二つの存在論的な区別が異なるということである。5

実際にクオリアを感じているため、私にとって心的な実在を疑うことはできない。よって、残る立場は心的なものだけが実在するという観念論的な一元論になる。この立場を取ることで物理主義と同様に心脳問題や心的因果の問題を回避できる利点があるが、心的なものしか実在していないとすると、実在は知覚している時だけ存在し、知覚をやめると消えるということになる。しかし、この点はこの立場を退ける致命的な問題点にはならないと思われる。なぜなら、聞く人がいない音はするのかという古くからの哲学の問題にあるなずなら、もし、知覚されない実在を認めると、この世界には誰にも感じることがなら、もし、知覚されない実在を認めると、この世界には誰にも感じることが存在を生じさせないためには、知覚している時だけ心的なものが実在するとのできない未知の感覚が無数に存在することになるので、そのような存在を生じさせないためには、知覚している時だけ心的な一元論では、無意識に関する問題点を回避することが難しいという指摘がある。6神経科学的な知見により、我々の精神過程の大半は無意識下で起こることが分かっており、クオリアしか実在しないとすると、無意識的な行動の因果関係の説明を

することが難しくなる. また,神経変性疾患や向精神薬の例からも明らかなように,クオリアや意識の状態は脳の物理的・化学的な状況により,影響を受けて変化するため,脳という物理的なものを切り捨て,心的な実在だけが存在するという立場では説得力に欠けると考え,この立場を退けることにする.

(3) 筆者の主張

以上のように、それぞれの立場に利点はあるものの、致命的な問題点が存在するため、心脳問題や心的因果の問題に取り組むためには、これらの主要の立場を避けて考える必要があると思われる.

チャーマーズは、クオリアは物理的なものに還元できないとし、物理主義を否定する。そして、クオリアをそれ以上還元できない実体としてとらえ、既存の物理学を拡張し、クオリアの振る舞いをも記述しようとする自然主義的二元論を主張する。つまり、チャーマーズがとる自然主義的二元論とは、現代の物理学にはクオリアが含まれておらず、そのため、意識のハードプロブレムを解くことは不可能であるが、その上で脳からクオリアが生まれていることは間違いないとし、クオリアを物理学における時空、質量・エネルギー、スピン、電荷等の他のものに還元できない世界の根底にある特性としてとらえることで、心的なものも含め、全てを自然科学的な方法で説明する立場ということである。そして、チャーマーズは、こうしてもたらされるクオリアと物理系の関係を支配する理論を精神物理法則と名付け、その候補として、情報の二相説が提唱される。

ここで, チャーマーズの提唱する自然主義的二元論が心脳問題や心的因果の問題を解く上で妥当なものであるかを検討する.

自然主義的二元論という考えは、クオリアが物理的なものに論理的に付随 しないことを根拠に物理主義を否定する結果として導かれるが、この点の基 本的な議論は次のようなものである.

- (1) In our world, there are conscious experiences.
- (2) There is a logically possible world physically identical to ours, in which the positive facts about consciousness in our world do not hold.
- (3) Therefore facts about consciousness are further facts about our world, over and above the physical facts.

(4) Therefore materialism is false. (The Conscious Mind, p. 109)

チャーマーズは自然現象にクオリアが論理的に付随しないことを, 哲学的 ゾンビや逆転クオリアの想定可能性を根拠に論証する. ここで, チャーマー ズの論証を支える哲学的ゾンビと逆転クオリアについて問題点がないか考え るが, 前者が確証する結論は後者のそれに比べて強いために, それぞれの想 定可能性によりもたらされる結論は微妙に異なる. そこで, まず, 哲学的ゾ ンビについて考える. この世界を見返してみると, 現実的に, 人間の認知, 学習、判断などの高度な精神活動には必ずクオリアが伴っており、人間など の高度な生命体では哲学的ゾンビは不可能であるという批判が存在するが、 その批判に対して、神経科学的な知見から反論を加える. ヒトの小脳は大脳 の 10 分の 1 ほどの大きさであるが、神経細胞の数は大脳の 200 億個に対し て,800 億個存在し,複雑な情報処理を行っている.そして,運動を調節し たりする認知機能を持つが, 小脳の活動が意識にあがってくることはなく, 複雑な認知課題をこなしている時に必ずしもクオリアが伴っている訳でない. そもそも、哲学的ゾンビの論証で問題になるのは、想定可能性であり、形而 上学的に可能であれば,現実に即していなくても構わない.続いて,逆転ク オリアについて考える. 逆転クオリアの場合は想定可能性についても哲学的 ゾンビの例のようにすんなりと受け入れることが難しいと思われる.例えば, 自分にとっての赤のクオリアが他の人間にとっての青のクオリアに見えてい るというような色についての逆転クオリアを持った人が存在することは想定 可能であるように思える. しかし, 同じ逆転クオリアでも, 快と不快が逆転 している人を想定することは難しいのではないかと思われる. そこで, 色と 違い快の場合はなぜ逆転クオリアがすんなりと受け入れないのか考えてみる と,色というのは例えば赤の場合,700nmの光を見た時に生じる赤いという クオリアという二つの関数によって定義されるのに対し, 快不快は例えば痛 みの場合、ある神経線維が興奮したときに起こる痛いというクオリアという 二つの関数に加えて,それを避けようとするという行動を引き起こすものと いう三つの関数によって定義されるからではないかと思われる.痛みの3つ の関数はデイヴィッド・ルイスの論文の主題にもなっている狂人の痛みと火 星人の痛みという思考実験によってそれぞれの違いをはっきり分かることが できる.

> There might be a strange man who sometimes feel pain, just as we do, but whose pain differs greatly from ours in its causes and

effects. Our pain is typically caused by cuts, burns, pressure, and the like; his is caused by moderate exercise on an empty stomach. Our pain is generally distracting; his turns his mind to mathematics, facilitating concentration on that but distracting him from anything else. (Mad Pain and Martian Pain, p. 216)

このように色のクオリアはそれを感じた時に何かの行動を引き起こすこと がないのに対し、快不快のクオリアは不快を避け、快をさらに受けようとす るという行動を引き起こすという点が逆転クオリアの想定可能性の違いにな っている.そもそも快不快の概念は生存における選択有利性によって説明が でき、生存に有利な行動を促進し、不利な行動を抑制するような生物が自然 選択によって残っているが、生存に不利なため淘汰された、痛みのような不 快を促進しようとする生き物も存在するだろうという反論もあるかもしれな い. しかし, 逆転クオリアの場合, 行動とは関係ないため, 選択有利性とも 関係ない、つまり、快不快が逆転クオリアになっている人間は他者から見分 けることができず, 実際に快を求めて行動しているように見えるものの, 本 人は不快のクオリアを感じており,もし,クオリアに機能があるならば,心 の中では、そのクオリアを避けたいと思いながらも快を求める行動をとって いるのかもしれない. 快のクオリアと記憶の関係について考えてみた場合, もし快が瞬間的なものであり、記憶にも残らないと仮定した場合、それを求 めるという次の行動を引き起こさないため、快のクオリアそのものと行動は 切り離すことができるという反論もあるかもしれないが、瞬間的であれ、ク オリアは全一的であり、快のクオリアはそれを感じたいという機能をそれ自 身で備えているように思われる.このように,チャーマーズの論証を支える 重要な思考実験である逆転クオリアは、クオリアの種類によっては、想定可 能性をすんなり受け入れられるものではないと考える.

チャーマーズは物理主義の否定から、クオリアを還元できない実体としてとらえた自然主義的二元論を主張し、情報を究極的な実在した情報の二相説を提唱する.しかし、この説は信憑性に欠けると考える.なぜなら、情報の二相説を推し進めると、情報処理をする全てのもの、例えば、サーモスタットもクオリアをもつはずであるという汎心論に近い説をチャーマーズは主張するが、先ほどの小脳の例からも明らかなように、複雑な情報処理をするだけでは、必ずしもクオリアが生じる訳ではなく、神経科学的な知見からは、ある段階からクオリアが生じてくる創発説の方が正しいのではないかと思わ

れるからである.創発説を支持する神経科学的な仮説として、ジュリオ・トノーニの統合情報理論が挙げられる.この仮説はある身体システムは、情報を統合する能力があれば、意識があるというもので、睡眠・麻酔・昏睡などの研究から明らかになったものである.創発説では、砂山のパラドックスが生じるという反論が予想されるが、睡眠から覚醒への例でも明らかなように意識の明瞭さには段階があり、統合情報理論では、Φと呼ばれる統合された情報量の単位と対応する.また、クオリアが物理学の他の基本的な特性と同じようなものだとした場合、情報処理を行う段階は無数に存在するが、個別の脳という段階で統一された意識が生じることに対する説明が難しいと思われる.なぜなら、個々の情報処理で生まれたクオリアが集まって個別の脳に対応する意識を作っているという説明では、クオリアは知覚した段階で、それ自体で完結するものであり、それ以上基本的なクオリアで還元できるようなものではないということとの矛盾が生じるからである.

このように、自然主義的二元論にも問題があるため、残る立場はチャママーズも心惹かれるところがあると述べている随伴現象説となるため、随伴現象説では痛みのクオリアと痛み止めの薬を飲むという行動が無関係とないと思われる。なぜなら、思考が行動の原因である場合も随伴する場合も同じように行動に思考が伴うので、どちらの場合も随伴する場合も同じように行動に思考が伴うので、どちらである。しかし、随伴現象説にていたが正しいと知ることはできないからである。しかし、随伴現象説ももにどいの逆転クオリアが見えていた人間が神経細胞等の物理的の性質は変わらず、次の瞬間に青のクオリアが見えるようにクオリアが逆転した場合、その変化に気づくと思われるが、これは、クオリアが脳という物理的なものに作用したということとなり、随伴現象説との矛盾になる。この問題はチャーマーズが提起した現象判断のパラドックスと同種の問題であると思われるが、私は現在、この問題に対して、説得力のある反論が用意できていない.

(4) 今後の展望

今までの議論で実体二元論,物理主義,観念論などは,利点は存在するものの,重大な問題点が存在するため,受け入れ難い立場であることが明確になった.また,チャーマーズの提唱する自然主義的二元論も心脳問題や心的因果の問題を解く上で非常に有用な考え方ではあるものの,クオリアの私秘性や全一性が置き去りにされており,そのまま受け入れることができない.

この論考を経て、私が最も有力な候補だと考えている立場は随伴現象説である.しかし、この立場も現象判断のパラドックスという大きな問題点を抱えている.この問題点をうまく解決することができれば、心脳問題や心的因果の問題の解決に向けて大きな前進となる.従って、今後は現象判断のパラドックスの解決の一点に絞って研究を行う予定である.

注

- 1. Dennett(1995), pp.34-5
- 2. ただし、リベット自身は、我々は意思に対して拒否権を行使することができるので、この実験結果によって、自由意志が否定されることはないとしている.
- 3. Dennett(1995), pp.101-38
- 4. ロジャー・スペリーの研究が行った、脳梁を切断した患者に対する研究より、それぞれの脳半球が独立した意識を持っていることが実証された.
- 5. 武田(2003), p98
- 6. 信原(2002), p25-6

東京大学大学院新領域創成科学研究科博士課程1年

(5) 参考文献

- [1] Chalmers, D, 1995, The Conscious Mind, Oxford University Press.
- [2] Dennett, D, 1991, Consciousness Explained, The Penguin Press.
- [3] Jackson, F, 1982, "Epiphenomenal Qualia.", *Philosophical Quarterly*, 32, 127-36.
- [4] Kripke, S, 1980, Naming and Necessity, Harvard University Press.
- [5] Lewis, D, 1980, "Mad pain and Martian pain.", Readings in the Philosophy of Psychology, 1, 216-22.
- [6] Libet, B, C. A. Gleason, *et al.*, 1983, "Time of conscious intention to act in relation to onset of cerebral activity: The unconscious initiation of a freely voluntary act", *Brain*, 106, 623-42.
- [7] Negel, T, 1986, The view from nowhere, Oxford University Press.
- [8] Ryle, G, 1949, The Concept of Mind, University of Chicago Press.
- [9] Searle, J, 1992, The Rediscovery of the Mind, The MIT Press.

新進研究者 Research Note

- [10] Tononi, G, 2012, Phi: A voyage from the brain to the soul, pantheon.
- [11] 信原幸弘, 2002, 『意識の哲学一クオリア序説』, 岩波書店.
- [12] 武田一博, 2003, *D.チャーマーズは心の唯物論を論駁したか*, 『沖縄国際大学総合学術研究紀要』, 6, 93-109.

新進研究者 Research Note 理由のために行為することと実践的推論 Acting for Reasons and Practical Reasoning

中根 杏樹

Abstract

"Practical reasoning", has been considered as indispensable for acting for reasons. Is this claim right? This paper deals with this question. In addressing this question, it should be noted that practical reasoning has two aspects: aspect as an act of mind and aspect as a propositional structure. I seek to answer the question positively from both aspects by developing Velleman's view. In my opinion, practical reasoning as an act of mind is that an agent take one's action to be supported by promises. As far as we describe our action in terms of our end, we act for reasons in virtue of practical reasoning.

(1) 研究テーマ

「実践的推論 (practical reasoning)」は、「理由のために行為する」ことの必要条件に関わると考えられる. 本稿の研究テーマは、実践的推論である.

(2) 研究の背景・先行研究

なぜ、理由のための行為を説明するために、「実践的推論」について考えることが助けになるのだろう。それは、理由のために行為するとき、行為者は、その理由から為すべき行為を導き出している、と考えられるからだ。「なぜ彼に道を教えていたの?」「道が分からず困っていたから、助けようと思って」。この例では、「道を教えることが彼を助ける手段であるがゆえに、当該の行為を行うべきである」という行為者の思考が示されている。「理由」は日常的に「根拠」とも言い換えられるが、この例では、その目的と手段が根拠となって、行為を為すべきものとして導いている。なんらかの命題を根拠として結論を導くことを「推論」と言うのだから、理由のための行為は実践的推論と切り離せない関係にあると考えられる。

この関係を捉えて、アンスコムは、特定の行為において、特定のことがらが理由であるのは、手段一目的、部分一全体の関係を表すような実践的推論

の前提として働く場合のみであると主張する(Anscombe, 1957, pp.79-80; 150-152 頁)i. 他の論者もこれに同意し、デイヴィドソンは、欲求と信念が当該の形式をとって行為を惹き起こすとき、それらの欲求と信念が基本理由となると考えていた(Davidson, 1963, p.5; 5 頁).

この見解は妥当だろうか. われわれは理由のために行為する際, つねに実践的推論にもとづいて行為を導いているのだろうか. この問いが, 本稿の中心的な問いとなる. だが, この問いは, 実践的推論が様々な側面を有するために, 答えるのが難しい. 実践的推論には心的側面と非心的側面がある. 一方で, 実践的推論は, 行為者の, 為すべき行為についての思考であると考えられる. しかし, 他方で, 実践的推論は, 前提を根拠に別の行為を結論づけるものであると考えられる. 根拠づけるという関係は, 人の心理ではなく, その内容である命題に由来するから, この側面においては, 実践的推論は命題的構造であると考えられる。 実践的推論には二つの側面があるため, 当該の問いには幾つかの答え方がある. 本稿は, 両者の関係を緊密なものとして理解し, 両者に関して肯定的に答えたい.

理由のための行為がつねに実践的推論にもとづくかという問いを考えるうえで筆者が焦点を当てたいのは、実践的推論の他の諸側面である。これはあまり意識されない点であるが、その役割に関して、実践的推論には二つの側面がある。一方で、実践的推論は行為を為すのがよい、役立つポイントを示す(Anscombe、1989、p.114; 199 頁). 他方で、実践的推論は行為の記述にかかわる。意図的行為はなんらかの記述のもとで意図的である。先の例では、当該の人物の行為は「人助けをする」という記述のもとで意図的であるが、「空気を振動させる」という記述のもとでは意図的でない。この行為記述に関して、実践的推論は、行為の記述が有する秩序であると考えられる(Anscombe、1957、p.80; 152 頁)iii. この点は問題ではないと思われるかもしれないが、実践的推論の心的側面についてどのように考えるかを左右する。本節の残りでは、心的側面に関する議論を紹介する。(3)節では、そこで紹介された D.ヴェルマンの見解を発展させるという仕方で、理由のための行為がつねに実践的推論にもとづくかという問いに答える道筋を示したい.

さて、実践的推論の心的側面とはどのようなものだろうか.実践的推論は 為すべき行為を導くような思考であると述べた.われわれはときに、「何を 為すべきか」という問いに答えるという目的をもって、行為の創案、結果の 想像といった多様な思考を行う.実践的推論とは、このような、「何を為すべ きか」の問いに答えるという目的をもって行われる心的行為であると考える かもしれない.ここから、理由のために行為するときには、事前にこうした 心的行為を行っているのだ、という単純な見解が導かれる.

この単純な見解はアルパリーらの反論に直面する.彼らは、心的行為であるような自発的な思考を「熟慮(deliberation)」と呼ぶiv.彼らは、熟慮が心的行為であるという事実をもとに、先の単純な見解(事前熟慮説)を批判する.彼らが指摘するのは、無限後退の問題である.熟慮はそれ自体が心的行為である.そうすると、事前熟慮説によれば、その熟慮のための事前の熟慮が行われていたはずである.熟慮のための熟慮もそれ自体が心的行為なのだから、その熟慮のために事前の熟慮が行われていなければならない.こうして、

このような考察から、彼らは、熟慮説には見込みがなく、理由のために行為することは、非自発的な過程によって可能だと結論づける(ibid、p.230). アルパリーらは、論文の最後で、「理由応答は関連する異なる態度間の〔…〕 論理的関係のゆえに生じる心的状態の移行以上のものは必要としない」のであり、それは因果的な移行であると述べている(ibid., p.238).

事前熟慮説は無限後退に陥る(Arpaly and Schroeder, 2012, pp.219-220).

さて、アルパリーらの論証を受け、彼らとは異なる方向で展開したのがヴェルマンである.彼は「実践的推論」の記述を与えるという役割に目を向け、これを自発的な心的行為、熟慮ではなく、非自発的な思考として理解する.彼によれば、実践的推論とは、行為中を通して、行為が意味をなすことを保証する思考であるv.彼の説明では、それは、行為中に行為を監督するという役割を担う.実践的推論は、「行為が意味をなすものであることを保証する」という監督役割を「行動の流れを行為概念の下に置く」ことによって果たす.行動を行為概念のもとに置くとは、たとえば、指を動かすという行動を「電気のスイッチを点ける」という概念のもとで記述することである.(Velleman、2015、pp.168-169).ヴェルマンの見解では、「われわれはつねに「何を為すべきか」に答えるという目的をもって行為の事前に心的行為を行っている」のではない、人が実践的推論を行っているということは特定の記述のもとで行為の意味を理解しているということであり、それで十分なのである.

(3) 筆者の主張

アルパリーらは、理由応答は関連する異なる態度間の論理的関係のゆえに 生じる因果的な心的状態移行以上のものは必要としないと言う.まず、彼ら は議論なしに理由応答を「過程」と想定している点に問題がある.さらに、理 由に応じることの説明として、これで十分な説明になっているか、という疑 問がある.理由のために何かを行ったと言えるには、理由となることがらを 理由として理解している必要があるように思われる.たとえそれが論理的関 係のゆえに生じるのだとしても,しかじかの状態が行為者に因果的に生じた というだけでは,理由のために行為していると言うのに不十分ではないか.

推論を心的行為として捉える論者はこの点を重視する。P. ボゴシアンは,推論はすべて自発的な心的行為であると述べる(Boghossian, 2014, p.2)。なぜなら,「推論はわれわれが行う何かであり,たんにわれわれに生じることではない。それは $[\cdots]$ 目的をもって行われる何かである」からだ(ibid., p.5)。彼の考えでは,この点を捉えるために,推論の理論は「見做し条件(Taking Condition)」に適合しなければならない。それは「推論は必然的に,思考者が自分の前提が自分の結論を支持すると見做しており,その事実のゆえに結論を引き出すということを含む」という条件である(ibid.) [強調は原著者].

アルパリーらは「見做し条件」に適合する説明を与えていないように思わ れる. 「見做し条件」は推論の理論が満たすべきものとして与えられている が,理由応答の説明一般は行為者が理由と見做して行為することを説明する 必要があるため、少なくともこれに類する条件を認めるべきだろう. だが、 ボゴシアンは見做し条件を「推論はたんに生じることではなくわれわれが行 うことだ」という点を捉えるために導入しているので、見做し条件を認める ことで、理由応答・推論は心的行為であるという見解を認めることになるよ うに見える.そうすると、無限後退の問題に陥るように思われる.しかし、 ボゴシアンが行うように「たんにわれわれに生じる」ことではないことから 「理由応答・推論は心的行為だ」ということをただちに導くことは、できな い. たとえば、「これは人助けだ」という信念は行為ではないが、それは行為 者によって信じられていることであり、たんに行為者に生じるものとしてで はなく、行為者に帰属する. 「前提が結論を支持すると見做す」ことを行為と 捉える必要はないだろう.したがって、理由応答・推論の理論は見做し条件 に適合する必要があるということは、無限後退の問題に陥ることなしに認め られる. そして、アルパリーらの見解は見做し条件に適合する説明を与えて いないので、十分な説明ではない.

つぎに、ヴェルマンの見解をみてみよう。彼の見解は、実践的推論が行為の再記述、行為の理解と関わるということを指摘している点で重要であると思う。「何を為すべきか」という問いに答えるという側面にのみ注目すると、それは心的行為であると考えるほうに傾いてしまう。だが、ヴェルマンの見解は「記述を与え行為を理解している」という点を考察に含むことで、心的行為ではないものとしての実践的推論がどのようなものかを把握する一つの道筋を示している。われわれは自分が何をしているのかを熟慮せずとも理解している。心的行為として行われないような実践的推論はこの理解の働きで

ある、と、このように考えることで、彼の見解は、アルパリーらの示す無限 後退の問題を回避し、さらに、見做し条件が捉える行為者の状態を考慮に入 れることができている.

とはいえ、彼の見解に欠落があるのは明らかだ.彼の説明で実践的推論の説明が尽くされるとしよう.そうすると、第一に、ヴェルマンは行為の記述を与えるという側面のみを強調するため、その見解は、これだけでは推論の説明にはならないように思われる.見做し条件が示すように、推論は、前提と結論の間に成り立つ根拠づけの関係を理解したうえで結論を引き出すものだと考えられるvi. 前提が結論を根拠づけるという点を説明に含む必要がある.この問題から派生して、第二に、彼の見解は、実践的推論の「為すべき行為を導く」という役割に説明を与えていない.この役割にも説明は与えるべきだろう.この役割は,前提が結論を為すべきものとして支持するということによって説明される.以上の問題から、「前提が結論を根拠づける」という実践的推論の側面をヴェルマンの見解に含むべきである.つまり、冒頭で述べたような非心的側面をこの見解に含む必要がある.

では、この側面とはどのようなものだろうか、それは、ある行為を別の行為を結ぶ関係、つまり自的一手段、全体一部分の関係(計算的(calculative)関係)を表す命題の構造だと考えられてきた(Anscombe, 1957; Schwenkler, 2015; Vogler, 2002). 目的一手段関係は、たとえば、風邪を治すこととネギを巻くことの関係で、全体一部分関係は、道路を渡りきることとそのための第一歩の関係であるvii. 例を用いて説明したい、ある人は「風邪を治す」ために以下のような計算的構造にもとづいて行為を導く.

(目的・大前提) 私の風邪が治る

(前提) 血流がよくなるならば、私の風邪が治る

(前提) 私が首にネギを巻くならば、血流がよくなる

(結論) 私は首にネギを巻く

この例では、目的の観点から、「首にネギを巻く」べきである.諸命題が計算的構造を形成するとき、前提はその行為を支持し、結論は前提の観点から「為すべき行為」として示される.一方で、計算的関係を表さないような命題(例:「ネギは美味しい」)は、「ネギを巻く」という行為を支持する根拠ではない.この構造に現れないがゆえに、当該の命題は根拠、理由として分節化されないのである.行為においてある命題が前提となり結論を根拠づけるためには、計算的構造を形成する必要がある.

さらに、計算的関係を表す命題の構造は、前提が結論を根拠づけ、結論を 導出するための構造であると同時に、ヴェルマンが重視する「行為の再記述」

を支える構造でもある(Anscombe, 1957, pp.46-47; 88-89 頁). これらの命題が、前提が根拠となり為すべき行為を結論づける計算的構造を形成しないのであれば、当該の行為に対して「風邪を治す」という目的による記述も与えられない. 「ネギは美味しい」という命題は、与えられた情報のかぎりでは、他の命題と記述を繋ぐ関係を持たないため、当該の行為に「風邪を治す」行為としての記述を与えることはできない. 一般化すれば、根拠、理由をもちだして目的から行為を再記述することが可能であるのは、計算的構造で諸命題が前提・結論として働く場合のみである.

この計算的構造を考慮にいれて、ヴェルマンの見解を修正するなら、どの ようになるだろうか. 先に、推論のポイントは「前提と結論の間に成り立つ 根拠づけの関係を理解したうえで」結論を引き出すことにあると述べた. ヴ ェルマンの見解にこの点を含めると,実践的推論は,前提と結論の間になり たつ根拠づけの繋がりを理解して,目的の観点から行為に記述を与えている ことである,ということになる. そして, いま見たところでは, その繋がり とは、計算的構造である. 実践的推論は、特定の記述を受ける行為に関して それがどのような計算的構造からその記述を与えられるのかを理解している ということである.このように言うことは、行為者が「根拠」といった概念を 用いて前提と結論の繋がりを把握していなければならないと言うことではな い. たとえば、最初の事例の人物は、「なぜ?」と問われれば、「人助けにな るから」と答え、「それが結論の根拠である」とは言わないかもしれない. し かし、鴻(2014)で述べられるように、問われればそれを答えに提示すると いうその傾向性によって、「人助け」を当該の行為の根拠として理解し、それ ゆえに行為していると行為者は示している. 理由として持ち出すという傾向 性を有していれば、根拠として理解していると言うのに十分である.

「為すべき行為を示すという役割」に関して言えば、計算的構造は再記述を支える構造であると同時に、目的、その他の前提の観点から為すべき行為を示す構造であるため、行為の記述間の繋がりを理解していることは、その行為を為すべきポイントを理解しているということである。実践的推論が「何を為すべきか」という問いに答えるという側面をもつと認めたとしても、それが心的行為であると言わねばならないわけではない。

以上から「あることがらが行為のための理由であるのは、実践的推論の前提として働く場合のみであるか?」という問いに、筆者は、非心的側面と心的側面の両方から、肯定的に答えたいと思う。われわれが理由のために行為するとき、その理由は、計算的構造にもとづいて、行為を支持するものとして分節化される。そして、目的の観点からの記述は、計算的構造によって可

能になっている.したがって、非心的側面から、当該の問いに肯定的に答えることができる.心的側面に関して、実践的推論を熟慮としてのみ理解すると、当該の見解は無限後退の問題に陥る.しかし、それを行為だと考える必要はない.行為を前提の観点から理解していることであると考えられる.ヴェルマンが注目した通り、理由のために行為している人は、目的によって行為を再記述しているはずである.そのかぎりで、行為者はその根拠づけの関係を理解してその行為を引き出している.したがって、心的側面からも、当該の問いに肯定的に答えることができる.

(4) 今後の展望

(3)節の見解を展開し、「では、その理解とは?」と問うことが必要であるのは明らかである。しかし、ここでは、これをメタ倫理学の議論にどう応用できるかということを考えたい。(3)節では、行為を説明する理由(動機理由)を主題にしていたが、行為を正当化する理由(規範理由)について考察するために用いることができる。メタ倫理学では、規範理由の源泉をめぐって、二つの立場が対立している。その立場とは、内在主義(internalism)と外在主義(externalism)である。B.ウィリアムズによって提唱された内在主義には、規範理由は健全な熟慮の経路に現れるという洞察がある(Williams、2001)。この洞察は、内在主義を直接的に支持するものではなくても、規範理由を考えるうえで重要な示唆であり、前提を加えれば、内在主義を支持する。本稿の見解から、規範理由に二つの仕方でアプローチすることができる。

第一に、「熟慮の経路」を本稿で述べた非心的な実践的推論として理解する道がある. つまり、規範理由は実践的推論の前提に現れると考える道がある. C.フォーグラーは、これに加えて、規範理由は行為者相対的であると考え、積極的に内在主義を支持する (Vogler, 2002. ch.8).

第二に、心的側面として理解し、それ自体の積極的な役割を主張する道がある. C.コースガードによれば、熟慮は、理由を与えるような実践的アイデンティティを統一し、その行為者の全一性を保つという役割を果たす (Korsgaard、2009、pp.125-126). 彼女の見解は、熟慮、理由、行為者性の間の連関を示唆しており、検討の価値のある道筋を示している.

i アンスコムはさらなる目的を目指した「意図を伴う」行為に限定して主張 している. 本稿でも「行為」は以下そのようなものだと理解してほしい.

ii 「命題的構造」という用語は Audi (2006, pp.86-87) に負っている. それは、妥当あるいは妥当でないとして論理的に評価する際にわれわれが言及す

るような、抽象的な命題の構造である(ibid.).

iii 「記述の有する秩序」については(3)節で説明する.

iv 本稿の範囲では、「自発的」をその行為者がなんらかの目的を理由として 了解していることを含むものとして用いる.そう読まなければ、アルパリー らの無限後退を示す論証は成り立たないように思われるからである.

v 以前の著作によれば, 行為が意味をなす (make sense) とは, それを理解できる (understand) ということである (e.g. Velleman, 1989, p.182).

vi このことを言いあらわし、アンスコムは「[ある命題から別の命題が帰結するという] そのことが分かる (seeing) という以外に、実際に推論をしている (inferring) ということが意味することが何か他にあるだろうか. 私は、ないと思う」と述べる (Anscombe, 1989, p.112; 196 頁).

vii 「計算的関係」という用語は、Vogler (2002) に負う. 「によって関係」と同様のものである. その関係を表す命題的構造を「計算的構造」と呼ぶ.

(慶應義塾大学)

(5) 参考文献

Anscombe, G. E. M. (1957). *Intention*. Harvard University Press. 〔菅豊彦訳(1984). 『インテンション』,産業図書.〕

--(1989), [2005]. "Practical Inference". in *Human Life, Action and Ethics*. Imprint Academic: pp. 109-148. [門脇俊介・野矢茂樹監訳(2010). 「実践的推論」, 『自由と行為の哲学』, 春秋社.]

Arpaly, N. & Schroeder, T. (2012). "Deliberation and Acting for Reasons". in *Philosophical Review* 121(2): pp. 209-239.

Audi, Robert (2006). Practical Reasoning and Ethical Decision. Routledge. Boghossian, Paul (2014). "What is inference?" in Philosophical Studies 169 (1): pp. 1-18.

Davidson, Donald (1963). "Actions, Reasons, and Causes", in *Essays on Actions and Events*. Oxford University Press. 〔服部裕幸・柴田正良訳 (1990). 「行為・理由・原因」, 『行為と出来事』, 勁草書房.〕

Korsgaard, Christine M (2009). Self-Constitution: Agency, Identity, and Integrity. Oxford University Press.

Schwenkler, John (2015). "Understanding 'Practical Knowledge'". in

Philosophers' Imprint 15.

Routledge.

Velleman, David (1989). Practical Reflection. Princeton University Press.

—— (2015). "Time for Action". in Time and the Philosophy of Action.

Vogler, Candace A (2002). Reasonably Vicious. Harvard University Press.

Williams, Bernard (2001). "Postscript" in *Varieties of Practical Reasoning*, Elijah Millgram (ed.). MIT Press: pp.91-97.

鴻浩介(2014). 「実践的推論の二つの位相」, 『哲学の探求』(41):27-48 頁

認知的侵入可能性と認識的影響 The Cognitive Penetrability and Epistemic Influence

原田夏樹

Abstract

Is it possible that what one believes or thinks influences what one sees? If so, then this relation is called the cognitive penetration of perception. In this paper, I'll do two things. First, I'll introduce some problems with the cognitive penetrability—an epistemic problem and a cognitive scientific one—and suggest there's confusion in the discussion of the cognitive penetration. Second, I'll refer to the cases in which perceptual experiences can't provide enough justification for believing such-and-such because of cognitive penetration. This is the case of the epistemic downgrade. Then, I'll argue that the cognitive penetration does not always lead to the epistemic downgrade.

(1) 研究テーマ

認知的侵入可能性(cognitive penetrability)とは,知覚以外の要素(信念,欲求,期待などの先行する心的状態)が知覚経験の内容に対して影響を及ぼすことである.認知的に侵入可能であるならば,「同一の外的条件のもとで同一の外的刺激に対して,二つの主体が異なる内容の視覚経験を持つこと」が可能である.一般に,信念や思考が知覚によって影響を与えられるのは当然のことであると考えられているが,信念などの心的状態が知覚に影響を与えることはありうるだろうか?本稿は認知的侵入可能性をめぐる問題の概観と,そこから生じる認識論的問題について論じることを目的としている.

そもそも、なぜ認知的侵入が問題となるのだろうか.この問題は、知覚の哲学において近年盛り上がりを見せており、それを認めるか否かによっていくつかの重大な帰結が導かれるi.

①例えば、観察の理論負荷性.「知覚観察はテストされている理論に対して中立的であり、その理論を支持したり退けたりする証拠となるものである」という経験主義者たちの想定に対して、ハンソンらは「知覚は我々の理解や知識を纏って成立しているのであって、純粋無垢の感覚与件などは存在しない」と主張した(Hanson 1958). 観察が、先行する心的状態によって影響されうるというのなら、当の観察に基づいて合理的に理論を決定することが困

難になってしまうのではないだろうか?

②正当化の問題.これは①を認識論一般のものとして広く捉え直したもの である. ここで重要なのが,「知覚はそれに基づく信念を正当化する」とか「知 覚は知識を与えうる」といった考え方である.ある主体が視覚経験を持つと き、ふつうはその経験に基づいて信念ないし判断が形成される.知覚経験は 信念を認識的に正当化するのである.そうして形成された信念が正しければ, 主体は世界についての何かしらの知識を持つことになる.知識というものは 究極的には知覚経験に基づいているのである.こうした知覚に基づく信念形 成ルートは、信念同士のそれと類比的に考えることが可能である. なぜか? それは、知覚経験がそれ自体で意味を持ち、信念や文に真偽を問えるのと同 様、正確であったり不正確であったりする状態として分析されるようになっ たことと関係している、現在の知覚の哲学者たちの多くが志向説を採ってい るが、これは端的に言えば「知覚経験は表象的・志向的内容をもつ」という ものである.知覚経験は、世界をあるあり方をしたものとして表象し、表象 される世界のあり方(表象内容)の観点から特徴付けられるのである(源河 2017a). 知覚経験が外界の出来事について正しいとか正しくないとか言える のならば,自然と次のような考えが浮かぶだろう.すなわち,ほかの信念や 推論過程を経由することなく,知覚経験は外界についての信念の正当化に寄 与するというものである.

しかし、背景的信念等によって知覚が影響を及ぼされうるというならば、知覚の持つ認識的な役割(正当化を与える、信念や知識の理由となる)は脅かされるだろう. 認知的侵入可能性を認めてしまうと、認識的な有害性が生じうることになってしまうのである. このことは、いわゆる現象的保守主義に困難を突きつける. この立場によれば、知覚経験を持つことはそれだけでしかじかのことを信じるための一応の(prima facie)正当化を与える、と考えられているからだ.

③心のモジュール説との衝突.これはフォーダーによって提唱されたものであるが、心的モジュールとは、特定の領野に特化してすばやく計算を進めるシステムのことである.フォーダーやピリシンらは心をコンピュータとして捉えようという認知観のもと、心というコンピュータがどんな風にして情報処理を行うのかを明らかにしようとした.彼らによれば初期段階の計算プロセスを行うところのモジュールは、認知的に侵入不可能である.つまり、知覚システムがモジュールであるならば、認知的に侵入不可能であることになる.

(2) 研究の背景・先行研究

そもそも「認知的侵入可能性」という術語を初めて用いたとされているのが、古典的計算主義者のピリシンであった. 認知的侵入可能性の問題の背景には、古典的計算主義の枠組みがあったのである. 古典的計算主義によれば、心は形式システム、つまり「離散的なアイテムをその形式的な性質に言及した規則に従って操作することで計算を進めるシステム」であると考えられた(戸田山 2004:30). 代表的な Fodor (1983) の心のモジュール説を参照しよう.

フォーダーがモジュールとして想定していたのはあくまで視覚・聴覚に関わる入力系(周辺系)であって、推論・解釈に関する中央系ではない.心的モジュールの特徴のうち、認知的侵入可能性を考えるうえで重要なのが「情報遮断性」である.前述の通り、知覚プロセスであるモジュールが情報的に遮断されているということは認知的に侵入不可能だということである.このことの根拠としてよく挙げられるのが、ミュラー=リヤー錯視だ.矢羽の向きが異なる二直線の長さが等しいと頭ではわかっていながら、依然として両者の長さは異なるものとして知覚されるii.

知覚システムがモジュールであるというフォーダーの仮説は、ピリシンのいう認知的侵入不可能性と密接に関わっている。というのも、彼はフォーダーのいうモジュールの情報遮断性という概念を、侵入不可能である視覚プロセスを説明する際に利用しているからである。また、初期からピリシンが試みていたのは、知覚プロセスを思考から分離することであった。さらにピリシンは、視覚プロセスを初期と後期に分け、認知的に侵入可能なのは後者であり、視覚プロセスの初期段階は侵入不可能だとした。彼のいう初期視覚(early vision)とは、刺激の開始から観察者に中心化された(viewercentered)対象の表象を構築するに至るまでの一連の視覚プロセスのことである。これに対して認知的に侵入されうる後期視覚(late vision)は、認識や同定を含む。一般に対象認識というものは、視覚的な情報と概念的情報及び長期記憶との相互作用によって担保されると考えられている。この直観を保持するためにピリシンは初期視覚という区分を設け、そこから認識的なプロセスを排除したのである(Raftopoulos&Zeimbekis 2015)。

一方で近年の知覚の哲学においては、認知的侵入可能性を支持する論考が数多く発表されている.以下では、経験科学的な実験データや哲学的な思考実験に基づく考察を紹介するが、そのいずれもが、色や形などの(初期視覚の圏内にあると思われる)低次の性質でさえ侵入されうることを示唆している.

I.「バナナは黄色である」という信念を予め持っていたために, モニター

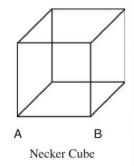
上のバナナの色を灰色に調整するという課題に対して、被験者は青みがかかったように調整してしまう(Hansen *et al.* 2006). これは、バナナについての記憶が色の視覚経験に侵入した例として考えることができる.

Ⅱ.同じ色の顔だが、その顔の各パーツが白人に典型的な特徴を持つか、 黒人に典型的な特徴を持つかによって異なった色に見える.つまり、白人の 場合はより明るく、黒人の場合はより暗く見える(Levin&Banaji 2006).こ れは、人種的なバイアスが色の明るさの視覚経験に侵入した例と考えられる.

Ⅲ. Siegel (2006) において考案された有名な思考実験で、以下のようなものがある. これまで松を見たことがないが、様々な種類の木々の中から松だけを選び抜く仕事に就き、教えてもらっているうちに松を見分ける能力が向上する. こうした能力を獲得する前と後での視覚経験においては、現象的

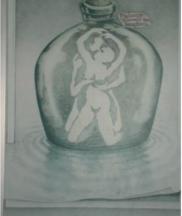
な差異が存する.これは,松に関する知識や信念が視覚経験に侵入したケースと考えることができるだろう.

IV. これらの多義図形は, 与えられている低次の性質(色, 形, サイズ, 位置)には変化が生じていないにもかかわらず, 違った内容の視覚経験を持つことが可能である(Macpherson 2012). 内容が異なるということは現象的性格も異なるということなので, 両者はまったく異なる知覚経験だということになるiii. これらは, 何を見ようとす





Duck/Rabbit Picture



Dolphins/Naked Couple

るかという意図や,見えなかった相を教えてもらうことによって得た知識が, 視覚経験に侵入していると考えられるだろう.

以上,認知的侵入が起こっているように思われる 四つの事例を紹介したが,実際にはどれも,認知的 侵入可能性を支持する根拠としては決め手を欠い

Macpherson (2012) より抜粋

ている. 例えばIVに対しては即座に異論を唱えられるだろう. すなわち, ここに現れている差異は現象的性格におけるそれではなくて, ただたんに判断や解釈の操作が異なるだけだ, というものである. 例えば, アヒルーウサギの二重アスペクト像については, アヒルである, ウサギであるといった二通りの解釈があるだけで, 現象的性格は同一であると考えることができる.

しかしながら、再反論も可能である.上記のような解釈モデルでは、ウサギ性は色や形など低次の性質に基づき、推論過程を経てなされた判断だと考えられている.しかし「ウサギである」という種性質も知覚されうるとした

場合はどうだろうか?それは非推論的に得られた知覚内容に含まれると言う ことができる.

Macpherson (2012) が指摘するように、直観的には、Ⅲに対しては「色や形といった低次の性質を注意深く見て推論・解釈する能力が向上した」という説明を、Ⅳに対しては現象的差異による説明を採るのが好ましく思われる. いずれにせよ重要なのは、両者とも認知的侵入によるものではないと考えることが可能であると言うことだ. 多義図形の視覚経験における現象的差異も、認知的侵入というよりは、注意の向けかえという観点からの説明がもっともらしい.

以上のことから、認知的侵入をはっきりと定義づけるのは困難である. どこまでが認知的侵入なのか、どの事例において認知的侵入があると言えるのかは、はっきり言ってケースバイケースであるというのが実情だ. 色経験におけるそれのように、明らかに認知的侵入があると言いたくなる事例もある一方で、学習による能力獲得という説明の方が適切であるような事例もある. また Machery (2015) では、認知的侵入可能性をめぐる議論が行き詰まっているということさえ指摘されている. 西村 (2017) ではこのことを考慮しつつ、大別して意味論的定義と因果論的定義が挙げられているが、論者の間での「認知的侵入可能性」の定義は依然としてまちまちであるiv.

(3) 筆者の主張

前節で見たように、認知的侵入可能性をめぐる議論は錯綜している.しかしながら、知覚経験が歪んでしまった場合に認識論的に何が起こるかを考えること自体は、哲学的に大変有意義なことである.本節では、認知的侵入を認識論的な問題として捉え直し、それがどんな認識的影響をもたらすのかについて扱う.すなわち、認知的侵入を「先行する心理状態によって知覚経験の内容が影響される場合」であると考えて議論を進めよう.この際、現象的性格も同様に変化しているかどうかはオープンにしておく.

ここで再び、冒頭の②で挙げた正当化の問題に戻ろう. 通常、「正当化されている(いない)」とか「合理的である(不合理である)」といったことは信念について言われることである. 知覚経験そのものについては、合理的だとか、信念と正当化関係に立つとは考えられてこなかった. しかし、認知的侵入(のようなもの)を認め、知覚経験そのものが歪められてしまうことがありうるとするならどうだろうか?知覚経験が信念形成に対して何か認識的な悪さをしている場合があるのではないだろうか?そのような場合、知覚経験も合理性という点で評価可能になる道も開けてくる. 例えば Siegel (2017)では、知覚経験が、先行する心理状態によって不合理なものとなってしまう

事例が挙げられている.

i. 怒りの思考実験. ジルはジャックが自分に対して怒っているのではないかと恐れている. そして実際にジャックに会った時, 彼女の恐れが, ジャックが怒っているように見える視覚経験を引き起こした. ここでジルは,「ジャックは自分に対して怒っている」という信念から出発し, それによって影響を受けた知覚経験を持ち, 最終的にその信念を強化するに至る. だが, 実際にはジャックは怒っていないのだとしたらどうだろうか. このケースでは先行する心理状態 (ジャックに対する恐れ) が, 知覚経験を不合理にしていると言えるだろう.

ii. 前成説. 精細胞の中に胚が含まれているという仮説を支持する人が顕微鏡を使って精細胞を見たら、細胞の中に胚があると報告した. 前成説についての信念が、それを支持するような視覚経験を引き起こしたのである.

iii.ペンチと銃.被験者に対象が提示され,それが銃であった場合には「gun」のボタンを押し、ペンチの場合には「pliers」を押す.対象を見る前に、白人か黒人の男性の写真が見せられる.白人よりも黒人を提示された場合の方がペンチを銃と誤認しやすい.「黒人は危険である」という偏見が、銃の経験を引き起こしたと考えられる.

シーゲルは上記のような経験を「乗っ取られた経験(hijacked experience)」と呼ぶが、彼女によれば、乗っ取られた経験を持つ主体がそれに基づいて信念を増強するのは不合理である。というのは、先行する心理状態によって悪影響を受けた乗っ取られ経験そのものが不合理だからである。繰り返すが、このことは「誤った信念から後続の誤った信念が形成されること」と類比的に考えることができるのである。シーゲルはこのように、経験が信念に対して与える正当化が弱められてしまうことを「認識的格下げ(epistemic downgrade)」と呼んでいる。

ここですぐに考えられる疑問が、「認知的侵入は**つねに**格下げを導くのか」というものであろうv. これについては否であると主張したい. ここで注意されたいのが、(1)で挙げた諸問題に加え、認知的侵入可能性の問題は知覚の許容内容をめぐる問題と密接に関わっているということだ. どういうことか.

知覚経験も内容を持つという志向説を採ることによって、「知覚内容にはどれだけのものが含まれうるのか」ということを問うことが可能となった. III で示唆されたように、知覚内容には、色や形といった低次の性質だけでなく、種性質や人格の同一性、他者の心的状態などの高次の性質も含まれうるのではないだろうか?ついで即座に生じてくるのが、我々が想定していた「認知的に侵入されたもの」は高次の性質ではないかという問題である. 高次性質

が知覚されうるならば(そして認知的侵入によって高次性質が知覚可能になる場合があるのだとすれば),認知的侵入は必ずしも格下げを導くのではないと主張できるだろう.「つまり,認知的侵入によって,知覚的に正当化されうる信念の種類が増えるのである」(源河 2017b:175).

(4) 今後の展望

今後の展望としては、三つが考えられる。まず認知的侵入の範囲について.前述のように、認知的侵入可能性を認めるかどうかは優れて経験的な側面を持つ. それゆえ、科学によって得られた実験データをその都度吟味していくという作業が哲学には求められるだろう。第二に、前節のように、認知的侵入可能性は高次性質の知覚可能性の問題とパラレルである。知覚による正当化を考えるときには、知覚の許容内容に関する問題も考えることが必要であるう。そして最後に、認知的侵入は主に視覚経験におけるものとして考えられているが、筆者は「他の感覚モダリティの場合についてはどういった議論が可能か」ということにも関心がある。

(慶應義塾大学)

i 以下の分類は Stokes (2013) を参照した.

ii ただし、情報遮断性の根拠としてこの錯視が本当に有力なのかについては議論の余地がある. というのは、未開の部族はミュラー=リヤー錯視に引っかからないという研究報告がいくつか挙げられているからである. このことは、むしろ我々が知覚学習によって獲得した何らかの能力によって錯視に引っかかっているという考えも可能であることを示唆しているのではないか. 例えば、Robins (2017)、Ch.2.1.を参照.

iii 現象的性格とは「その経験を持つことが主体にとってどのような感じか (what it is like)」という観点から特徴づけられる、意識に与えられる一人 称的にアクセス可能な何かである.一般に現象的性格と知覚内容は対応関係にあると考えられている.

iv もちろん、身体運動や外的な刺激によって(例えば薬を投与されて幻覚を見ている等)主体の認知状態が変化することで内容に変化が生じる場合は認知的侵入ではないといった程度にはコンセンサスが得られている.

v なお Siegel (2017) では、基準値以上の正当化が与えられる「認識的格上げ (epistemic upgrade)」についても検討されている.

(5) 参考文献

- 源河亨, 2017a,『知覚と判断の境界線:「知覚の哲学」基本と応用』,慶應義 塾大学出版会.
- ----, 2017b,「認知的侵入(不)可能性」,『ワードマップ心の哲学』所収, 信原幸弘編,新曜社.
- 戸田山和久,2004,「心は(どんな)コンピュータなのか」,『シリーズ心の哲学Ⅱロボット篇』所収,信原幸弘編,勁草書房.
- 西村正秀, 2017,「運動知覚の認知的侵入可能性」, 彦根論叢 412:52-67.
- Fodor, J. A, 1983, The Modularity of Mind, Cambridge, [『精神のモジュール形式--人工知能と心の哲学』伊藤笏康・信原幸弘訳,産業図書, 1985.]
- Hansen, T, M. Olkkonen, S. Walter, & K. R. Gegenfurtner, 2006, "Memory modulates color appearance", Nature Neuroscience 9:1367-8.
- Hanson, N. R, 1958, Patterns of Discovery, Cambridge. [『科学的発見のパターン』, 村上陽一郎訳, 講談社, 1986]
- Levin, D. &Banaji, M, 2006," Distortions in the Perceived Lightness of Faces: The Role of Race Categories", Journal of Experimental Psychology: General, 135:501-512.
- Machery, E, 2015," Cognitive Penetrability: A no-progress report", In Zeimbekis & Raftopoulos (2015).
- Macpherson, F, 2012," Cognitive penetration of colour experience", Philosophy and Phenomenological Research 84:24-62.
- Pylyshyn, Z, 1980, "Computation and Cognition: Isssues in the Foundations of Cognitive Science". Behavioral and Brain Sciences 3:111-32.
- Robins, P, 2017, "Modularity of Mind", Stanford Encyclopedia of Philosophy (https://plato.stanford.edu/entries/modularity-mind/).
- Siegel, S, 2006, "Which Properties are Represented in Perception?", Perceptual Experience. In T. Szabo Gendler & J. Hawthorne (eds.), Oxford, 481-503.
- ----, 2017, The Rationality of Perception, Oxford.
- Stokes, D, 2013, Cognitive penetrability of perception, Philosophy Compass 817:646-63.
- Zeimbekis, J. & Raftopoulos, A. (eds.), 2015, The Cognitive Penetrability of Perception: New Philosophical Perspectives, Oxford.

真理の表現的機能と反映原理について

林 大智

Abstract

Deflationism about truth stresses the two aspects of truth. First, truth has no substantial properties. Second, it is a useful tool for expressing, e.g. general assertions. Mainly because of the Gödel phenomena, however, there is some problem with these requirements in formal theories of truth.

Following Horsten and Leigh (2017), this paper claims that this difficulty can be partially overcome at least for the typed theory of truth with the help of iterated reflection principles. Furthermore, we show that it is possible to add restricted reflection principles to the Friedman-Sheard theory without paradox, and thus the theory partially satisfies the deflationists' requirements, too.

研究テーマ

形式的な真理の理論における、デフレ主義と真理の表現的有用性の関係

1 真理のデフレ主義と一般化の機能

真理のデフレ主義者は、おおまかには、真理は実質的な(あるいは重要な)概念ではないものの、表現上の有用性をもつ、ということを主張する。特に、真理の有用性の一つとして「一般化」と呼ばれる機能が広く認められてきた1。例えば、ある人が「今日は晴れているか、晴れていないかのどちらかだ」などといった、具体的な排中律のあらゆる例を全て認めているとする。このような例は無数に多く考えられるため、その人の態度は、それぞれを無数の連言でつないだ形で表す必要がある。だが、代わりに真理述語を用いれば、たった一文、「あらゆる排中律の例は真だ」によって表現できることになる。

形式的な真理理論においてこの表現的機能を実現しようとする場合、一般にその理論の演繹力は強くなる。だが他方で、真理の非実質性を保証するには、その理論が弱いほうが望ましい。こうしてデフレ主義者は、少なくとも自らのテーゼに形式的な基盤を与えようとする際にジレンマに陥ってしまう。本稿ではこの問題に焦点を当て、どのように軽減されうるかを考察する。

形式的な真理理論において主に対象とされるのは、一階述語論理の算術用の言語 \mathcal{L} と、その再帰的に公理化された算術の理論 \mathcal{B} である。以下では簡単のため、 \mathcal{L} の論理的記号は「と \land と \lor のみとし、その他は自然に定義する。また、 \mathcal{L} は定項記号 $\overline{\mathbf{0}}$ と後続者関数 \mathbf{S} 、足し算、掛け算用の二項関数記号+、

 \times のみ持ち、関係記号は等号=のみとする。各 $k<\omega$ について、 \mathcal{L} の項である 数項 \bar{k} は、 $\bar{0}$ に S を k 回適用した $S(S(...S(\bar{0})...))$ の略とする。B は通常の統語 論を展開できる必要があり、ここでは B=PA(ペアノ算術)とする。表現のコ ーディングを固定し、表現 E に対して、「E⁷はそのゲーデル数とする(誤解 の恐れがない場合、このカッコは適宜省略する)。また、例えば「x は £ の閉 項である/文である」といった述語は、それぞれ \mathcal{L} の式 $\operatorname{Term}_{\mathcal{L}}(x)/\operatorname{Sent}_{\mathcal{L}}(x)$ で 定義できる。また、Provs(x)は再帰的に公理化された理論 S の標準的な可証 性述語とする。 \mathcal{L} の評価関数 「 $^{\circ}$ 」は、 \mathcal{L} の項(のゲーデル数)に対して、 その値を返し、これも PAで定義できる(値の存在と一意性が PAで示せる)。 また、他の再帰関数についても同様である。例えば £ の変数 x (のゲーデル 数)と L の式 φ(のゲーデル数)に対して∀xφ(のゲーデル数)を返す関数 は∀で表される。一般に、関数gがPAで定義できるとき、gはこれを表す、 定義された関数記号とする。以下で考える他の言語 \mathcal{L}' についても、 $Sent_{\mathcal{L}'}(x)$ などと、同様に £式で定義できる 2。また、ここで PA は我々が既に受け入れ ているものとしての役割も持っており、ベース理論と呼ばれる。∠に新しい 一項述語記号 T を加えてできる言語を \mathcal{L}_T とする。今、PA に \mathcal{L}_T の公理や規 則を加えて新たな LTの(真理) 理論 Sを作る(以下、考察する真理理論は全 て、ベース理論 PA に真理理論の言語における公理や規則を加えて得られる とし、PAへの言及は省略する)。そのようなSの例としては、タルスキに由 来する次が有名である。

定義 2 \mathcal{L}_T 理論 the disquotational theory, DT は、 $\{T(\phi) \leftarrow \rightarrow \phi \mid \phi$ は \mathcal{L} 文 $\}$ 。 だが、DT は上に述べた一般化の機能の点で不満足である。例えば、 \mathcal{L} の排中律の例は全て真である」(を形式化した \mathcal{L}_T 文) でさえ、DT からでは示せない(Halbach, 1999, p.10)。より強い真理理論として、次も有名である。

定義 3 L_T 理論 the compositional theory, CT は、次の公理からなる。

CT1 $\forall s, t(\operatorname{Term}_{\mathcal{L}}(s) \land \operatorname{Term}_{\mathcal{L}}(t) \rightarrow (\operatorname{T}(s = t) \leftarrow \rightarrow s^{\circ} = t^{\circ}))$

CT2 $\forall x (Sent_{\mathcal{L}}(x) \rightarrow (T(\neg x) \leftarrow \rightarrow \neg T(x)))$

CT3 $\forall x,y(Sent_{\mathcal{L}}(x \land y) \rightarrow (T(x \land y) \leftarrow \rightarrow T(x) \land T(y)))$

 $CT4 \quad \forall \ v, x(Sent_{\mathcal{L}}(\forall \ vx) \rightarrow (T(\forall \ vx) \leftarrow \rightarrow \forall \ t(Term_{\mathcal{L}}(t) \rightarrow T(x(t/v)))))$

(CT4 の x(t/v)は、式 x の各自由変数 v を一斉に項 t で置き換えた結果)

CT は DT を含み、「L の排中律の例はすべて真である」といった文も示すことができ、さらに PA と同じ L 文を示す。つまり、CT は PA 上保存拡大であり、この性質が真理の非実質性の基準として重要だとする議論もある 3 。

だが、度々問題となるのは、証明と真理の関係である。我々が受け入れている理論 U から示された、 \mathcal{L} のどの文についても、(デフレ主義によれば) 真

と言えるはずである。よって、真理の一般化の機能からすれば、「Uから示された \mathcal{L} 文は全て真」という主張を示せるべきだろう。これは形式的に $\forall x(Sent_{\mathcal{L}}(x) \to (Prov_{\mathcal{U}}(x) \to T(x)))$ として表せ、Uの \mathcal{L} 文用の大域的反映原理, $(GRP_{\mathcal{U}})$ という(以下、特に理論 Uの中身に関心がないときは、単に $(GRP_{\mathcal{U}})$ と呼ぶ。)。ベース理論 PAの場合で言えば、PAから示された (\mathcal{L}) 文は全て真といえるはずなので、 $(GRP_{\mathcal{U}})$: $\forall x(Sent_{\mathcal{L}}(x) \to (Prov_{\mathcal{U}}(x) \to T(x)))$ で表現される。しかしながら、実際には、理論 $(CT + GRP_{\mathcal{U}})$ は(この+は、理論に公理や式の集合を加えて帰結で閉じて新たな理論を作る操作を表す)、PAから示せない (\mathcal{L}) 文を導くため、(CT) の保存拡大性から、 $(GRP_{\mathcal{U}})$ は (CT) で示せない。

こうした問題のため、非実質性か、表現的機能のどちらかの完全な実現は妥協する必要がある。そこで、本稿では保存拡大性を破棄したとして、GRPをどのように正当に獲得できるか考えることにする。まず、単純に CT+GRPPAを PA の真理理論とみなすことは不十分に思われる。というのも、この新しい理論において新たな L 文が示せたのならば、今度はそれらも含めた L 文全てに対する GRP、すなわち、GRPCT+GRPPAも導出できるべきだという要求が再び生じるからである。こうして要求を次々に満たすには、結局 GRPの追加を無限に繰り返す必要がある。このことの問題点は、無限回の追加というよりも、望ましい主張を無根拠に前提することでしか得ていないというアドホックさにあるだろう。そこで、真理への直観や理解を前提としない原理に基づいて理論を強くし、そこから GRPを示していくことが求められる。そうした原理の追加は依然として無限回必要になるかもしれないが、GRPの単なる追加のような自明さはないといえる。

以上の動機のもと、本稿では次の結果を得た。まず、GRPを非自明に得るという方針が、CTの場合に成功することを定理 1 で確認する。次に、型無しの真理理論のうち、代表的な理論の一つである FS 理論に焦点を当て、この方針を制限することで、理論の無矛盾性を維持したまま、GRPに関して定理 1 と類似の結果が得られることを、定理 2 と定理 3 で確認する。

2 反映原理の反復適用(型付きの真理理論の場合)

GRP に替わる原理の提示には、Horsten & Leigh (2017)の議論を採用する。 それによると、私たちは、すでに受け入れている理論 S においてある式が証明されたならば、それもまた受け入れるはずである。この態度は、S の統の的反映原理(uniform reflection principles)、 $RPs:\{\forall x(Provs(\phi(\dot{x}))\rightarrow\phi(x))\mid \phi(x)\}$ は S の言語の、高々x のみ自由変数にもつ式yという集合の形で表すことができるだろう(ここで $\phi(\dot{x})$ は、 $\phi(x)$ にx に等しい数項を代入した式のゲーデル 数を返す関数。また、具体的なSに関心がないときは、RPsを単にRPと呼ぶ)。すると、Sを受け入れることのうちには、暗黙的にこのRPsを受け入れることが含まれる。この暗黙的な受け入れは、真理概念を持たずとも定式化できる、一般的な態度である。よって、単なるGRPの追加よりも好ましいA。本稿はこの方針をGRPの獲得に利用する。

定義 4 理論 S に対して、理論 $RP_n[S]$ は次のように帰納的に定義される。まず $RP_0[S] = S$ 。次に、 $RP_{n+1}[S] = RP_n[S] + RP_{RPn[S]}$ とする。そして、 $RP_{<\omega}[S]$ は理論 $\cup_{n<\omega}RP_n[S]$ とする。

すると、CTについては上の要求が満たせることがわかる:

定理1 各nに対して、RPn+2[CT]はGRPRPn[CT]を示す。

証明 Horsten & Leigh (2017, pp.223-4)と同様に示せる。まず、各 \mathcal{L} 文 φ に対して、 RP_{n+1} [CT] $\vdash Prov_{\mathbf{RPn[CT]}}(\varphi) \rightarrow \varphi$ であり、CT は DT を含むので、 $\varphi \rightarrow T(\varphi)$ が示せる。よって、 RP_{n+1} [CT] $\vdash Prov_{\mathbf{RPn[CT]}}(\varphi) \rightarrow T(\varphi)$ で、さらにこれは PA 内でも示せ、 $PA \vdash \forall x (Sent_{\mathcal{L}}(x) \rightarrow Prov_{\mathbf{RPn+1}} \cap (Prov_{\mathbf{RPn[CT]}}(\dot{x}) \rightarrow T(\dot{x}))$ つ)。 ゆえに、 RP_{n+2} [CT] における RP の例を用いて $GRP_{\mathbf{RPn[CT]}}$ を得る。

よって、型付きの真理理論の場合、RP の利用という非自明な仕方で、各 GRP、つまり「CT から示された \mathcal{L} 文は全て真」「CT+RPcT から示された \mathcal{L} 文は全て真」、・・・等々が RP $_{<\omega}$ [CT]から示せることがわかる 5 。

3 型無しの真理理論 (FS) と反映原理の制限

次に型無しの真理理論、つまり \mathcal{L}_T 文への T の適用も許す理論を考える。ここでは特に the Friedman-Sheard theory, FS を考察する。特に、Halbach (2011) による定式化は次:

定義 5 \mathcal{L}_T 理論 FSN は、定義 3 の CT1 と公理 FS2~FS4 と、 \mathcal{L}_T 式用の全ての帰納法の例からなる。ただし、各 FSi(2 \leq i \leq 4)は、定義 3 の CTi の Sent $_{\mathcal{L}_T}$ に置き換えたもの。

次に、NEC とは \mathcal{L}_T 規則で、任意の \mathcal{L}_T 文 φ から $T(\varphi)$ の導出を許す。CONEC は逆に、 $T(\varphi)$ から φ の導出を許す規則である、そして、FS とは、FSN を NEC/CONEC で閉じて得られる理論とする。また、NEC/CONEC の適用をそれぞれ n 回未満に制限することで得られる、FS の部分理論を FS_n という。ここで FS_0 は $PA+\mathcal{L}_T$ 式の帰納法全てとする。

ここで FS に対する自然な要求は、CT と同様に FS に RP を追加していくことで、A GRP が示せることである。だが、BC GRP の定理により、FS は GRP GRP

そのため、少なくとも∠文に関して、矛盾を導くことなく上のような一般

 \mathcal{L} にない、新しい真理述語 T_n $(n<\omega)$ をとり、 $\mathcal{L}_{< k}$ とは、 \mathcal{L} に、各 $n< k \leq \omega$ について、 T_n を加えてできる言語とする。特に $\mathcal{L}_{< 0}$ は \mathcal{L} と一致する。

定義 6 $\mathcal{L}_{< n}$ 理論である $RT_{< n}$ とは、各 i < k < n なる i と k について、次の公理、 $RT1_n \sim RT7_n$ と、 $\mathcal{L}_{< n}$ 式用の全ての帰納法の例からなる: RTj_n ($1 \le j \le 4$)は、それぞれ定義 2 の CTj において、 $Sent_{\mathcal{L}}$ を $Sent_{\mathcal{L}< k}$ に、T を T_k に置き換えたものとする。また、 $RT5_n \sim RT7_n$ はそれぞれ次:

 $RT5_n \forall x (Sent_{\mathcal{L} < i}(x) \rightarrow (T_k(\underline{T}_i(\dot{x})) \leftarrow \rightarrow T_i(x)))$

 $RT6_n \forall x (Sent_{\mathcal{L} < \mathbf{i}}(x) \rightarrow (T_k(\underline{T}_i(\dot{x})) \leftarrow \rightarrow T_k(x)))$

 $RT7_n \forall x(T_k(x) \rightarrow Sent_{\mathcal{L} < k}(x))$

そして、 $RT_{<\omega}$ は理論 $\cup_{n<\omega}RT_n$ とする。

Halbach は各 $RT_{< n}$ に対し、その各公理を、FSの定理へと変換する関数を定めた。特に本稿では、 $\mathcal{L}_{< 1}$ 式から \mathcal{L}_T 式への再帰関数 h を用いる。これは、各 $\mathcal{L}_{< 1}$ 式 ϕ における $T_0(t)$ の出現を、それぞれ $Sent_{\mathcal{L}_{< 0}}(t) \wedge T(t)$ に置き換えて、 \mathcal{L}_T 式を出力する関数である(Halbach のものより若干簡略化したが、本稿には影響しない)。すると、各 \mathcal{L} 式 ϕ で $h(\phi) = \phi$ 。(Halbach,1994 p.322) 補題 各 $\mathcal{L}_{< 1}$ χ ϕ に対し、 $RT_{< \omega}$ $\vdash \phi$ と FS $\vdash h(\phi)$ は互いに同値。

証明 左から右は Halbach(1994, p.322)による。本稿の $RT_{< n}$ の場合、定義 6 で $RT7_n$ を新たに加えたが、Halbach の証明はそのまま実行できる。

次に、右から左については、Halbach が L 文の場合のみ示したが、その証明を利用して、 $L_{<1}$ 文用の証明を得たい。まず、Halbach(1994, pp.323-4)の結果は次のようである。 L_T 式から $L_{< n}$ 式への再帰的関数 g_n を、n に関して帰納的に作る。大雑把には、各 g_n は、 L_T 式に対し、そのうちで他の真理述語に含まれる回数が少ない真理述語から順に、大きい指標をつけていき、型の大小を守らせるようにして、 $L_{< n}$ 式に変換する関数である、 g_0 は、各 L_T 式に対して、そのなかの全ての Tt の形の出現を 0=1 に置き換えた式を出力する

関数とする(「0=1」の代わりに、適当な算術文としてもよい)。 g_{n+1} は次:各 \mathcal{L}_T 式 φ に対し、

$$g_{n+1}(\phi) = \begin{cases} \phi & \mathcal{L} \ \textit{のあるターム} \ \textit{s,t} \ \textit{について、} \phi \ \textit{が} \ \textit{s=t} \ \textit{の形のとき} \\ \hline 0 = \overline{1} & \phi \ \textit{が} \ \mathcal{L}_T \ \textit{の式ではないとき} \\ T_n(g_n(t)) & \phi \ \textit{が} \ \textit{Tt} \ \textit{の形のとき} \end{cases}$$

 φ が複合式の形のときは g_{n+1} と一、 \wedge 、 \forall をそれぞれ交換する。

そして、次が成立する(Halbach, 1994 p.323) : 各 n と \mathcal{L}_T の式 φ について、 $FS_n \vdash \varphi$ ならば、 $RT_{<2n} \vdash g_n(\varphi)$

これを用いて補題の右から左を示す。そのために、 $\mathcal{L}_{<1}$ 文 φ について、FS ├h(φ)であると仮定する。すると、ある n で FSn ├h(φ)で、特に n≧1とでき る。よって、Halbach の結果から、 $RT_{<2n} \vdash g_n(h(\varphi))$ を得る。 φ は $\mathcal{L}_{<1}$ 文なの で、真理述語は $T_0(t)$ の形でのみ出現しうる。よって、 $h(\varphi)$ において、この真 理述語は、 $Sent_{\mathcal{L}<0}(t) \wedge T(t)$ に置き換わっている。さらにこれは、 g_n の適用に よって、 $Sent_{\mathcal{L}<0}(t) \wedge T_{n-1}(g_{n-1}(t))$ に置き換わる。今、 $g_n(h(\varphi))$ におけるこれ らの出現を、それぞれ $T_0(t)$ に置き換えたものを φ 'とおく。 φ 'が $g_n(h(\varphi))$ と RT < 2n で同値であること・・・①、φ'がφ自身と一致すること・・・②、を示す。 そうすれば、 $RT_{<2n} \vdash \varphi$ 、つまり、 $RT_{<\alpha} \vdash \varphi$ が示されるので、右から左が示さ れたことになる。まず①をφの複雑さの帰納法で示す。複合式の場合と算術 的原子式の場合はそれぞれ帰納法の仮定と翻訳関数の定義より明らかなので、 原子式 $T_0(t)$ の場合に $T_0(t)$ と $Sent_{\mathcal{L}<0}(t) \wedge T_{n-1}(g_{n-1}(t))$ が $RT_{<2n}$ で同値であ ることのみ示す。まず、 $Sent_{\mathcal{L}<0}(t)$ ならば $t=g_{n-1}(t)$ が PA で示せ、 $Sent_{\mathcal{L}<0}(t)$ $\wedge T_{n-1}(g_{n-1}(t))$ は、 $\mathrm{Sent}_{\mathcal{L}<0}(t)\wedge T_{n-1}(t)$ に同値。さらに、k=n-1と i=0の場 合で定義 6 の RT5n,6n を用いて Sent_{ℓ<0}(t)→(T_{n-1}(t)←→T₀(t))が示せるの で、定義 6 の RT7n より、①が従う。② も φ の複雑さの帰納法で示せる。 □ 次に、FSとRT<ωそれぞれに対し、制限された反映原理を考える。

定義 7 FS に対し、制限された(restricted)反映階層 $rRP_n[FS]$ は n に関して帰納的に定義される。 $rRP_0[FS]$ は FS のままとする、n>0 では次:

 $rRP_{n+1}[FS] = rRP_n[FS] + \{ \forall \ x (Prov_{rRPn}[FS](\phi(\dot{x})) \rightarrow \phi(x)) \mid \forall \ x \phi(x)$ は、ある $\mathcal{L}_{<1}$ 文に h を適用して得た \mathcal{L}_T 文 $\}$. そして、 $rRP_{<\omega}[S]$ は $\cup_{n<\omega} rRP_n[S]$ とする。同様に、 $RT_{<\omega}$ 用の制限された反映階層 $rRP_n[RT_{<\omega}]$ については、 $rRP_0[RT_{<\omega}]$ は $RT_{<\omega}$ のまま。次に、 $rRP_{n+1}[RT_{<\omega}] = rRP_n[RT_{<\omega}] + \{ \forall \ x (Prov_{rRPn}[RT_{<\omega}]) + \{ \forall \ x (Prov_{rRPn}[RT_{<\omega}]$

そして、 $rRP_{<\omega}[RT_{<\omega}]$ は $\cup_{n<\omega}rRP_n[RT_{<\omega}]$ とする。

定理 2 各 $n < \omega$ と $\mathcal{L}_{<1}$ 文 ϕ について、 $rRP_n[RT_{<\omega}]$ $\vdash \phi$ と $rRP_n[FS]$ $\vdash h(\phi)$ は互いに同値である。さらに、このことは PA 内で証明できる。

証明 nの帰納法で示す。n=0の場合は補題による。

定理 3 各 n で、rRP_{n+2}[FS] ト GRP_{rRPn[FS]}である。

よって、各 $h(\varphi j)(1 \leq j \leq m)$ は、 $\forall x(Prov_{\mathbf{RPi}[\mathbf{RT} < \varphi]}(\varphi j'(\dot{\mathbf{x}})) \rightarrow h(\varphi j'(\mathbf{x})))$ の形なので、これは $\forall x(Prov_{\mathbf{RPi}[\mathbf{FS}]}(\underline{h}(\varphi j'(\dot{\mathbf{x}}))) \rightarrow h(\varphi j'(\mathbf{x})))$ と同値で、後者は $\mathbf{rRP}_{i+1}[\mathbf{FS}]$ から示せる。ゆえに、三段論法から $\mathbf{rRP}_{i+1}[\mathbf{FS}] \vdash h(\varphi)$ で、左から右側が示された。反対側は、今の議論をほぼ下から上にさかのぼることで同様に示せる。さらに以上は PA で示せる。よって、 $\mathbf{n}=i+1$ の場合も良い。

定理 2 より、特に \mathcal{L} 文について両者は同じ文を示す。 $rRP_{<\omega}[RT_{<\omega}]$ の無矛盾性は、例えば Leigh(2016)の KF への分析からわかる。よって、 $rRP_{<\omega}[FS]$ も無矛盾とわかる。また、定理 1 と同様に、次が成り立つことがわかる:

証明 定理 1 と同様に示すには ϕ : \forall $x(Sent_{\mathcal{L}<\mathbf{0}}(x)\rightarrow (Prov_{\mathbf{rRPn}[\mathbf{FS}]}(x)\rightarrow T(x)))$ に 同値な \mathcal{L}_T $\dot{\chi}$ ϕ 'であって、ある $\mathcal{L}_{<1}$ $\dot{\chi}$ ψ について $h(\psi)$ の形になっているもの を与えればよく、 ϕ ': \forall $x(Sent_{\mathcal{L}<\mathbf{0}}(x)\rightarrow (Prov_{\mathbf{rRPn}[\mathbf{FS}]}(x)\rightarrow (Sent_{\mathcal{L}<\mathbf{0}}(x)\wedge T(x))))$ が、 $h(\forall$ $x(Sent_{\mathcal{L}<\mathbf{0}}(x)\rightarrow (Prov_{\mathbf{rRPn}[\mathbf{FS}]}(x)\rightarrow T_0(x))))$ の形で、さらに ϕ と ϕ 'は論

よって、無矛盾な $\mathbf{rRP}_{<\omega}[\mathbf{RT}_{<\omega}]$ から、「 \mathbf{FS} で示せた $\mathbf{\mathcal{L}}$ 文はすべて真」、「 \mathbf{FS} に一回 \mathbf{rRP} を加えた理論で示せた $\mathbf{\mathcal{L}}$ 文はすべて真」等々が示せるとわかる。

結論と今後の課題

理的に同値なので良い。

本稿では、統一的反映原理の反復的な追加が、大域的反映原理の導出に有効であると主張し、さらに、型無しの真理理論で、特に ω 矛盾である FS においても、反映原理の制限によって、L文用の GRP が無矛盾のまま導出できることを示した。だが、型無しの利点を保持するには、L_T 文を含む GRP についても同様の結果が得られるべきだと思われる。特に、L_{< ω} 式一般に関して上の補題に近い結果が得られるのかについては現在考察中である。

型無しの真理理論としては、他にも the Kripke-Feferman theory, KF が有名である。これは ω 無矛盾なため、FS の問題は生じないが、定理 1 と同様の仕方で \mathcal{L}_T 文用の GRP を得られるかは定かでない。そこで、定理 1 の証

明を模倣するには、CT における L 文の集合のように、KF で既に双条件文が示せている集合に限定して GRP を示していくことが考えられる。そのための RP の追加によって KF がどれほど強くなるかなどは今後の課題とする 6。

注

- 1. デフレ主義の歴史的展開や、真理に関する一般化の機能の哲学的問題については、例えば Båve(2006)で広く考察されている。
- 2. PA 内での統語論の詳細は、例えば Kaye(1997, Ch. 9.2)を見よ。
- 3. デフレ主義と保存拡大性の関係については藤本(2014)を見られたい。
- 4. 査読者の指摘通り、もし GRP もまた暗黙的な受け入れの態度の定式化になっているといえるならば、RP から GRP を得るという本論の議論は冗長となる。一点補足すると、本論では「理論の帰結を暗黙的に受け入れる」態度が真理概念の把握に先立って可能である、という直観に依拠している。これが正しければ、その態度の定式化は真理概念を用いずになされるべきであり、GRP のような定式化は冗長となるはずである。少なくともこの直観への評価が定まらない限り、RP から GRP を得ることには意義があると思われる。
- 5. 査読者の指摘通り、以上は $U \vdash GRP_U$ なる理論 U を求める代わりに、 GRP_U の要求が生じるたびに、それが示せる U の適切な拡大を与えるものである。
- 6. 本稿の作成にあたり、佐野勝彦氏と匿名の査読者からは、全体の構成や多くの技術的誤りについて適切な助言を頂いた。ここに感謝申し上げる。

参考文献

- Båve, A. 2006, Deflationism: A Use-Theoretic Analysis of the Truth-Predicate, Dissertation, Stockholm University.
- Halbach, V. 1994, "A System of Complete and Consistent Truth," *Notre Dame Journal of Formal Logic*, 35(1):311-27.
- Halbach, V. 1999, "Disquotationalism and Infinite Conjunctions," *Mind*, 108(429):1-22.
- Halbach, V. 2011, Axiomatic theories of truth, Cambridge University Press.
- Horsten, L. & Leigh, G. E. 2017, "Truth is simple," Mind, 126(501):195-232.
- Kaye, R. 1997, Models of Peano Arithmetic, Oxford University Press.
- Leigh, G. E. 2016, "Reflecting on truth," If CoLog Journal of Logics and their Applications, 3(4):557-94.
- 藤本健太郎 2014 「デフレ主義と保存性」『哲学雑誌』 129(801):158-31 有斐閣 (北海道大学)

セラーズが言う自己についての「報告」とはどのようなものか? What kind of "report" about self argued by Sellers?

過能洋平

Abstract

In this paper, I examine how privacy and first person authority of self-confidence held by subjects are established. I do it by focusing on the "reporting" use of a language found in "report" that Sellars described in "Empiricism and the Philosophy of Mind". I think that the self-attribution of reported contents plays an important role because in the arguments presented by Sellars, "report" means the report on the mental states of self. When we "report" the sensory impression, we take the sensory impression as attributable to the self. At the same time, "reporting" the sensory impression expresses the self-knowledge about the sensory impression. Based on this observation, first, I review the Sellars's argument on how "reporting" use of a language is established. Then, I analyze the characteristic of the self-attributions. Finally, I consider the process of how privacy and first person authority are established from the self-attribution.

(1) 研究テーマ

セラーズの類比説を分析し擁護する

Analyze and defend Sellars's analogy theory

(2) 研究の背景・先行研究

・背景:自己知の持つ性質として一般に、不可謬性と自己告知性、直接性がある。また主体がそれに対して私秘性や一人称権威を持つと考えられるだろうi。とはいえ、自己知がこうした性質を持つということ、主体がそれらに対して私秘性や一人称権威を持つことは前提されており、これらのことがいかにして成立するかについてはあまり論じられていないように思える。筆者が疑問視しているのは自己知や、それに対して主体が持つとされるこうした性質は何に由来するのか、またどのようにして成立するのか、ということである。これらを検討するにあたり、アメリカ分析哲学者のWirflid Sellars が"Empiricism and Philosophy of Mind"(以下 EPM と略記する)のなかで論じた自己についての「報告」という言語使用を考察することが解明につなが

ると思われる。

• 先行研究

セラーズ研究の基礎文献の一つは Knowledge, Mind, and the Given である。著者である DeVries と Triplett は感覚印象に関する言説について次のように述べている。「印象言明について報告する役割が適用されると、私たちが内観を通じて知識を獲得しうると、私たちは自分たちの印象を報告しうると、私たちがそれらに対する特権的アクセスを持つと言われうると、そしてこれらの印象の出来事はそうした出来事を持つ個人にとって私秘的であると、人が伝統と調和して述べることのできる意味がある」…と述べている。ただ、印象言明が報告という役割を獲得するとどうしてそうなるのか、明確には述べていない。印象言明に報告するという役割が適用されることでなぜ私たちが内的出来事に対する一人称権威を持つと、述べることができるのだろうか?

また同じく基礎文献である Wilfrid Sellars において deVries はこの報告を非推論的なものと特徴付けているiii。この非推論的という特徴をブランダムは条件反射として特徴付けている。「すなわち、以前は推論されることだけが可能であったものを非推論的に報告する条件反射(多分、彼の思考に関係するある究極的には発見可能な神経生理学的出来事に依存しているもの)を訓練によって他の人のうちに発達させることができるのである」iv。

非推論的報告に関して、DeVries と Triplettvは次のような説明を与えている。未熟な野鳥観察者は最初、見つけた鳥をその大きさ、形、色といった手がかりをもとに図鑑などで調べることで、どんな鳥か確定する。見つけた鳥の特徴と図鑑に記載されている特徴とを比較することでその鳥の種類を確定する。その野鳥観察者は観察の経験を積むことで図鑑を調べずとも、次第にパッとみただけでその鳥を確定することができるようになる。

感覚印象を条件反射的に報告するようになるとしても、やはりこの「報告する」という役割を持つ言語使用の習得により、どうして感覚印象の私秘性やそれに対する一人称的権威が成立するのかあまり明瞭ではないように思える。

・セラーズの議論の概略:ここではセラーズがいう「報告」という言語使 用が成立するまでの議論を整理する。

セラーズは EPM の第 12~16 部において、自身が「神話」vi、と呼ぶ物語 (いわゆる「天才ジョーンズの神話」)を通じて、内的出来事に関する積極的 な主張を展開している。その神話の中にライル人たちと呼ばれる人々が登場する。彼らは、心や内的、思考といった概念を持っていない。そのため彼ら

は物理的対象やその性質について指示する事はできるが、内的出来事や直接 経験について語る事はまだできない。セラーズはそれらについて語ることが できるようになるために、彼らの言語使用や概念を追加拡張していくことを 通じて、自身の見解を示している。

ライル人たちは常に声に出して考えながら知的な振る舞いをしていた。ある時彼らが声に出して考えずとも知的な振る舞いをすることにジョーンズは気がつく。そこで彼は、普段声に出して考えているときに発話するのと同等の「内言」が行為者のうちに生起し、それから始まる内的過程の到達点として観察可能な言語行動が生じる、という理論を考えた。これを使って、ある主体 A の振る舞いを見ている別の主体 B は「A が「p」と考えている」という文を用いて彼の振る舞いを記述する。この「p」が理論的に措定される「内言」にあたるものである。次に B が同様の振る舞いをするとき同じく「私は「p」と考えている」という文を用いて自身の心的状態を報告する。同時にB はこの文を自己帰属することで自分の心的状態を理解していることになる。そして B によるこうした文の使用を行動による証拠が強く支持するときに、B は他者たちから肯定され支持しないときには否定されることを通じて「合理的に信頼できる自己の記述」をするようになる。この心的状態に関する自己記述が「報告」という役割を果たしているのである。

次にジョーンズは感覚印象の理論を考える。ライル人たちがそれまで持っていたのは思考の枠組みに属する知覚の概念である。これは端的に言えば、「赤い三角であるものを見ている」や「テーブルが茶色であることを見たり、ピアノが調子外れであることを聞いたりする」viiという言語使用で表されている、「であるのを見る」の言語使用である。これは観察可能な物理的対象についての言説である。しかしライル人たちは「印象や、感覚や、感じなどのようなもの、約言すれば、哲学者が「直接経験」の見出しでまとめてしまうような事象へ」viii言及できない。つまり、ライル人たちは物理的対象について語ることはできても、それの自分への現れといったものについて語ることができない。というのも彼らはまだ現れという概念を持っていないからである。

このような状態のライル人たちに見間違いや錯覚、幻覚が生じたと考えてみよう。ジョーンズはそれらついて語れるようになるために感覚印象の理論を提案する。彼は感覚印象を内的理論的出来事--知覚主体の状態として想定されたもの--として導入する。ixジョーンズは感覚印象について内的レプリカの領域という考えをモデルとしている。このモデルはレプリカが知覚者の中に生起することであって、レプリカをさらに知覚することではない。主体

の内に生起すると措定されるレプリカのモデルとなるものは固有の性質を持つ物理的存在である。赤い三角という性質はそのように記述される物理的存在者だけが持つ。赤い三角形の印象自体が赤い三角だとは限らない。こうしたモデルと措定されるレプリカの関係は「赤い三角の印象は赤い三角形をしたウェハースに、その度合いをきちんとした形で特定することはできないが、類比的であるということである」×。この類比の本質的特徴は「視覚印象は、可視的対象の色や形が互いに類似したり異なったりする仕方と構造的な類似を有する体系の中で、互いに類似したり異なったりする関係に立っている」xiことである。

そしてジョーンズは仲間たちにこのレプリカについて語ることを訓練する。彼は、「同胞が赤い三角形の印象を実際にもっているとき、そしてそのときに限り、「私は赤い三角形の印象をもっている」と言うように訓練するのである」xii。例えばある主体 A が「赤い三角形であるものを見ている」と述べるなら、それは A にとっての対象の現れの報告ではなく、物理的対象についての客観的な記述である。一方で A が「赤い三角形の印象をもっている」と述べる場合、これは対象について記述しているのではなく、その対象の A にとっての現れについて報告している。A のこの発話を聞いている別の主体 B は A の内的状態に「赤い三角形の印象が生起している」と措定し、A に帰属させる。次第に A はこの発話をするとともに自身のうちに「赤い三角の印象」が生起していると措定し、自己帰属するようになる。

(3) 筆者の主張

私の考えでは、「報告」という言語使用には自己帰属が深く関与している。帰属というのは通常、x は y に帰属する、x は y に帰属される、という使用を持つ。性質や属性、機能がある物理的対象に帰属される。赤さがこの対象に帰属されると述べる場合、この対象が赤さを有していると述べることに等しい。帰属には 3 つの要素がある。①性質や機能を持つ存在、②性質や機能自体、③①と②の関係を理解するもの。通常の帰属は①、②が一致しており③がそれらとは独立にある。ある対象(①)に赤い三角という性質(②)をある人が帰属する(③)という形になっている。自己帰属はこれら3つが同一である場合である。

発話から内言や感覚印象を帰属する前段階として、より単純素朴に物理的対象に性質や属性を帰属する段階がある。ある対象に大きさ、硬さや色といった性質を帰属させる。対象 O に性質 x を帰属させることは、O が x を持つ

と S が理解することである。自己帰属はある性質や機能の帰属先を自分にすることで可能となると思われる。

「報告」がジョーンズの神話の中で使われるのは自分の「内言」の発話訓練の時である。ある主体 A の「p」という発話が振る舞いによる証拠と合致している際、別の主体(A の発話の聞き手)B が肯定することとともに A はp という内言が自分に生じていたと措定し自己帰属する。また振る舞いによる証拠と「p」という発話が合致していない場合、戸惑いと訂正が発話者に返される。自己記述することは記述内容を自分に帰属することと言える。

主体 A が発話せずに振舞っている時主体 B はその振る舞いのみから A の内言を類推するしかない。実際にどのような内言(=思考)があるのか類推ではなく報告できるのは A のみである。そしてこのような状況になった時、A の「報告」には独特な権威が生じる。他の事柄、例えば物理的対象についてであれば、標準的状況における標準的観察者なら、誰でもその対象について記述し、どんなものであるか他者に等しく報告できる。しかし内言の場合、それを報告できるのはその内言を持つ当の主体のみである。

では感覚印象の「報告」の場合はどうか?内言の場合と同じ構造で考える。セラーズによると内言を報告する訓練は「赤い三角形の印象を実際に持っているとき、そしてそのときに限り、「私は赤い三角形の印象を持っている」と言うように訓練するのである」。主体が「赤い三角形の印象を」持つのは実際に赤い三角である対象を見ている時である。その時に「赤い三角の印象を持つ」と述べることを肯定する。そうでない時に、この発話を否定することで自己記述するように訓練する。

ここまでは内言の場合と同じだが、実際にあるものと生起している感覚印象が異なる場合はどうだろうか?青い球体であるものを見ているときに「私は今赤い三角の印象を持つ」と述べるならば戸惑いが返されるが、強い訂正はなされないだろう。発話者は自分の感覚印象に関する発話に対する応答から、自分の感覚印象が少なくとも他の人の感覚印象とはズレているということを理解して、レプリカが正常に成立するように試みたり、あるいは「報告」を変更したりするだろう。

現在自分にどのような感覚印象が成立しているかを「報告」できるのはその人だけである。物理的対象に関する報告は誰でもできるが、内に生起している感覚印象に関して報告できるのはその感覚印象が生じている当の主体のみである。一方で、他者はある人に感覚印象が生じていると措定できるだけである。この言語使用の違いにより、感覚印象の「報告」に独特な権威が成立すると思われる。

ある感覚印象を「報告」することはその感覚印象を自己帰属しているとともに、それに関する自己知を表明していると言えるだろう。人は自分に生じていると非推論的に知っている感覚印象について語ることができる。自分に生じている感覚印象を記述することはそれを自分に帰属させていることである。こうした言語使用は感覚印象の一人称的言説である。

物理的対象について記述する言語使用は「であるのを見る(感覚する)」の語りを当てはめ、感覚印象について記述する言語使用には「のように見える」の語りを当てはめる。感覚印象の一人称的権威はこうした現れについての語りをし始めた時に成立する。またそうした語りを習得することとともに自己知が成立する。物理的対象についての語りは対象についてあるいはそれに関する知識を述べているにとどまる。他方で感覚印象について語ることは、発話者が持つ特定のパースペクティブからの現れについて語っている。感覚印象の「報告」は感覚印象についての自己知の表現であるとともに、主体の現在の内的状態についての報告である。

(4) 今後の展望

主体が自己知に対してもつ一人称的権威について、「報告」という言語使用の特徴に注目して検討した。今後さらにこの方向での考察を進める。そして自己知自体についてセラーズがどのような立場に立っているかの検討とととはした。と考えている。セラーズ自身が直接的に自己知について論じたことはほとんどなく、また二次文献においても論じられることは同じく少ない。ひまず、彼が認識論において合理主義に立っているように、自己知においいも合理主義に立つと筆者は仮定している。もしそれでうまく解釈が通らないいる。記識論に関するセラーズの主張との齟齬が生じると思われる。その場合、セラーズ解釈、また彼がとるべき立場との整合性を考えていく必要がある。これ自体セラーズ解釈に貢献すると思われる。考察を進めた結果、新たなセラーズ解釈を導き出せるかもしれない。またセラーズが自己知についなり独特な立場をとっていると判明したなら、自己知の議論への貢献につながるかもしれない。自己知からセラーズの議論を再構成するのは、これまでのセラーズ研究であまり検討されていない観点である。

i 金杉武司 (2014) pp.167-8

- ii deVries, Willem A. and Triplett, Timm., (2000), pp.176-77
- iii deVries, Willem A., (2005), p.210
- iv Brandom, Robert. (1997), p.176
- v deVries, Willem A. and Triplett, Timm.,(2000),p.153
- vi Sellars, (1997) p.91
- vii Sellars, (1997) p.108
- viii Sellars, (1997) p.108
- ix このレプリカの生起は知覚者の状態とされるが、生理的状態として導入されるのではない。しかしセラーズはこれが生理的状態と判明する 余地を認めている。
- x Sellars, (1997) p.112
- xi Sellars, (1997) p.112
- xii Sellars, (1997) p.115

(5) 参考文献

- Brandom, Robert. (1997), Study Guide, in Wilfrid Sellars: *Empiricism* and the Philosophy of Mind, Harvard.
- ・Sellars, Wilfrid., (1997), Empiricism and Philosophy of Mind With an Introduction by Richard Rorty and a Study Guide by Robert Brandom, Harvard University Press. (浜野研三訳), (2006), 『経験論と心の哲学ウィルフリド・セラーズ』、岩波書店
- deVries, Willem A. and Triplett, Timm., (2000), Knowledge, Mind, and the Given: Reading Wilfrid Sellars's Empiricism and the philosophy of mind, including the compelte text of Sellars's essay, Hackett Publishing Company.
- · deVries, Willem A., (2005), Wilfrid Sellars, McGill-Queen's University Press.
- ・三谷直澄 (2010)「マクダウェルはセラーズをどう理解したのか?:「見えるの語りの選言主義的解釈をめぐる一考察」、信州大学人文科学論集 44 号、pp.1-20
- ・三谷直澄(2011)「セラーズにおける意図の分析について」、信州大学人 文科学論集 45 号 pp.1-21
- ・村井忠康 (2012)「知覚と概念-セラーズ・マクダウェル・「描写」-」、科学哲学 45-2 号、pp.99-114
- ・金杉武司(2014)「自己知と自己認知」、信原幸弘・太田紘史編『シリーズ新・心の哲学 I 認知篇』、勁草書房、pp.167-206.

(所属:千葉大学大学院人文公共学府人文公共学専攻)

状況づけられたエージェントの推論活動 -アブダクションと常識推論をめぐって-

森 悠貴

Abstract

The main theme of this note is to explain why we can behave in goal directed and reactive way, and what kind of features of us make it possible, by referring to recent research results of philosophy, cognitive science and artificial intelligence. Besides that, we must take into consideration that our behavior is always influenced and restricted by our own body and the environment surrounding it. Specifically, this note deals with abduction and common-sense reasoning as important reasoning we are using on a daily basis. Because these nonmonotonic reasoning allows us to adjust the amount of information and revise information.

(1)研究テーマ

我々人間は合理的な行為を遂行することのできる行為主体・エージェント(agent)である。すなわち我々は、何か一定の目的を立て(その場合、我々はその目的を達成することが、自分にとって益になると普通信じているが)、その目的のためにはどのような手段を講じるのが適切であるかを考慮し(その際、多様な情報を収集し、それらに基づいて論理的帰結を導いたり蓋然的な推測を立てたりして)、実現可能な諸手段の内から、最も適当と思われるものを選択し、それを実行している。このような我々の合理的行為は、一方で、人間に固有の特性と考えられ、伝統的な認識論の文脈においては、この営みの根底にある知性(intellectus)・理性(ratio)といった(我々人間が本来的に備えているような)能力がどのようなものであるか、ということが哲学的な問いであった。他方で、現代では、人工知能の発展に伴って、上記のような我々の活動のあり方をシミュレートし、より効率的な体系を作り上げるという課題が現実的なものとなってきた。

本稿では、以上のような 特徴的な活動のあり方を、目的志向的(goal directed)で、かつ環境対応的(reactive)という二つの特徴付けから捉えることにし、この二つの特徴を満たした活動を行いうるシステムないし行為主体・エージェントを、GDR 系[注1]と呼ぶことにする。そして、我々のGDR 系としての営みがどのようなアーキテクチャ(設計原理)によって実現

しているのかを、伝統哲学的な文脈からではなく、人工知能や認知科学の近年の動向を踏まえて掘り下げることが、本研究のテーマである。

実際に具現している GDR 系のなかで、最も典型的なものは人間であると言ってよいだろう。そして、人工知能分野の課題とは、我々人間の合理的な、GDR 系としての振る舞いの再現であると位置づけることができるだろう。GDR 系を再現するにあたって、初期の人工知能分野において採用された設計思想は、まずもって人工知能に対して、達成すべき課題や周囲の環境についての可能な限り完全な情報を与えることで、状況を(一方的に)制御できるシステムをつくる、というものであった。このことから、当時の研究者たちの間で、行為に先立って完全な状況把握を行うことが合理的振る舞いの条件だと思われていたこと、また、典型的 GDR 系である我々人間も同様に、完全な状況把握から合理的行為を導いていると考えられていたことが窺える。

例えば、迷路を抜けるという課題を果たすための人為的な GDR 系(すなわち人工知能)を設計する場合には、迷路の地図を端的に与えてしまう、あるいは、そうした完全な地図を作成する手順を組み込んでおくといった手法が、当然要求され、また最適なものであると考えられてきた。もちろん以上の述べ方は過度に単純化したものであるが、このように GDR 系として振る舞うためには、(伝統哲学の認識論が我々の合理的な活動の根幹に理性・知性といった能力があることを想定していたように) 行為主体があらかじめ詳細で俯瞰的な情報や正確な学問的知識を有していることが不可欠あるいは望ましい条件だと考える立場をここでは、GDR 系についての理想主義的見方と呼んでおこう。

ところが興味深いことに、近年における認知科学・進化心理学・進化言語学などの発展、さらには人工知能研究における知識表現論の進展などによって、GDR系を理想主義的観点に基づいて実現することには著しい困難があるということが明らかになってきた。(その挫折のもっとも有名なものがフレーム問題であると筆者は考える)

これを受けて、現代の人工知能分野や認知科学の動向に基づいて GDR 系を捉えると、以下の重要な点が見えてくる。

まず、【1】我々GDR系が行う情報の入手・処理は非常に局所的・短期的で身体依存的・環境依存的である。この特徴は人工知能開発とその周辺の諸分野で強く意識され、人間的行為主体の捉えなおしの中で重視されるようになったため、状況づけられた認知(Situated Cognition)という標語で表現され、広く用いられている。

次に、【2】我々GDR系が遂行する行為は、われわれと我々を取り巻く環境

との相互作用(interaction)と捉えるべきである。というのも、伝統的な行為論の文脈において、行為とは、行為主体の意図が帰属された運動であり、行為主体から対象への一方的なものである(interaction ではなく action)、と特徴づけられていた。しかし、【1】のような特徴を持つ GDR 系が営む行為をより精確に分析するためには、行為主体の意図を出発点に行為を考えるのではなく、環境との相互作用(interaction)を出発点にして分析をすべきである。

そして、【3】状況依存的な我々GDR系が、環境との相互作用によって合理的に振る舞うことができるのは、その背後に(無意識に活用され、その重要性が見出されにくいような)洗練された情報処理メカニズムを持っているためである。我々は先に述べたように環境依存的ではあるが、それは環境に一方的に支配されているのではなく、その中で能動的に対応策を練るというリアクションをとりながら様々な創造的活動を行っている。例えば、他者との情報交換、目標達成に成功した他者の模倣、典型例からの予測、行為に失敗した後の分析・学習などは、環境の中に組み込まれ限られた情報しか得ることのできない我々が、GDR系として振る舞うために適応(adaptation)を重ねた結果会得し、体系化した行為であると言えるだろう。

そしてこのような環境へのリアクション・適応こそが、人間が GDR 系として振る舞う(以下では GDR 的に振る舞う、等と略記する)ために本質的に不可欠なメカニズムであり、その特性について、周辺分野(認知科学や人工知能等)の研究成果を参照しながら探求することが本研究の目標である。

(2)研究の背景・先行研究

1956年のダートマス会議にて、コンピュータシステム上で人間の知的な振る舞い(合目的行為遂行)の再現を目標とする研究領域が人工知能(artificial intelligence)と名付けられた。前節で提示した研究のテーマにおいて理想主義的見方と呼んだ考え方は初期の人工知能研究(J.McCarthyや P.Hayes など)に見ることができる。その一方で、Situated Cognitionの観点に立つ研究としては、M.Tomaselloによる人間同士のコミュニケーションの分析や、R.Langackerの像図式(Image Schema)などの、進化言語学や認知言語学の現代的分析を念頭に置いている。そして哲学上の先行研究としては、P.Gärdenforsによる人工知能の哲学的基礎付けの試みや Conceptual Space論、そしてこの始原であると考えられるWittgenstein後期の言語ゲームやプロトタイプ理論が挙げられる。

(3)筆者の主張

先ほど紹介した近年の認知科学の動向では、我々の感覚知覚や情動といった情報処理の手法に注目が集まっている。しかし、本論文はこの潮流に倣うのではなく GDR系である私たちが行う推論について掘り下げる。推論は(研究背景で紹介したところの) 理想化された理性・知性によって可能になる(いわば感覚知覚や情動の対極にある) 知的営みである、という認識が一般的であった。しかし、我々が状況づけられている行為主体でありながら、GDR的な振る舞いが可能であるということをより正確に分析するには、推論に関しても理想的な見方を捨てるべきであると筆者は考える。このことを自覚できるようになったのは、近年の周辺学問(特に人工知能分野)の進展によるところが大きいだろう。と言うのも、我々が日常的に(しかも無意識的に)行っている、一見不完全に見える情報処理は、それだけで直接に観察すると単なる欠陥や不完全さのように見えてしまい、実はその根底に、我々自身が環境に対する積極的リアクションとして講じる方策・戦略といったものが隠されていることを見落としてしまいがちだからである。

そこで本論は【A】アブダクション・仮説形成(Abduction)、【B】フレーム問題(Frame Problem)と、それに密接に関係する常識推論(Commonsense Reasoning)について詳しく検討する。これらは、我々が日常的に用いる典型的な推論・情報処理であり、なおかつ、非妥当な推論が我々の GDR 的振る舞いにおいて(おそらく不可欠で)重要な役割を果たしていることを示しているにも関わらず、今までそのメカニズムや重要性が明確にされていなかったものであると言えるだろう。アブタクションや常識推論が我々の GDR 的振る舞いに対しどのような貢献をしているか、また、その働きから見てとれる哲学的興味はどのようなものか、以下で具体的に見ていくことにする。

3-A アブダクション

アブダクションとは大まかに言えば、命題 Pによって表される事実が、行為主体に知られた際に、その事実が成立している理由を最も適切に説明すると思われる仮説 Hを案出したり探し出したりして、次の行為を行うための前提として暫定的に採用すること、言い換えれば、観測した事実 Pが成り立っていることをうまく説明できるいくつかの一般化の候補の中から、もっとも適切であると思われるものを選び、その一般化が背後で成り立っているという仮説を採用することである。

我々が日常生活においてアブダクションを行う典型事例を見てみよう。

【例 1】A 氏には、市役所に出かけるという用事があり、ついでにその途

中にあるレストランで食事をとろうと考えている。市役所に行く途中でレストランに寄るか、市役所から帰る途中でレストランに寄るかの選択肢があるが、A氏はレストランでの待ち時間をできるだけ少なく済ませたいと思っている。そんななか A氏が市役所に行く途中でレストランの様子を見ると、当のレストランの前に行列がないことに気付く。実はそのレストランにはいつも順番待ちの列ができていて、食事ができるまで待たされるのが常なのだが、その日は様子が違うようだった。そこで A氏は「店の前に順番待ちの行列がないのは、おそらく今レストランがちょうど空いているためである」「従って今レストランに寄れば、待ち時間をかけずに食事ができる」と考え、市役所に行く前にレストランに寄ることにした。

つまり、A氏はいつも行列ができているレストランの前に行列がないことを観測し、その事実を説明する一般化のうち「店が空いているならば、店の外に行列はない」という一般化が適していると考え、仮説として採用した。そして、その一般化が現状に適用できるということを前提にして、往路でレストランに寄るという行為を選択した。もしA氏が採用した仮説が正しかった場合、市役所へ行くついでに、時間を無駄にせず効率的に食事をとることができただろう。しかし、仮説が間違いである場合も当然ありうる(実はレストランが休みで誰も並んでいないだけであった場合など)。

従来のアブダクション分析においては、我々がどのようにして仮説を立てているのか、そして複数たてた仮説の中からもっとも適切なものをどのように判別しているのか、といった問いが議論の中心となっていた。しかし、アブダクションに関して筆者がより注目すべきであると考えたのは、(i)状況づけられた行為主体である我々は、必ずしも妥当な推論に基づき行動しているわけではない、という点と、(ii)我々は有限である手持ちの情報をアブダクションによって拡張し、一定の仮説が立てられれば、(仮にその仮説の正確性が保障されていなくても、またさらに、未決定の事項が保留にされたままであっても)行為の実行に踏み切っている、という点である。先ほど指摘したように、状況づけられた行為主体である我々は身体的制約、環境的制約によって、合理的な振る舞いに必要な情報をすべて正確に直接入手することはそもそも不可能である。そのような中で、GDR的に振る舞うためには、不確実・未決定な事項を保留にしたままでも行為の遂行へ踏み切ることは避けては通れない。

そこで我々は、身体によって観測した有限で断片的な情報を、仮説を設定することによって拡張し、次の行為のモティベートに利用している。さらに、もし仮説が間違いだった場合でも、観測に基づいて仮説の棄却をすることで、

(仮説を立てなかった場合よりも)手持ちの情報を増やすことが可能である。しかし、アブダクションは、我々の GDR 的振る舞いにとってもっとポジティブな意味があると筆者は考える。それは、観察した状況についての情報に基づいて、行動を決定するという、我々の状況適応に貢献している点である。我々がある目標を達成しようとするとき、その成功が見込める手段はおそらく複数あることが常である。しかしその複数ある手段のうち、行為主体がどのような状況に置かれているかによって達成の容易さや実現度が変動する。そんな中、どの手段を選ぶべきか(目的実現のための手段にとって、状況がどのように影響を与えているか)はすぐ明らかになる事ではない。そこで我々はアブダクションを通して得られた仮説に基づいて、どの手段が最も目的達成に適切であるか模索しているのではないだろうか。

つまり、アブダクションとは、我々の目的達成のためにどのような手段を選択すべきかを決定する基準・条件についての仮説を、観察した状況に基づいて提供する推論であり、状況づけられた行為主体である我々にとって、GDR的に振る舞うために不可欠なものである。

以上が、本論文で強調したい、従来指摘されてこなかったアブダクションの重要な役割である。

3-B フレーム問題と常識推論

フレーム問題とは、人工知能分野から提示され、哲学者によって盛んに議論された、我々の合目的な振る舞いを人工知能で再現する際の困難についての議論である。しかし、フレーム問題がどのような問題であるか、どのような点で哲学的に重要であるか、ということについては精確な考察がなされて来たとは言えず、この点は M.Kamermans と T.Schmits によっても指摘されている。詳細は参考文献に任せるが、言ってしまえばフレーム問題とは、妥当な推論だけを行う人工システムに、我々人間が GDR 的に振る舞うために行う推論(つまり、非妥当な論理的導出を含む推論)を再現させようとしたところ、追加しなければならない公理が大量に要求され計算爆発に陥ってしまう、という問題である。そしてこの問題を巡って、(a)人間にはフレーム問題は起こらない。従って人間には人工システムによって還元不可能な知的能力が備わっている、といった議論や(b)言語を用いた、つまりシンボル操作による表象主義的なアプローチでは人間の振る舞いの再現は不可能であり、コネクショニズム的アプローチをすべきである、といった議論が哲学者の間で行われていた。

しかし実際には、その後の人工知能研究の進展を参照すると、ある種の情

報の度外視の原理を組み込んだ論理的言語(Circumscription や Default 論理など)を採用することで、計算爆発は回避することができ、さらにこうした特別な種類の論理的原理とは、実は、状況づけられた GDR 系である我々が、環境へのリアクションとして日常的に用いている論理(いわゆる常識推論)そのものであることが明らかになってきた。従って、従来哲学者が行ってきたフレーム問題解釈は的外れなものであり、むしろ、フレーム問題が示唆している哲学的に興味深い問いとは、①我々が日常的に行う常識推論とはどのようなものであるか、そして、②なぜ常識推論が我々の GDR 的振る舞いにとって役に立つのか、というものであるだろう。そして本研究の文脈からみると、デフォルト推論(Default reasoning)、閉世界仮説(Closed world assumption)などの常識推論は、(我々が行為を行う際に利用できる仮説を創出できるという点でアブダクションが有用であったのに対し)我々が行為をする際に考慮しなければならない情報を減らすことができる、という点で我々の行為に貢献していると考えられる。

我々は実際、日常的な推論において(I)(重要な影響を持たない事実に関しては)情報を得るまでは無いものと見なす、(Ⅱ)報告を聞いてから対処すればいいこと(もしくは当面の目的遂行に関して関係ない話)は考慮しない、(Ⅲ)もし新しく考慮すべきことが発生したという証拠を入手したら、その時に古い情報を棄却し新しい情報を採用する、といった特徴的な情報処理を行っている。このような情報処理が可能であるのは、我々が常識推論を駆使しているからであると筆者は考えている。つまり、常識推論には、(直接証拠を得ていない情報に対して、暗黙に、標準的な状況が成り立っていることを想定したり、考慮すべき状況は成り立っていないことを想定したりすることで)、目的達成のために扱わなければならない情報を減らす働きがある。そしてこれらの想定が正しくなかった場合でも、新たに情報が明示されれば、それに応じた手持ちの情報の改訂を可能にする。この働きこそが、我々の GDR 的振る舞いに対する常識推論の貢献である。

常識推論は我々の行為の遂行にとって不可欠なものであり、GDR的振る舞いを人工知能で再現するならば避けて通ることはできないだろう。このことを明らかにしているという点で、フレーム問題は依然として興味ある問いを我々に投げかけている主題である。

(4)今後の展望

かつての哲学史において、我々が行う妥当でない推論の重要性は指摘されてきたが、その重要性の所在はあまり明らかにされてこなかった。しかし本

研究を通して、状況づけられた主体が GDR 的に振る舞うためには、非妥当な推論を駆使することで、(情報を求めるときは不確かな仮説でも利用し、確かな情報であっても明示的に必要になるまで無視するといった) 考慮する情報量を調節する、もしくは状況に応じた手持ちの情報の改訂を行う、ということが情報処理上必須であることが明らかになった。

人工知能上での再現といった応用に耐えるような、我々の合理的行為者としての在り方についての分析をするためには、本研究が明らかにしたことが重要になってくると筆者は見込んでいる。さらに、人工知能分野や認知科学の成果と哲学的な分析のそれぞれの立場を双方向的に考慮しながら、我々の情報処理の傾向や状況適応の仕組みについて探究することは興味深い課題であると言えるだろう。

[注 1] "Goal directed"は行為論で、"Reactive"はプログラミング言語の意味論で使われている一般的意味をそれぞれ念頭においている。

(5) 参考文献

- · Dennett, D, 1984, Cognitive Wheels: The Frame Problem of AI. In The Philosophy of Artificial Intelligence edited by Boden, A, M, pp. 147-170.: Oxford U.P.
- · Gärdenfors, P,2004, Conceptual space: MITpress.
- · Gärdenfors, P,2014, Geometry of Meaning: MITpress.
- · Harmelen, F, V; Lifschitz, V; Porter, B, Eds. 2008, Handbook of Knowledge Representation: Elsevier
- · Kamermans, M; Schmits, T, 2004, The History of The Frame Problem
- · Langacker, R, 1987, Foundations of Cognitive Grammar vol. I & II: Stanford U.P.
- Langacker, R, 2008, Cognitive grammar -A basic introduction: Oxford U.P.
- · Wittgenstein, translated by G.E.M. Anscombe L, 1963 Philosophical Investigations: Basil Blackwell, Oxford
- · McCarthy, J; Hayes, P, 1969,

Some philosophical problems from the standpoint of artificial intelligence

- · Reiter, R, 1980, A Logic for default reasoning. Artificial Intelligence 13: pp.81-132.
- · Tomasello, M, 2010, Origins of Human Communication: MITpress

(首都大学東京)

様相の形而上学における傾向性主義を退ける

大畑浩志

Abstract

This paper aims to critically examine the position called
Dispositionalism in metaphysics of modality, which is the view to
explain the possibility by the disposition of individuals.

Dispositionalism is parsimony compared to other position like modal
realism using possible worlds and has the advantage of being
naturalistic in that it can explain the possibility from familiar object to
us. Dispositionalism, however, cannot handle alien properties well. It
does not go on as a consistent theory, regardless of whether it admits
alien properties or not. From the above, it is shown that
Dispositionalism is not a hopeful position.

(1) 研究テーマ

様相の形而上学における傾向性主義を退ける

(2) 研究の背景・先行研究

2-0 はじめに

本稿は、様相の形而上学における傾向性主義と呼ばれる立場を批判的に検討することを目的とする。傾向性主義とは、可能性や因果といった様相概念を、個体がもつ傾向性によって説明しようとする立場である。傾向性主義は、可能世界を用いる様相実在論のような立場に比べ、存在論的に倹約であり、また、われわれに身近な対象から可能性を説明できるという点で自然主義的であるという利点をもつ。しかし傾向性主義は、エイリアンな性質をうまく処理することができない。傾向性主義は、エイリアンな性質が存在すると主張しても、あるいは存在しないと主張しても、どちらにせよ一貫した理論として成立しない。以上より、傾向性主義はこれまでの見方に代わる有力な立場とはならないことを示す。

2-1 傾向性とは何か

様相の形而上学において、傾向性主義はどこに位置するのか。まず、必然 性や可能性、偶然性といった真理様相を、可能世界の存在にコミットして分 析する立場は、広い意味での可能世界に関する実在論と呼ばれる。可能世界についての実在論は、可能主義的実在論と、現実主義的実在論へとさらに分岐する。前者は可能世界を具体的対象として認める立場であり、主な支持者として D・ルイスがいる。後者は可能世界を抽象的対象として認める立場であり、主な支持者として A・プランティンガや、R・アダムスがいる。一方、様相の分析に可能世界を用いない立場は、可能世界に関する反実在論と呼ばれる。可能世界に関する反実在論には、虚構主義、規約主義、本質主義などの立場があるが、傾向性主義・もここに含まれる。つまり、傾向性主義とは、いかなる意味においても可能世界の存在を認めず、可能性や必然性の説明を与えようとする立場である。なお、現実主義的実在論と可能世界についての反実在論はともに「存在するものはなんであれ現実に存在する」というテーゼを共有しており、現実主義の陣営としてひとくくりにされる。よって、傾向性主義も現実主義のひとつである。

では、傾向性主義が依拠する傾向性とはどのような性質なのか。傾向性は、G・ライルのような経験主義者によって、しばしば反事実的条件法によって分析されてきた。このとき、xがもつ水溶性という傾向性は、「xが水に投入されたならば、xは水に溶けるという発現(manifestation)をもたらしただろう」という形で説明される。反事実的条件法による傾向性の分析のメリットは、傾向性という目に見えない性質の存在を、真理条件のはっきりした命題のかたちで説明できることである。還元主義的な分析が成功すれば、沸騰性(effervescency)や脆弱性(fragility)といった語を含む文の命題はすべて、経験的な命題へと翻訳される。

しかし、本稿で焦点を当てる B・ヴェター[Vetter,2015]のような傾向性主義者は、傾向性は反事実的条件法のような他の様相概念によって分析できない、原初的な概念なのだと主張する。その根拠は、反事実的条件法を採用する限り、傾向性はつねに刺激条件と発現のペアで個別化されるが、それは間違っており、傾向性は「~できる」というその発現のみによって個別化されなくてはならないというものだ。論証は以下の通りである。

ヴェターによれば、反事実的条件法による傾向性の分析は、それが刺激条件と発現のペアで考えられている限り、成功しない。たとえば脆弱性という傾向性を考えてみよう。脆弱なグラスは、ハンマーで叩かれても、二階から落とされても、長い時間をかけて一定の力を加えられても、壊れてしまうだろう。また、作りはしっかりしているが非常に古い脆弱な椅子は、急激な気温の変化によって壊れてしまうだろう。このとき、脆弱性を説明するために急激ひとつの刺激条件を選ぶことはできない。もし脆弱性を説明するために急激

な気温の変化を選んだならば、グラスは脆弱性をもたないことになるだろう。 なぜなら、壊れやすいグラスであっても、気温の変化には耐えてしまうから である。同様に、ハンマーによる衝撃を選んだならば、古い椅子は脆弱性を もたないことになってしまう。古い椅子は気温の変化には弱いかもしれない が、衝撃には強いということがありうるからである。

さらにヴェターは、条件法分析の刺激条件は、質的にだけではなく、量的にも一意的に決定できないと指摘する。条件法の刺激反応として、ハンマーによる衝撃を選んだとしよう。そのような場合、脆弱な物体を壊すのに十分な力は、1Nなのか、8Nなのか、140Nなのかといったことを決定することはできない。衝撃の値を恣意的に決めてしまえば、質的な決定に伴う困難と同じ問題が発生してしまう。ヴェターは、条件法分析の刺激条件を特定できないという問題は、脆弱性のような特定の傾向性だけでなく、ほぼすべての傾向性に当てはまると主張している。ii

こうした議論から、傾向性は端的にその発現のみによって個別化されなければならないという主張が導かれる。傾向性は、一般にひとつの条件法による特徴づけでは捉えられず、無数の条件法に対応づけられる。傾向性は、個体がもつ「~できる」という能力にほかならず、個体に局在的な(localized)可能性として捉えられる。傾向性のこの特徴を活かし、ヴェターを含めた傾向性主義者たちは、因果や反事実的条件法といった他の様相概念の説明項として、傾向性を用いる。よって、傾向性はそれ以上分析されない、原初的な様相概念となる。

2-2 傾向性を拡張する

ヴェターによれば、傾向性とは個体がもつ「~できる」という能力を端的に示す性質であり、個体に局在的な可能性である。iiiこのような理解のもと、ヴェターは傾向性を潜在性(potentiality)という概念へと拡張する。たとえば脆弱性という潜在性は、壊れることができるすべての個体がもつ性質である。コンクリートの橋であっても、それが壊れることも可能である以上、その橋は脆弱性という性質をもつ。このように解釈し直された傾向性は、「歩くことができる能力」や「楽器を演奏することができる能力」といった能力(ability)や力能(power)を包摂する。

ヴェターは、潜在性として理解された傾向性は、傾向性を表現する語の文脈依存性(context-sensibility)や曖昧さ(vagueness)をうまく捉えることができると主張する。植木鉢は、レンガと比べれば壊れやすいが、シャンパングラスと比べれば壊れにくい。脆弱性は文脈によってさまざまな仕方で帰属されるが、それは植木鉢もレンガもシャンパングラスもすべて「壊れることが

できる」という潜在性をもっているからにほかならない。それらが比較できるのは潜在性に度合いがあるからであり、われわれは普段その潜在性の度合いが高いものを指して傾向性をもつと考えているのだとヴェターは主張する。

潜在性あるいは力能として理解された傾向性は、すでに広範囲の可能性をカバーしている。ヴェターや A・ボルジニ & E・ウイリアムス [Borghini&Williams,2007]といった傾向性主義者は、さらに様相概念全般を傾向性に還元するため、反復的潜在性(iterated potentiality)ないし高階の傾向性(higher-order disposition)を導入する。これらは、現実の個体がもつ傾向性が発現し、何らかの事態が成立したならば、その事態はさらに別の傾向性をもつだろうという考えに支持されている。たとえば、ガラスのコップは脆弱性という傾向性をもつが、もしガラスのコップが割れたならば、その破片は人を傷つけることもできる。したがって、ガラスのコップは脆弱性という一階の傾向性をもち、凶器となるという二階の傾向性をもつ。ivこのように、傾向性の発現による可能性のレンジを広げることで、適切な傾向性が適切に配置されれば、ケネディがオズワルドに撃たれなかったかもしれないというような複雑な事態ですら、傾向性で説明できるとボルジニ&ウイリアムスは主張する。

2-3 ヴェターによるエイリアンな性質の説明

傾向性主義にとっての困難のひとつは、エイリアンな性質が存在しえたことを説明できないことだ。エイリアンな性質とは、これまで現実世界に存在してきたあらゆる個体に例化されてこなかった性質である。エイリアンな性質は、ルイス[Lewis,1986]が言語的代用主義を批判する際に導入した。vルイスは、言語的代用主義であってもエイリアンな性質の存在自体は認められるが、そうした性質が複数あった場合、それぞれの見分けがつかないことを指摘した。つまり、言語的代用主義は、エイリアンな性質のみが異なる可能世界を、ひとつの可能性としてひとくくりにしてしまう。さて、一見したところ、傾向性主義もエイリアンな性質にはうまく対処できない。傾向性主義は、エイリアンな性質の見分けがつかないどころか、そうした性質が存在することができたという可能性そのものを説明できないようにみえる。なぜなら、傾向性主義は、現実に存在する個体の傾向性から構成される可能性しか認めないからである。

傾向性主義であっても、なんらかの個体は未知の性質を生み出す傾向性を もっていると言うことによって、エイリアンな性質を説明できるかもしれな い。しかし、これまで現実世界に存在してきたあらゆる個体に潜在的にも例 化される可能性のなかった性質が存在しえたことは説明できない。もちろん、 その性質をもつ可能的対象、つまりエイリアンの存在可能性も説明できない。 ヴェターは、こうしたエイリアンな性質に対して、傾向性主義に利用可能 な二つのオプションを提示している。まず、(1)傾向性主義の利点は損なわ れるが、エイリアンな性質を認めるという選択肢、次に(2)エイリアンな 性質をそもそも認めないという選択肢である。順に検討してゆく。

(1)においては、普遍者あるいは命題という抽象的対象が、エイリアンな性質を発現する潜在性の担い手として想定される。エイリアンな性質を \mathbf{F} とするならば、 \mathbf{F} に対応する普遍者 \mathbf{F} -性が存在する。普遍者 \mathbf{F} -性は、これまでけっして例化されたことがなく、また現実に存在する個体はすべて \mathbf{F} -性を例化する潜在性を有していないとしよう。それでも、 \mathbf{F} -性自身は、それが普遍者である以上、例化されるという潜在性をもつ。ヴェターは、命題を用いても同様の論証を行う。「性質 \mathbf{F} が存在する」という命題は、たとえそれがつねに偽であったとしても、真となる潜在性をもつ。ヴェターは、普遍者や命題という抽象的対象に訴え、現実に存在する対象のみによって、エイリアンな性質への潜在性のルートを確保しようとする。

しかしこのような戦略は、われわれの周りに存在する具体的対象からすべての形而上学的可能性を説明するという傾向性主義のポリシーに反している。なぜなら、(1)によるアプローチがうまくいくためには、プラトン主義的に普遍者や命題を解釈した上で、さらに潜在性の担い手の幅を抽象的対象にまで広げなくてはならないからである。このような困難から、ヴェターは(1)よりも(2)の選択肢を好んでいる。

(2)の戦略は、エイリアンな性質をそもそも認めないというものだ。な ぜエイリアンな性質がないといえるのか。それは、ヴェターによれば、傾向 性主義は以下に示す潜在的例化原理(Principle of Potential Instantiation)を 自然に受け入れられるからである。

あらゆる普遍者は、少なくとも潜在的に例化されなければならない。Fである性質の普遍者が存在するのは、Fであるか、潜在的にFであるか、または潜在的にFであるような何かを存在させるような、そうした個物が存在するときに限る。

PPI は、すべての普遍者は例化されていなければならないという $D \cdot P - \Delta A \cdot P = D \cdot P =$

上、現実の対象に(潜在的にも)例化されることはないので、PPIが認められれば、エイリアンな性質の普遍者は存在しない。よって、エイリアンな性質は存在しない。当然、エイリアンな性質を例化する存在、つまり可能的対象としてのエイリアンも認められない。

(3) 筆者の主張

傾向性主義の二つのオプションに対しそれぞれ批判を加えたい。おおざっぱにまとめると、以下の通りだ。(1)の戦略をとれば、存在することが可能なすべての性質の普遍者がこの現実世界に存在することになり、プランティンガが認めるような、必然的存在者としてのこのもの性と見分けがつかなくなる。(2)の戦略では、PPIに問題があり、何がエイリアンな性質であり、何が正しく認められる性質なのかの区別がつかなくなる。

- (1)の問題は、ヴェター自身が認めている困難に加え、エイリアンな性質も含むすべての性質が、すでに普遍者ないし命題の形で、この現実世界に詰まっていると考えられているということだ。このような考えは、プランティンガ[Plantinga,2003]が導入したこのもの性(haecceity)の考えに近い。viプランティンガは、現実世界にはユニコーンもエイリアンも存在しないが、「ユニコーン性」や「エイリアン性」といった性質は必然的に存在し、それは現実世界においても利用可能な性質だと考えた。傾向性主義の魅力は、その存在論的な倹約性と、われわれに身近な個体から可能性を説明できることとであったが、いまや、その立場は自らが批判する相手と同じになっている。しかも、このようにパワフルな普遍者や命題を認めてしまうだろう。そうであれば、われわれの直観に沿うように可能性をすくい上げるというタスクにおいて、傾向性主義は可能世界に関する実在論よりも力不足なのは明らかであるから、傾向性主義が他の現実主義に対して有利である点はなくなってしまう。
- (2)の批判に移ろう。(2)が問題なのは、何がエイリアンな性質であり、何が正しく認められる性質なのかの区別がつかなくなるということである。アームストロングの例化原理であれば、このような問題は発生しない。なぜなら、現実に例化されている性質はすべて認められ、かつ、現実に例化されていない性質は端的に認められないからだ。何がエイリアンな性質で、何がそうでない性質なのかの区別は、現実世界をみるだけで明らかである。しかし、PPI はこのような区別を単純につけることができない。次のような事例を考えよう。時点 t において、ある個体 a が F という性質を例化する潜在性をもっており、かつ、a 以外のすべての個体は F を例化しておらず、F を例

化する潜在性ももっていない。このとき、PPI によれば、F はエイリアンな性質ではないことがわかる。次に、その後の時点 t'において、その個体 a が消滅するとしよう。このとき、性質 F の担い手は潜在的にもいなくなるため、PPI によれば、性質 F は存在しなくなる。よって、時点 t'においては、性質 F はエイリアンな性質だと判定されるだろう。しかし、例化されるかもしれなかった性質はエイリアンな性質ではないため、性質 F はエイリアンな性質ではない。時点 t'においては、性質 F がエイリアンな性質なのか、そうでないのかを区別することができない。時点 t'から時点 t をふり返ったとしても、a が実際に F を例化していない以上、性質 F がエイリアンな性質だったのかどうか、つまり例化される可能性のあった性質かどうかはわからないのである。vii

(4) 今後の展望

本稿では、様相の形而上学における傾向性主義を批判した。ただし、本論でも指摘した通り、エイリアンな性質にまつわる困難は、傾向性主義だけでなく、現実主義全般が抱えている問題である。私自身は、現実主義的実在論、とりわけプランティンガによるこのもの主義+事態としての可能世界をとる立場がもっともらしいと考えているため、こうした立場で、エイリアンな性質をいかに説明するかというのは今後の課題となる。また、傾向性は可能性の説明には使えないことがわかったとして、それでも傾向的性質はあるのか、それともカテゴリカルな性質しか認められないのかということも問題である。現時点では、アームストロングがそうであるように、現実主義的実在論はカテゴリカルー元論を採用するべきであると考えている。

i カテゴリカルな性質の存在を認めず、すべての性質は傾向性であると主張する傾向性主義(汎傾向性主義 Pandispositionalism ないし傾向性一元論 Dispositional monism)とは、区別されなければならない。様相についての傾向性主義の目的はさまざまな様相を傾向性によって説明しようとするだけである。様相についての傾向性主義者の中には、カテゴリカルな性質の存在を認める者もいれば、認めない者もいる。

ii より正確には、確定者(determinate)ではなく、確定可能者(determinable)に基づく傾向性はすべて刺激条件を特定できない。

iii すべての傾向性主義者が、傾向性を能力として捉えているわけではない。 D・ジャコブ[Jacob,2009]のような傾向性主義者は、傾向性を、反事実的条件 法の Truthmaker として考えている。だがヴェターによれば、このようなア

プローチは、傾向性が反事実的条件法によって分析できないのと同じ理由で 退けられる。

iv ヴェターによれば、高階の可能性は、さらなる別の傾向性への(for further dispositions)傾向性であって、傾向性の傾向性(dispositions of dispositions)ではないと厳密に規定している。しかし加地(2016)は、ヴェターによる反復的潜在性の使用は混乱していると指摘している。

v 言語的代用主義とは現実主義的実在論の立場のひとつであり、あらゆる 関係や性質を名指すことのできるパワフルな言語を想定し、そうした言語に よって記述された文の極大無矛盾集合によって可能世界を構成する。

vi プランティンガはこのもの性ではなく個体本質(Individual essence)という語を用いているが、両者はしばしば同じものとして扱われる。正確には、このもの性に対するひとつの理解、あるいはバリエーションとして、個体本質を考えるのが正しいと思われる。

vi 何がこれから生まれてくる性質で、何がエイリアンな性質であるかが、普遍者の世界で定まっているならば、このような問題は起こらない。しかし、傾向性主義にとって、こうした強い形而上学的コミットメントは認められないであろうし、認めたとしても、それはエイリアンな性質が存在しないことをトリビアルな形で述べているに過ぎない。

(大阪市立大学)

(5) 参考文献

Armstrong, D.M. (1978). A Theory of Universals. Cambridge: Cambridge University Press.

Bird, A. (2009). Natures metaphysics: laws and properties. Oxford: Clarendon Press.

Borghini, A., & Williams, N. E. (2007). A Dispositional Theory of Possibility. *Dialectica*, 62(1), 21-41.

Jacobs, J. D. (2009). A powers theory of modality: or, how I learned to stop worrying and reject possible worlds. *Philosophical Studies*, 151(2), 227-248.

Lewis, D. K. (1986). On the plurality of worlds. Oxford: Blackwell.

Plantinga, A. (2003). Actualism and Possible Worlds. Essays in the

 $Metaphysics\ of\ Modality, 103-121.$

Vetter, B. (2015). Potentiality: from dispositions to modality. Oxford: Oxford Univ. Press.

Wang, J. (2014). The Modal Limits of Dispositionalism. $No\hat{u}s$, 49(3), 454-469.

加地大介. (2016)「力能と様相」『埼玉大学紀要 (教養学部)』,53-2

死の害のタイミング問題

佐々木渉1

Abstract

According to the deprivation approach, why death is bad for the person who dies is not answered by its positive features, but by what it deprives of her. In the metaphysics of death, this approach is a widely accepted explanatory principle of the death's harm. However, deprivationists face so-called the timing puzzle(TP). TP is the problem concerning the question when the death is bad for the person who dies. There have already been proposed five candidates to solve TP. In this paper, I point out that subsequentism, which seems to be achieving a leading position in five, cannot be maintained. Also, I defend a variation of concurrentism and to support this idea, I suggest that the deprivation might be considered as some sort of temporal process.

1 研究テーマ

死の形而上学では死がなぜ悪いのか、死が悪いとすればその害はどのようなものかという問題が探究される。死の害がどのようなものか、一つの有力な学説は剥奪説と呼ばれる考え方である。剥奪説は、死が悪い理由を死がわれわれから望ましいものを剥奪することによって説明する。これによれば、死が私たちにいかなる経験ももたらさなくとも、死が剥奪するものによっては死が悪いということを説明できる。しかし剥奪説は、死の主体が死の害をいつ被るのかという問題に直面する。この問題はタイミング問題とよばれ、すでにいくつかの解決策が提案されてきた。現時点で最も有力なのは死後説である。本論文では、死後説には問題があることを指摘する。そしてその問題点は剥奪説を比較説として表現することに関わっていることを指摘する。さらに、問題点を回避するために同時説を擁護することを試みる。その際、これまで剥奪を瞬間的な出来事として仮定してきたように思われる剥奪説に対し、剥奪を時間幅のある出来事として捉える方法を模索する²。

2 研究の背景・先行研究

2.1 剥奪説

死の害についての剥奪説 (deprivation approach) は Nagel (1993) によって最初に提案された。ネーゲルによれば「死が悪いのは、死がもたらす積極的な特徴

によるのではなく、死が私たちから奪うものの望ましさによる」という (Nagel 1993 p. 64)。剥奪説は、死が悪い理由に、不快な経験のようなものを必要としないため、死が私たちにいかなる経験ももたらさないにも関わらず悪い理由を説明できる。また私たちは若くして死ぬことを老いてから死ぬことよりも不幸だとみなすことがあるが、剥奪説によれば、若くして死ぬ人は、老いて死から死ぬ人よりもより多くの生の良きもの (長く生きていれば享受できたであろう利益や価値) を剥奪されていると説明することができる。剥奪説はこのように死に対する私たちの一部の直観をよく捉えており、魅力的な説明原理であると言える。

2.2 比較説

多くの文献では剥奪説の考えをより明確に表現するために、可能世界を用いて価値を比較するという方法がとられる。この方法は比較説 (Comparativism, Luper 2007) と呼ばれ、たとえば次のように定式化される 3 (Johansson 2013, p. 257)。

S にとっての可能世界w における出来事E の全体的な価値

- =S にとっての w の内在的価値
- -E が生じていない、w に最も近い可能世界 w^* の S にとっての内在的価値

比較説は、剥奪説の考えをより明確にするだけでなく、価値を計量することで、 実際にケース同士を比較することができる利点がある。ここでこのように剥奪 説や比較説をとることは死の害について探究を進めるうえで多くの利点をもた らす。しかし、剥奪説や比較説はタイミング問題と呼ばれる哲学的な難題に直 面する。

2.3 タイミング問題

死の害のタイミング問題は、死の主体が死の害をいつ被るのかと言う問題であるが、その際エピクロスの次の論証が問題となる。

 \cdots 死は、 \cdots (中略) \cdots じつはわれわれにとって何ものでもないのである。なぜかといえば、われわれが存するかぎり、死は現に存せず、死が現に存するときには、もはやわれわれは存しないからである。(エピクロス 1959 p. 67)

死の害についてのタイミング問題の困難は、通常の害についてのタイミング問題と比較するとわかりやすい。たとえば、私が道で転んだとする。私は転んでできた傷に痛みを感じるだろうし、自分の不注意に落ち込むかもしれない。

あるいは病院に行くために、ハイキングに行く予定をキャンセルしなければならないかもしれない。そうした身体的、精神的苦痛や、転んだという出来事が剥奪したハイキングの楽しみの喪失は、転んだ傷が癒えるまで続くかもしれない。そしてまさにその期間、私は転んだという出来事による害を被っていると言える。害を被るタイミングについてのこうした考え方は、常識的であるとともにもっともらしい。

しかし、死の害については同様に考えることはできない。なぜならエピクロスの言う通り、死者はもはや存在せず、一見して害を被る主体が存在しないように思われるからである。タイミング問題の解決をめぐっては、死の主体が害を被る時点に応じて、生前説、同時説、死後説、永久説、無時間説の5つの解決策が主に唱えられている。この問題についてのそれぞれの論者の見解は鋭く対立しており、それぞれの見解に長所と短所が見出されている状況である。

2.4 5 つの解決策

この節では、タイミング問題についてすでに提案されている 5 つの解決策それぞれについて、長所と短所をまとめ、論争状況を整理する。

死後説 (subsequentism) 死後説 4 の主張は明確である。人は死によって害を被る。害を被る時点は死が発生した後の時点だと考えるのがもっともらしい。したがって人は死後に害を被る。死後にどのような害を被るかは比較説に基づいて、もし死ななかった場合に得られた価値の差し引きで説明できる。

死後説の問題点はもちろん、死の主体が存在しない時点で、死の害を被ることができるのかという点である。このことを正当化するために死後説の擁護者は死者が死後も性質の帰属者になることを強調する (Ruben 1988 pp. 211-215)。たとえば「ナポレオンが 1988 年に賞賛される」という文において、ナポレオンは 1808 年に死んだにも関わらず、1988 年において「賞賛されている」という性質が適切に帰属させられている (Ruben 1988 p.212)。そして「死の害を被る」という性質もこのように考えれば、死後に死の主体が害を被ることは不思議なことではない (鈴木 2011 p. 19)。概ねこのようにして死後説の擁護者は死が死ぬ主体にとって害であること、あるいは悪いことであることが、死後の時点においてその主体に帰せられることに問題はないと結論づける。

<u>生前説 (priorism)</u> 生前説 5 は、生前説は死の主体が死の害を被るのは、死によって否定された主体の欲求が発生した時点であると考える 6 。たとえば、私はこの論文を投稿したいと考えているが、不幸にもその前に死んでしまったとし

よう。このとき私の「この論文を投稿したい」という欲求は、死によって挫折させられてしまう。このとき私がいつ害を被ったかといえば、挫折させられた「この論文を提出したい」という欲求を持っていた時点だと考えるのが自然だと生前説論者は主張するのである。ここで生前説は、死後の出来事が生前の私に因果的に作用するとは述べておらず、逆向き因果にコミットしている訳ではない。生前説論者は、害を被る時点において主体の存在を確保する代わりに、ある時点の福利がその時点の出来事だけで決まるということを否定するのである。

しかし生前説は、死後の出来事が死ぬ主体にとって害であるためには、生前にその出来事に対して欲求を持っていなければいけないということになる。また Johansson(2013) によれば、欲求の充足は生前に抱かれていたものだけが対象となるため、もし生き延びていたら抱いていたであろう様々な欲求はすべて対象外となってしまうという問題もある (Johansson(2013) pp. 261-2)。

生前説にはこのように不可解な点があるものの、死後説が直面する主体の問題を回避できるという大きな利点がある。エピクロスの論証を重視するならば、 生前説は有力な候補になると思われる。

同時説 (concurrentism) 同時説は Lamont(1983) によって主張される立場である。ラモントはまず生前説と永久説が受け入れられないことを指摘した上で、死以外の剥奪の害は、剥奪が行われる時点で被ると考えられるため、死の害も死と同時に被られると考えられると結論づける。私見では、ラモントの主張の問題点は、死後説が考慮されていないことである。もし死と同時に剥奪が行われるとしても、その時点で死の主体が存在しないならばそれは死後説と同じである。それゆえ、ラモントの指摘は同時説を支持する強い理由とはならないように思われる。

永久説 (eternalism) 永久説 7 によれば人は死の害を永久的に (eternally) 被る。永久説論者がこのように主張する理由は次のようなものである。たとえばリンゼイという女性の死について二つの可能世界同士の価値関係を比べたとき「それらの世界が特定の価値関係にあるのならば、リンゼイが存在するときだけでなく、彼女が存在しない時点でも、それらはその価値関係にある」(Feldman 1993 p. 320)。

永久説の主張のポイントは、比較説を用いた結果、死が悪いものだと判明 するとすれば、それはすべての時点においてそうであるという点にあるが、 ${
m Luper}(2007)$ の指摘するように、これは「死の害をいつ被ったのか」という問

題を「死が害であったということが、真であるのはいつか」という問題と混同することに起因している (Luper 2007 pp. 128-9)。したがって、永久説はタイミング問題の答えとしては適切ではないといえる。

無時間説 (atemporalism) 最後に、無時間説 8 を検討する。無時間説によれば人は死の害を被るが、それは「どの時点でもない」(Johansson 2013 p. 216)。 Johansson(2013) によれば、無時間説をとる理由は、他のどの立場も少なからぬ理論的コストを支払う必要があるが、無時間説はそうしたコストを免れているからだという (Johansson p. 267)。しかしその代わりに無時間説は、死の害を他の害と違って特別なものと考える。無時間説は、死の害が他の害と異なる理由を説得すべきであると批判される (Johansson p. 268, 鈴木 2011 p. 21)。

以上により、タイミング問題の答えとして現状有力な立場は死後説、生前説 そして、無時間説であると言える。

3 筆者の主張

この節では、タイミング問題の答えとして現状有力である死後説、生前説そして、無時間説を再検討し、いずれの立場も難点を抱えることを指摘する (3.1節)。そのうえで、対案として、同時説の復権を提案する (3.2 節)。

3.1 死後説、生前説、無時間説の難点

まず死後説と生前説を検討しよう。両者はトレードオフの関係にある。死後説がエピクロスの論証をきっぱり否定し、死を被る主体が害の時点に存在するという要件を外すことで、それ以外の直観を救おうとするのに対し、生前説は死を被る主体を害の時点に確保する代わりに、ある時点の内在的価値がその時点のみの出来事で完結するという直観を手放してしまう。この点、生前説と死後説の論争は二つの立場が守ろうとする直観のどちらをとるかという論争に終始し、膠着してしまいそうである。

しかし、生前説が欲求の充足という考えに依存している点を考慮すると、この考えに依存せずに比較説と適合する死後説は、生前説にまさる。剥奪説や比較説が私たちの直観をうまく説明する説明原理であることはすでに述べた。一方、欲求の充足による説明では、死後にありえた出来事の全てを比較することはできない。そのため、生前説は比較説とはうまく適合しない可能性がある。また、私の理解では、生前説をとると、たとえば私が死後に悪口を言われることが害であるかどうかは、生前に悪口を言われたくないと思っていたかどうかに依存してしまう。しかしこれは不可解であろう。したがって、生前説と死後説では、

死後説の方が見込みのある立場であるといえる。

他方、死後説にも独自の難点がある。第一に、死後のいつの時点まで死の害を被るのかという問題がある。これには十分生きたと言える時点だと答えることができる。しかしそうすると、死者が死後のある時点から突然害を被らなくなるということになり、奇妙である。他方これを回避するためには、死者は死後、無限に害を被ると主張する必要があるように思われるが、これも受け入れがたい立場であるように思われる。第二に、エピクロスの議論を重大に受け止める人々にとっては、死後説は受け入れられないという問題点がある。エピクロス主義者は、死後説論者が自説の擁護のために持ち出す死後の性質帰属の例を否定してしまうように思われるからである。

ここで、一つの方法は無時間説をとることである。しかし私の考えでは、無時間説は、タイミング問題に答えること自体を否定する立場であるように思われる。また死は時間的な現象であることには疑いがないのだから、死の害も時間の中に位置付けられるものとして考えられる方が良い。そこで私は無時間説を支持するのではなく、同時説の復権を提案する。

3.2 同時説の検討

Lamont (1983) の見解では、同時説を支持する強い理由にはならないことはすでに述べた。しかし同時説は別の仕方でも擁護できる可能性がある。そこで難点となるのは、死と同時の時点で死の主体が存在するのかどうかという点である。私の見解では同時説において死の主体が存在しないと考えられてしまうのは、死が瞬間的な出来事だと考えられているせいであると思われる。しかしこれではあらゆる出来事の時点を決めることが困難である。そもそも死が瞬間的な出来事であるということは自明ではない。むしろ死は時間幅を持って進行するプロセスであるように思われる 9 。死が時間幅を持つプロセスであれば、同時説に関わる上記の難点は解決可能であるように思われる。死はその死が進行する期間において主体から望ましいものを剥奪すると考えることができるからである。

4 今後の展望

上述の通り、同時説は、死を瞬間的な出来事ではないと考えることで、その 難点を回避できる。一方で、このように再定式化された同時説では、比較説に 基づいて死の害を説明しようとすると、死の害は死のプロセスで生じた個々の 出来事が生じなかった場合に得られた利益に制限されてしまう。しかしこれで は、死の害は不当に小さくなってしまう可能性がある。ここで一つの解決方法は、剥奪説の定式化として、比較説をとることをやめることである。たとえば死によって剥奪されるものを、能力や可能性、潜在性だと考えることができるかもしれない。この見解に基づけば、私がたった今死んだ場合に剥奪されるのは、私がもし今死なずに生きのびてこの論文を投稿できた世界と、そうならなかった世界の内在的価値の差ではなく、この論文を投稿する可能性や能力そのものである言える。こうすることで、同時説に基づいても、死の害は適切に評価されうる。死の害をこのように考える立場については、同様の見解が Hanser(2008) などですでに指摘されており、この考えを同時説の擁護に用いることの見通しは、暗くないように思われる。

注

- ¹ 大阪大学大学院人間科学研究科、e-mail: wataru.sasaki.001@gmail.com 本稿を修正する際に、匿名の査読者から有益なコメントをいただいたので感謝 したい。
- ² 本論文では「人は死ぬと存在しなくなる」という終焉テーゼ (Feldman 2000) を前提とする。また死の害や死が悪いというときには、もっぱら死ぬ主体にとっての価値を論ずる。
- ³ 比較説については、Feldman(1993 pp. 314-316)、Feit(2002, pp. 366-373)、Luper(2007, pp. 82-88) など多くの論者が独自の定式化を提案しているが、以下の議論ではこれらの定式化の違いは影響しない。
- 4 死後説を支持する文献には $\rm Bradley(2009)$ 、 $\rm Ruben(1988)$ 、 $\rm Feit(2002)$ 、鈴木 (2011) などがある。
- 5 生前説を支持する文献には Pitcher(1993)、Feinberg(1993)、Li(1999)、Luper(2009) などがある。また Pitcher(1993) によれば、アリストテレスもこの立場を取っているという (Pitcher 1993 pp. 163–4)
- 6 このように生前説論者は死の害を比較説的というよりも欲求充足的に捉えている。実際 Picther(1993) はもっぱら欲求の充足に注目していて、剥奪説や比較説を主張していない。しかし Luper(2007) は生前説は比較説の死の害と両立すると述べている (Luper(2007) p. 136)。

- 7 永久説を支持する文献には ${
 m Feldman}(1993)$ がある。また永久説は時間や持続の形而上学における永久主義 $({
 m eternalism})$ とは全く別の主張である。
- 8 無時間説を支持する文献には Nagel(1993)、Silverstein(1980, 2000)、Johansson(2013) などがある。ただし、Nagel(1993) や Silverstein(1980, 2000) の立場は不定説 (indifinitism) と呼ばれることがある (Luper 2007 p. 136)。 Johansson(2013) によれば両者の違いは、不定説が、害を被る時点はあるが決められないと主張したり、害を被る精確な時点を決められないあるいは、タイミング問題に答えはないと主張するのに対し、無時間説は「どの時点でもない」を正確な答えだと主張する点にある (Johansson 2013 p. 266)

 9 たとえば、 $\mathrm{DeGrazia}(2017)$ によれば、死の定義についての哲学的議論の中には、死を時間的に幅のあるプロセスだと考える立場や、束概念としての死を定義するという立場があるという。

文献

- [1] Bradley, B. (2009) Well-Being and Death Oxford University Press.
- [2] エピクロス (1959) 『エピクロス 教説と手紙』 出隆、岩崎允胤訳、岩波書店
- [3] DeGrazia, D. (2017) "The Definition of Death" *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition)
- [4] Feldman, F. (1991) "Some puzzles about the evil of death" in Fischer(eds) pp. 305–326
- [5] Feldman, F. (2000) "The termination thesis" Midwest Studies in Philosophy 24 (1):98-115.
- [6] Feit, N. (2002) "The time of death's misfortune" Noûs 36 (3):359-383.
- [7] Fischer, J. M. (ed.) (1993) The Metaphysics of DeathStanford University Press.
- [8] Johansson, J. (2013) "The Timing Problem" In Fred Feldman Ben Bradley (eds.), The Oxford Handbook of Philosophy of Death pp. 255– 273.
- [9] Hanser, M. (2008) "The metaphysics of harm" *Philosophy and Phenomenological Research* 77 (2):421-450.
- [10] Lamont, J. (1998) "A solution to the puzzle of when death Harms its victims" Australasian Journal of Philosophy **76** (2)pp. 198-212.

- [11] Li, J. (1999) "Commentary on Lamont's when death harms its victims" Australasian Journal of Philosophy 77(3) pp. 349–357.
- [12] Luper, S. (2009) The Philosophy of Death Cambridge University Press.
- [13] Nagel, T. (1970) "Death" in Fischer(eds) pp. 60-69
- [14] Pitcher, G. (1984) "The Misfortunes of the Dead" in Fischer (eds) pp. $159{-}168$
- [15] Ruben, D-H. (1988) "A puzzle about posthumous predication" *Philosophical Review* 97 (2) pp. 211-236.
- [16] Silverstein, H. S. (1980) "The evil of death" Journal of Philosophy 77(7) pp. 401-424.
- [17] Silverstein, H. S. (2000) "The evil of death revisited" Midwest Studies in Philosophy 24 (1) pp. 116-134.
- [18] 鈴木生郎 (2011)「死の害の形而上学」科学基礎論研究 **39(1)** pp. 13-24

行為の合理化の論理構造——主たる理由の時間的・様相的特性を巡って

笹本もも

Abstract

Davidson points out that an action can be rationalized by specifying its primary reason. We consider which logical language we should adopt when formulating the primary of an action. Although Davidson is content with classical first order predicate logic, we construct several examples which clearly show the necessity to introduce some extension of Prior's temporal logic: hybrid logic. We will explain the reason and the philosophical outlook of analyzing primary reasons in terms of hybrid logic.

(1) 研究テーマ

(1-1)「行為の合理化」という考え

哲学において、行為(action)は、長い間、重要な研究主題となってきた。それだけでなく、近年それはますます大きな関心を集めるようになっている [1]。このことには、認知科学・進化心理学・人工知能・応用倫理といった哲学の周辺分野の急速な発展など、色々な理由があるだろう。だが、とりわけ重要なきっかけとなったのは、D.デイヴィドソンによる行為の合理化(rationalization of action)についての考察だと思われる [2]。行為の合理化とは、一言で言えば、ある行為者がなぜある行為を行おうとするのか(行ったのか)という問いに対する答え(説明)を与えることである。これが「合理化」と呼ばれるのは、もしも我々が当該行為の理由を適切に与えることができれば、その行為にはもっともな理由があり(with good reason)、我々はこの行為者を(当該行為との関係において)合理的である、理解可能であると認めることができるからである。

デイヴィドソンのこうした考えは、自己や他者の行為に理由を求めたり与えたりする我々の何気ない日常的営みが、実は自己と他者を適格な行為主体つまり合理的主体として認め合うための本質的な手続きであることを示した点で、行為の理由という主題への哲学者の関心を改めてかき立てたと言ってよいだろう。とはいえ、一口に「理由」と言っても何でもよいわけではない。例えば、ある行為における行為者の身体運動の生理学的な記述が与えられれば、なぜその身体運動が生じたかの理由は明らかになるとしても、行為の理

由が明らかにされたことにはならず、合理化は果たされていないだろう(不 合理な行為であっても、身体運動の生理学的記述は可能である)。この点につ いて、デイヴィドソンは次のような考えを提起した。すなわち、行為の合理 化として働きうるような理由とは、基本的には、彼が「主たる理由(primary reason)」と呼ぶ種類のもの、すなわち、「賛成的態度 (pro-attitude)」と「信 念 (belief)」のペアでなければならない、ということである。用語法だけか ら見るとこれらは少々奇異にも思えるが、実際の用法を見れば、デイヴィド ソンの考えが十分納得のいくもの――我々が日常的に行っている行為の合理 化の在り様に適合しており、また同時に、アリストテレス以来の行為論(手 段と目的の特定によって人間の行為を説明する考え)にも則ったもの--であ ることがわかる。なぜなら、基本的に、(1)賛成的態度とは、当該行為におい て目的となっている事柄(より詳しくは、これはトークン的ではなくタイプ 的なものである)を実現しようとするこの行為者の心的態度のことであり、 (2)また、ここでの信念とは、現に着手された(されようとしている)この行 為(これはタイプではなく、トークン的なもの、個別的出来事である)が、 (1)における目的を実現するための一つの手段となっているはずだという、行 為者の信念のことだからである。

(1-2) 主たる理由をいかに定式化すべきか

本研究では、主たる理由を与えることで行為の合理化が行われるという、 以上のようなデイヴィドソンの基本的考えを受け入れた上で、さらに進んで 次のような問題を考えたい。

一般に、主たる理由は、具体的にどのような仕方で(どのような言語表現を用いて)与えられるべきだろうか。言い換えれば、当該行為における賛成的態度と信念のそれぞれをできるだけ適切・精確に定式化するためには、どのような論理的道具立てを用いるべきだろうか。この問題は、一見そう思われるほど簡単ではない。この点は、実践的三段論法(practical reasoning)の形式に即して考えてみるとわかりやすい。(デイヴィドソン的に捉えられた)実践的三段論法とは、1)ある行為の主たる理由のうち、賛成的態度を大前提とし、2)信念を小前提とし、3)それらから、当該行為が行為者にとって「望ましい(desirable)」という結論が導き出される、という形の推論である。すなわち、図式化して書けば、次のようになる[3]。

実践的三段論法

大前提 (主たる理由のうちの) 賛成的態度の定式化:

行為者Sは、あるタイプの事態Pに対して、賛成的態度を持つ。

小前提 (主たる理由のうちの)信念の定式化:

Sは、この個別的なQする行為は、Pを実現するための一手段であると信じている。

結論 当該行為の望ましさ

ゆえに、この個別的なQする行為は、Sにとって望ましい。

直観的に言って、この型の推論事例が与えられ、そのとき大前提と小前提がともに真であるならば、結論もまた真であること(つまり、この推論図式が妥当であること)は、ほとんど明らかだと思われる。それだけでなく、実際に我々は、自己の意思決定においても、他者の行為の評価においても、こうした推論を当たり前に行っていると考えられる。このように我々が実践三段論法を重要な仕方で適用しており、そのような実践三段論法において主たる理由が中心的構成要素として機能していること、まさにこの事実こそが、行為の合理化は主たる理由を特定することで果たされるというデイヴィドソンの考えに大きな説得力を与えていると言えるだろう。

そこで、先に提起した問いに戻ろう。実践三段論法は、一体どのような論 理的言語を用いて定式化されるのが適切だろうか。ごく単純に考えて、まず は、様々の態度動詞に対応する様相演算子と、価値評価・選好(preference) を表すための様相演算子が必要となるだろう。なぜなら、大前提には賛成的 態度を表現する態度動詞、小前提には「信じる」という態度動詞、そして結 論には「望ましい」という評価語が、それぞれ登場するからである。とはい え、態度動詞や評価語に対応する様相演算子を含んだ論理の研究は、いくつ か存在しているが、ようやく緒に就いたばかりであり、現状でその本格的な 応用を行うには不十分のように思える[4]。実際、デイヴィドソン自身は、行 為文の論理形式の分析において、単なる一階古典述語論理を用いる(ただし そこには、出来事たちの集合の上を走る個体変項の導入という、興味ある拡 張が導入されてはいるが)にとどまり、いま言及したような様相論理的な演 算子の導入といったことには一切関心を示してない。このことには、クワイ ン以来の様相論理批判を受け継ぎ、論理学上の外延主義を取ろうとする彼の 哲学的立場が働いているであろうが、しかしその点を措いても、彼の態度に は無理からぬものがあると言えるかもしれない。

しかしながら、以上のようにして、態度動詞や評価語の本格的な論理的取り扱いをペンディングすることに同意したとしても、実は依然として、次のような重要な問題が残されていることを見過ごしてはならない。すなわち、大前提における P (行為者の心的態度ではなく、その態度の対象・内容とな

っているタイプ的な事態) や小前提における Q(「この個別的な Q すること」という仕方で登場する、当該行為トークンの特徴づけとしての Q) を適正に定式化する段階で、時間に関わる概念に訴えることが必要となる。この意味で、やはりある種の様相論理的・時間論理的言語を採用することを考えなければならないからである。——だがその点に入る前に、こうした本稿のテーマの背景となっている先行研究・関連研究について簡単に確認しておこう。

(2) 研究の背景・先行研究

デイヴィドソンの合理化がどうして重要であるかについては以上で触れた通りである。本研究は、直接には一特に実践三段論法は--ルポワらの解釈[3] に依拠している。この後述べる仕方は実践三段論法のより適切な形式化を行うことを目的としている。

他方本研究の背景をなしている重要な要素は、以下に述べる通り、様相論理とハイブリッド論理である。様相論理はよく知られている通り、通常の論理に様相演算子□と◇を足したものであるが、特に本研究において重要なのは、以下で説明する通り、時間に関わる基本概念を様相論理で表現しうる、いわゆる時間論理である。時間論理の言語自体は、A.N.プライアーの展開した、過去についての存在量化子<P>、過去についての全称量化子[P]、未来についての存在量化子<F>、未来についての全称量化子[F]にあたる様相演算子が導入された言語である。それぞれの直観的な意味は、かつて一度(once)、過去ずっと(hitherto)、今後いつか(sometime)、未来永劫(henceforth)となる。しかしながら、本研究で一層重要になるのは、近年急速に開発されている、より進んだ表現力を持った時間論理の言語、すなわちハイブリッド論理と呼ばれる言語である[5]。本項では、様相論理とハイブリッド論理について、詳細なシンタクス・セマンティクスに立ち入る余裕はないが、何故それが主たる理由の定式化に需要な役割を果たし、哲学的な意義を持つのか、またどのような哲学的興味を持つのかについて事項で紹介したい。

(3) 筆者の主張

主たる理由を適切に記述するという観点では、通常の古典一階述語論理では明らかに不十分となるような具体例を考えてみよう。

例 1: コンサート鑑賞のため武道館に向かう太郎

現在の日時は2017年12月2日16時である。太郎は最近とても忙しく働いており、日付感覚が曖昧である。今、たまたま予定がなくなったため昼寝

をしていたところ、友人からの電話に起こされる。太郎が今から一時間後までに武道館に来れば、余ったチケットを用いてコンサートを観賞できると提案をされる。太郎は誘いに応じ、時計も見ずに武道館へ向かう。

太郎が今行っている武道館への移動は合理化可能である。しかし、この行 為の主たる理由を次のように定式化するのは適切ではない。

2017年12月2日17時に始まるコンサートの鑑賞をしたい。(賛成敵的態度)

この武道館への移動は 2017年 12月 2日 17 時から始まるコンサートの鑑賞を実現する手段であると信じている(信念)

何故、この定式化は適切ではないのか。それを具体的に追及するために、行為主である太郎の行為、知っていること、知らないことを列挙してみよう。

- ① 太郎は今(2017年12月2日16時)武道館に向かっている。
- ② 太郎は武道館に一時間後に着くことを知っている。
- ③ 太郎は一時間後に武道館であるコンサートが始まることを知っている。
- ④ 太郎はこのコンサートを鑑賞したがっている。
- ⑤ ただし、太郎はコンサートが始まる日時を知らず、今が何日の何時であるかを知らなかった。

ここから下記のようなことが考えられる。

賛成的態度の定式化に現れた日時を太郎自身は知らない。彼が知っていたのは今から一時間後にコンサートが始まるということであり、2017年12月1日17時にコンサートが始まるとは知らなかった。それでは、賛成的態度において、この当該のコンサートについて記述するためには、太郎の視点から一時間後に始まるコンサートと書くのが適切だろうか、それとも2017年12月2日17時に始まるコンサートと使用の脈絡から独立な表現(典型的には固有名)で書くのが適切だろうか。

この疑問に答える前に、別の簡単な例をもって考えてみよう。例えば、今登ろうとしている山が富士山であることを知らない人が、あの山に登りたいと富士山を指さした時、この人が登りたいのは富士山であり、彼の賛成的態度を「富士山に登りたい」と定式化しても問題はないのではないか、ということが考えられるだろう。

このように、賛成的態度を記述するにあたって、脈絡独立な表現で書いて 差し支えない事例は少なくない。しかしこの場合ではそうではない。何故な らこの場合は、太郎が行きたがっているコンサートが太郎自身によってどのように特定されているか(太郎がどのような情報状況に置かれているか)、が 重要になるからである。

このことを明確にするために、例1を少々変えて考えてみよう。

例 2:日付を勘違いする太郎

太郎は最近とても忙しく働いており、日付感覚が曖昧である。しかし太郎は 2017 年 12 月 9 日に、彼の好きなバンドのコンサートがあるので行きたいと思っていた。たまたま予定がなくなり昼寝をしていたところ、友人からの電話に起こされる。太郎が今から一時間後までに武道館に来れば、余ったチケットを用いてコンサートを観賞できると提案をされる。太郎は誘いに応じ、時計も見ずに武道館へ向かう。

ただし、2017年 12月9日に武道館でコンサートが行われるが、それは太郎が行きたかったコンサートではない。太郎は彼の行きたいコンサートが始まる日時を間違えており、本当は 2017年 12月2日 17時に始まるのであった。しかし、彼の友人は太郎の行きたかったコンサートとその日時を正確に把握しており、まさに 2017年 12月2日 16時に太郎を誘う電話をしていた。

この例において、もし太郎が時計を見て現在の日時を知ったならば、彼は友人の誘いを断り、コンサートに行かなかっただろう。ここでのコンサートに向かう太郎の行為の説明をしようとするのならば、あくまで今から一時間後のコンサートと特定したので行こうとしたのだと理由づけなければならないだろう。彼は誤って一週間後にコンサートがあると思い込んでいたので、日時を特定した賛成的態度を与えると、彼の行為の説明にはならず、実践三段論法が妥当なものにならないのである。ここで重要なことは、彼は賛成的態度の内容 P において、インプリシットな時点への参照が存在し、これ自体は当人の知識にも信念にも入っていないということである。現に彼が今ここで行ったことを説明するには、今から一時間後という捉え方が、彼の行動の説明にとって不可欠であり、そのように述べないと理由にならないのだ。

これを表すために(2)で述べた時間論理の拡張であるハイブリッド論理を使うことができる[6]。ハイブリッド論理と時間論理の言語に更に「時点命題変項」と呼ばれる自由変項 x,y,z...を付け加えた言語であり、x の直感的な意味は目下の時点は x だということである。つまり x はその x が使用されている時点や状況といったものを脈絡依存的に表す物に他ならない。さらに時点命題変項に適用される関数記号として()+1 を導入しよう。例えば x+1 は目

下xより一時間後であると解釈される。この道具立てを用いると よって、以下のような定式化の方がより適切であるだろう。

太郎は一時間後に始まるコンサートの観賞をしたい x<f>(太郎) (賛成的態度)

太郎はこの武道館への移動は一時間後に始まるコンサートの鑑賞を実現す る手段であると信じている(信念)

太郎にとってこの移動は望ましい

もちろん、一時間後に始まるコンサートも、2017年 12月 2日に始まるコンサートも、同じコンサートではあるが、後者のコンサートには太郎は行かないのである。

このように行為の合理化をおこなうならば、当該のコンサートを客観的に記すだけでは正確ではなくなってしまう。 賛成的態度や信念の内容が行為主と無縁なものになってしまい、行為の理由の説明にならなくなってしまうからである。行為の理由は、本質的に行為主の視点を込みにしないと、適切な理由を与えることができないのだ。内容を定める為には、行為主が持っている情報・観点を加味しなければならない。現在の行為主の状況との関わりが明らかなように--先の例で言うならば「x から一時間後」のように-主たる理由の内容を表現することが肝要である。

行為を行う際の我々の欲求内容や信念内容が、どれほど深く周囲の状況・環境・脈絡を前提し、それらに依存して形成されているかが、この分析から示されている。

(4) 今後の展望

以上で述べたように行為の説明を定式化するに当たって、ハイブリッド論理に代表される時間的様相論理の言語を用いるのは非常に有用である。また逆に、このような適切な論理を用いれば、一般に行為主に帰属される理由というものが、まさに一つの振る舞いを行為として成立させる本質的要件の一部であるこが、明瞭に見て取れるようになると言ってよいと考えられる。既に述べた通り、主たる理由において、行為の環境依存性・状況内在性が重要な仕方で表立ってくる。今後、これらをいっそう詳しく考察し、また、賛成的態度に現れる態度動詞についても概念的分析を施して、広範な諸行為の合理化について時間的様相的論理を用いた解明を行いたい。

注釈

- [1] 参考文献(9)(10)(11)参照
- [2] ①に収められている諸論稿、特に Essays(1)-(5)を参照
- [3] 参考文献(3)参照
- [4] A.S.Rao & M.P. Georgeff, "Decision Procedures for BDI Logics", Journal of Logic and Computation, 8(3), 1998, pp.293-344 等の先 行研究がある。
- [5] 参考文献④参照
- [6] ただし"next time"演算子を使う等の他のやり方もある。これは査読者から指摘を受けた。

(5) 参考文献

- ① Davidson, D. (1980). "Essays on Actions and Events." Oxford university press.
- ② Fagin, R. et al. (2003). "Reasoning About Knowledge." MIT Press.
- 3 LePore, E. et al. (ed.) (1985). "Action and Events—Perspective on the philosophy of Donald Davidson." Basil Blackwell.
- 4 Areces, C. "Hybrid Logics" in [5]
- ⑤ Blackburn, P. et al. (ed.) (2006). "Handbook of Modal Logic." Elsevier Science.
- 6 Prior, A. N. (1957). "Time and Modality." OUP Oxford.
- 7 Prior, A.N. (1967). "Past, Present, and Future." Oxford Univ Pr on Demand.
- ® Broome, J. (2013). "Rationality through Reasoning." Wiley-Blackwell.
- Bratman, E. M. (2013) "Intention, Plans, and Practical Reason."
 Center for the Study of Language and Inf.
- Wilson, George and Shpall, Samuel, "Action", The Stanford Encyclopedia of Philosophy (Winter 2016 Edition), Edward N. Zalta (ed.),
 URL = https://plato.stanford.edu/archives/win2016/entries/action/.
- Wallace, R. Jay, "Practical Reason", The Stanford Encyclopedia of Philosophy (Summer 2014 Edition), Edward N. Zalta (ed.), URL = https://plato.stanford.edu/archives/sum2014/entries/practical-reason/>.

(首都大学東京大学院. 日本学術振興会)

田中大海

Abstract

Kanovich, Okada and Terui proved a three layered representation, which states a semantic relationship between intuitionistic linear logic and classical linear logic. In addition, they also verified there exists a sound and faithful embedding between them in terms of the three layered structure, which is very similar to Gödel translation. Modifying their methods, we showed that the three layered representation and embedding theorems hold for other substructural logics. In this report, we will focus not on the technical detail of these results but on the possibility of additionally extending them in a framework of substructural logics.

1 はじめに

相意味論(phase semantics)は Girard [3] によって導入された、線形論理に対する健全かつ完全な代数的な意味論である。特定の意味論の考察によって、それに対応する統語論上の体系の様々な性質を示すという旧来的なアイデアが、依然として線形論理の文脈においても有効であるということが、90年代にかけて相意味論を扱った様々な論文において示されてきた。岡田 [6, 7]によるカット除去定理の証明、Lafont [5] および岡田-照井 [8] による有限モデル性(そしてそこから導かれる決定可能性)の証明などがその代表例である。これらの結果に鑑みるに、完全性が成り立つという点で、相意味論は整合意味論 (coherent semantics) などの他の線形論理の意味論に対し、少なくとも証明可能性に関わる現象の分析において優っている、といえる。例えば、岡田によるカット除去を含意した強完全性の証明は、ゲンツェン流のカット除去アルゴリズムを与えてはいないものの、高階の線形論理に対するカット除去定理を容易に示すことに成功している。

意味論的な動機に即した研究としては、これらの論文からはやや後の時期のものとして、Kanovich-岡田-照井 [4] が挙げられる。そこでは、任意の直観主義相空間が、ある古典相空間の部分空間になるという興味深い結果が示されている。Kanovich らはこの結果を用いて、いくつかの系を導いている。第一に、直観主義相空間に対する新しい閉包作用素の定義である。様々な閉包作用素の定義を与えることによって直観主義相空間を定義するという試みが、Abrusci [1] や岡田 [6, 7] によってなされてきたが、それらはいずれも二階の定義として与えられてきた¹。しかし、直観主義相空間が古典相空間の部分空間であるという性質から直ちに、一階の定義を与えることが可能となる。第二に、直

観主義線形論理 ILL がある特殊な直観主義相モデルのクラスに対して完全であることが示されている。このクラスのモデルは準古典的 (quasi-classical) 相モデルと呼ばれ、外見上は非常に古典相空間に近いものの、 \bot が一般にはその台集合の部分集合にならないという点で、古典相空間よりも一般的な構造を持っている。さらに第三の系として、ILL から古典線形論理 LL への、証明可能性に関する埋め込み定理を導くことができる。これ以前に、Schellinx [9] が、 \bot と 0 を言語から取り除いた直観主義線形論理は古典線形論理と完全に証明能力が一致するという結果を示していたが、Kanovich らの埋め込みは翻訳こそやや複雑なものの、線形論理の結合子および定数をすべてカバーした結果であるという点で Schellinx の結果を拡張させることに成功している。

我々は、Kanovich らの証明したこれらの定理および系と同様のものほぼす べてが、部分構造論理の広範囲において成り立つということを予想している。 さらに、我々はこれらの命題が一部の部分構造論理に対して成り立つことを 示すに至っている。ここでは紙面の制約上、それらに対して証明を与えるこ とはできないため、本研究を発展させた完成版において詳細な証明を与える ことにしたい。したがって我々はそれらの証明の詳細に対してではなく、そ れらが Kanovich らの挙げた結果とどういった点で対応しており、どういった 点で異なっているのか、ということの輪郭を描くことに本稿の目的を限定す る。特に注目すべき点は、直観主義的な部分構造論理から古典的な部分構造 論理間への埋め込みすべてが、直観主義論理から様相論理 S4 への埋め込みと してよく知られるゲーデル翻訳に酷似しているということである。Kanovich らの研究および今回得た結果によって、ゲーデル翻訳が部分構造論理という、 より包括的な視点の中で、一般性を持った現象として理解される可能性が示 唆されているのである。個別に部分構造論理の体系を考察するという段階を 経た今、我々は次の問いに対してアプローチする:三層表現定理やゲーデル 翻訳等の命題は部分構造論理という枠組みの中でどこまで一般化することが できるのか。

2 部分構造論理間の相意味論的関係と埋め込み定理

[4] では既に、**ILL** から **LL** への翻訳とゲーデル翻訳の類似性が指摘されている。実際、ある固定された命題変数 p_0 に対して、 $\&p_0$ を様相演算子 \Box と見なすことで、それはゲーデル翻訳の定義とほぼ一致する 2 。

ここでは特に、次の二つの体系、アフィン論理(affine logic)と縮約的線 形論理 (contractive linear logic) を例にとって、それらの相意味論的分析の、 埋め込み定理の証明への応用を概観する。以下が **ILL** の始式および推論規則 である:

$$\frac{A,B,\Gamma\vdash\Pi}{A\otimes B,\Gamma\vdash\Pi}\otimes l \quad \frac{\Gamma\vdash A}{\Gamma,\Delta\vdash A}\otimes B \otimes r \quad \frac{\Gamma\vdash A}{A\multimap B,\Gamma,\Delta\vdash\Pi}\circ l$$

$$\frac{A,B,\Gamma\vdash\Pi}{A\otimes B,\Gamma\vdash\Pi}\otimes l \quad \frac{\Gamma\vdash A}{\Gamma,\Delta\vdash A\otimes B}\otimes r \quad \frac{\Gamma\vdash A}{A\multimap B,\Gamma,\Delta\vdash\Pi}\multimap l$$

$$\frac{A,\Gamma\vdash B}{\Gamma\vdash A\multimap B}\multimap r \quad \frac{\Gamma\vdash \Delta}{1,\Gamma\vdash \Delta} 1l \quad \frac{1}{\vdash 1} 1r \quad \frac{1}{\bot\vdash}\bot l \quad \frac{\Gamma\vdash}{\Gamma\vdash \bot}\bot r$$

$$\frac{A,\Gamma\vdash \Delta}{A\oplus B,\Gamma\vdash \Delta}\otimes l \quad \frac{\Gamma\vdash A}{\Gamma\vdash A\oplus B}\oplus r_1 \quad \frac{\Gamma\vdash B}{\Gamma\vdash A\oplus B}\oplus r_2$$

$$\frac{A,\Gamma\vdash \Delta}{A\otimes B,\Gamma\vdash \Delta}\otimes l_1 \quad \frac{B,\Gamma\vdash \Delta}{A\otimes B,\Gamma\vdash \Delta}\otimes l_2 \quad \frac{\Gamma\vdash A}{\Gamma\vdash A\otimes B}\otimes r$$

$$\frac{A,\Gamma\vdash \Delta}{A\otimes B,\Gamma\vdash \Delta}\otimes l_1 \quad \frac{B,\Gamma\vdash \Delta}{A\otimes B,\Gamma\vdash \Delta}\otimes l_2 \quad \frac{\Gamma\vdash A}{\Gamma\vdash A\otimes B}\otimes r$$

$$\frac{A,\Gamma\vdash \Delta}{A\otimes B,\Gamma\vdash \Delta}\circ l$$

ここで、A,B は **ILL** の論理式を、 Γ,Δ,Π は論理式の多重集合をそれぞれ表す。また、 $\Gamma\equiv A_1,\ldots,A_n$ のとき、 $!\Gamma\equiv !A_1,\ldots,!A_n$ を意味し、交換規則は 省略して用いるものとする。直観主義アフィン論理 **IAL** は **ILL** に次の規則を認めることで定義される 3 。

$$\frac{\Gamma \vdash \Delta}{A, \Gamma \vdash \Delta} Wl$$

これとは対照的に、直観主義縮約的線形論理 ILLC は ILL に無制限の縮約を認めることで得られる。

$$\frac{A, A, \Gamma \vdash \Delta}{A \Gamma \vdash \Delta} Cl$$

次に IAL と ILLC それぞれが、それに対応する古典的な体系といかなる関係にあるのかを、意味論的立場から見ていきたい。我々は次の二つの定理が成り立つことを確かめた:

定理 **2.1** 任意の直観主義アフィン相空間 \mathcal{M} に対して、準古典アフィン相空間 \mathcal{M}_a と古典相空間 \mathcal{M}_a が存在し、次を満たす:

- 1. $\mathcal{M} \sqsubseteq \mathcal{M}_q \sqsubseteq \mathcal{M}_c$
- 2. M_q から M への相同型写像 h が存在する。

定理 2.2 任意の直観主義縮約的相空間 M に対して、準古典縮約的相空間 M_q と古典縮約的相空間 M_c が存在し、次を満たす:

- 1. $\mathcal{M} \sqsubseteq \mathcal{M}_q \sqsubseteq \mathcal{M}_c$
- 2. M_a から M への相同型写像 h が存在する。

ここで、 $\mathcal{M}_1 \sqsubseteq \mathcal{M}_2$ は \mathcal{M}_1 が \mathcal{M}_2 の部分空間であることを意味し、写像 h が \mathcal{M}_1 から \mathcal{M}_2 への相同型写像であるとは、h が \mathcal{M}_1 から \mathcal{M}_2 (\mathcal{M}_1 と \mathcal{M}_2 は それぞれ \mathcal{M}_1 と \mathcal{M}_2 の台集合を指す)への全単射であって、相空間上の演算を保存する写像であることを意味する。上記二種類の定理に関してそれぞれ、1 は三つの空間が部分空間関係に関して三層構造 (three layered structure) をなしていることを示している。2 は直観主義相空間と、それに対して構成される準古典相空間が、相空間として同じ構造を持つことを示している。特に 2 によって、特定の準古典相モデルのクラスに対して、対応する体系が完全であることが帰結する。1 と 2 を合わせて、アフィン相空間、または縮約的相空間に関する三層表現定理 (three layered representation theorem) と呼ぶ。いま述べたように、次の結果を系として得る:

系 2.3 IAL の任意の論理式 A に対して、**IAL** で A が証明可能。 \iff 任意 の準古典アフィン相モデルにおいて、A は充足される。

系 2.4 ILLC の任意の論理式 A に対して、**ILLC** で A が証明可能。 \iff 任意の準古典縮約的相モデルにおいて、A は充足される。

三層表現定理において特に注意すべき点は、アフィン相空間の三層構造が、通常の相空間や縮約的相空間の三層構造とは、完全に対応していないということである。それはアフィン相空間に対する三層表現定理において、古典相空間が一般にアフィンではないということから明らかである。上記の系のように、準古典相モデルのクラスに関する完全性定理の段階ではこの差異は問題とはならないものの、三層構造を使って我々が示した埋め込み定理においては、以下で示されるように対応関係に差異が生じる。線形論理の任意の論理式 A に対して、 A^{\bullet} を次のように帰納的に定める:

$$p^{\bullet} := p \& \mathbf{1} \qquad (B \multimap C)^{\bullet} := (B^{\bullet} \multimap C^{\bullet}) \& \mathbf{1}$$

$$(B \star C)^{\bullet} := B^{\bullet} \star C^{\bullet} \quad (\star \in \{ \otimes, \&, \oplus \}) \qquad (!B)^{\bullet} := !B^{\bullet}$$

$$\mathbf{1}^{\bullet} := \mathbf{1} \qquad \qquad \bot^{\bullet} := \bot \& \mathbf{1}$$

$$\mathbf{0}^{\bullet} := \mathbf{0}$$

このとき次が成り立つ:

系 2.5 IAL の任意の論理式 A に対して、A が **IAL** で証明可能である。 $\iff A^{\bullet}$ が **ILL** で証明可能である。 $\iff A^{\bullet}$ が **LL** で証明可能である。

系 2.6 ILLC の任意の論理式 A に対して、A が ILLC で証明可能である。 $\iff \varphi(p_0) \multimap A^\circ$ が古典縮約的線形論理 LLC で証明可能である。ここで、 p_0 はある固定された命題変数を示し、 $\varphi(p_0) \equiv !p_0 \otimes !(p_0 \otimes p_0 \multimap p_0)$ 。

ここで A° とは [4] で定義されていた翻訳とまったく同じものを指す。したがって上記の系のうち二つめが、Kanovich-岡田-照井の埋め込み定理の、縮約的線形論理におけるカウンターパートになっている。ここで、 $\varphi(p_0)$ の代わりに、 $\psi(p_0) \equiv (p_0 \& 1) \otimes ((p_0 \otimes p_0 \multimap p_0) \& 1)$ という論理式を用いることで、ILLC と LLC の乗法・加法的断片についての埋め込みを得る。すなわち、ILLC と LLC の乗法・加法的断片をそれぞれ、FLec および GLc とすると、次の命題が成り立つ:

命題 2.7 $\mathbf{FL_{ec}}$ の任意の論理式 A に対して、A が $\mathbf{FL_{ec}}$ で証明可能である。 $\iff \psi(p_0) \multimap A^\circ$ が $\mathbf{GL_c}$ で証明可能である。

この命題の $\psi(p_0)$ に $\varphi(p_0)$ を代入した命題は成り立たない。なぜならば、様相演算子! は $\mathbf{GL_c}$ の乗法・加法的断片の言語には含まれていないためである。次にアフィン論理に関する埋め込み定理の注意に入りたい 4 。第一に、系 2.5 では \mathbf{IAL} から古典アフィン論理への埋め込みではなく、 \mathbf{IAL} から \mathbf{LL} への埋め込みが主張されている。ここではその理由について考えたい。まず定理 2.1 を強めた次の命題 (†) が成り立つと仮定する。

- (†) 任意の直観主義アフィン相空間 \mathcal{M} に対して、準古典アフィン相空間 \mathcal{M}_q と古典アフィン相空間 \mathcal{M}_c が存在し、次の 1 と 2 を満たす:
 - 1. $\mathcal{M} \sqsubseteq \mathcal{M}_q \sqsubseteq \mathcal{M}_c$
 - 2. M_q から M への相同型写像 h が存在する。
- (†) と定理 2.1 の唯一の違いは、1 と 2 を満たすようなある古典相空間 M_c に対して、 M_c がアフィンであるかどうかという点にある。系 2.5 と 2.6 は相空間の三層構造を反映した結果であるから、もし (†) が成り立つとすれば、IAL から古典アフィン論理への埋め込みが系として得られると考えるのは自然である。しかしここでは、(†) が一般には成り立たないことを示す。次の補題は簡単に示すことができる。

補題 2.8 $\mathcal{M}_c = (M_c, \bot)$ を古典相空間、 $\mathcal{M} = (M, Cl)$ を $Cl(X) := X^{\bot\bot}$ であるような \mathcal{M}_c の部分空間とする。(すなわち \mathcal{M} は準古典相空間。)このとき、次の 1 と 2 は同値である:

1. # はアフィンである。

$2. \perp \cdot M \subseteq \perp_{\circ}$

任意の古典相空間は準古典相空間であるから、上記の (\dagger) が成り立つと仮定すると、補題 2.8 より $M_q=\perp^\perp=M_c$ が成り立つ。したがって、 M_q と M_c は同じ相空間となる。しかしこのとき、2 によって 1AL は古典アフィン相モデルに対して完全であることになり、矛盾する。したがって、定理 2.1 を強めた (\dagger) は成り立たない。第二に、この埋め込み定理では 1AL から 1LL への埋め込みが考えられている。この埋め込み関係に関しては、相意味論的な分析を経ずに、純粋に証明論的な議論によって示すことが可能である。しかしこの関係は、相空間上の関係としては明示的に現れていない。したがって相意味論的に本質的なのは、1AL から 1L への埋め込みであると考えられる。

3 今後の展望

本稿では紙面の都合上、ここで紹介した定理に対して証明を与えることを しなかった。そこで第一の展望として、これらすべての定理に対して証明を与 えた論文を完成させることを目標とする。またさらなる発展として、我々は 部分構造論理におけるゲーデル翻訳の一般化の可能性を考えている。ある体 系に対して成り立つ統語論上の性質を、より広い枠組みの中で捉えなおすと いう試みが近年盛んになってきており、そうした研究は部分構造論理におい て一層の発展を見せている。例えば照井[10]がそのひとつである。そこでは 個別の体系に対してカット除去の成否を検証するのではなく、部分構造論理 においてカット除去が成立するための意味論的条件が考察されている。部分 構造論理におけるこのようなひとつの趨勢のもとで、Kanovich らの挙げた結 果や、それをもとに我々が示した結果を捉えなおすと、三層表現定理やそこ から生じるいくつかの系が、部分構造論理においては、現在までに知られて いる以上に高い一般性を持っているという可能性が浮かび上がってくる。そ して、個別的ないくつかの事例においてそれらを検証するという我々のこれ までの研究によって、三層表現定理や埋め込み定理をさらに一般化した形で 述べるという試みが現実味を帯びてきている。興味深いことに、このような 動機のもとで部分構造論理において埋め込み定理を拡張しようという研究が 既に Galatos-小野 [2] によって行われている。そこではグリベンコの定理(二 重否定翻訳)が代数的にどのような性質として特徴づけられるかが考察され ており、[10] と同じく代数的手法が強力な役割を演じている。これらの先行 研究を端緒として、部分構造論理においてゲーデル翻訳を一般化するという 我々の試みは、代数モデルの考察によってこれまで以上に発展していくと考 えられる。

注

 1 例えば岡田は直観主義相空間として次のような定義を提案した:Mを可換モノイド、Dを次の条件を満たす M 上の部分集合族とする。

- 1. D は任意の \bigcap について閉じている。
- 2. 任意の $X \in \wp(M)$ 、 $Y \in D$ に対して、 $X \multimap Y \in D$ 。

このとき、任意の $X \in \wp(M)$ に対して、

$$Cl(X) := \bigcap \{Y \in D \mid X \subseteq Y\}_{\circ}$$

直観主義相空間がこのような二階の (そして非可述的な) 定義によって特徴づけられるのに対して、Girard によって最初に与えられた古典相空間の閉包作用素は一階の定義により特徴づけられる:M を可換モノイド、 \bot を M 上の固定された集合とする。このとき任意の $X \in \wp(M)$ に対して、

$$Cl(X) := X^{\perp \perp}$$
°

 2 ただし部分構造論理においては、直観主義論理や ${f S4}$ の場合と違って、結合子や定数に対して乗法的/加法的という区別がなされることに注意する必要がある。特に、 ${f 1}$ と ${f T}$ 、 ${f L}$ と ${f 0}$ などの定数は、直観主義論理や ${f S4}$ では区別されていない一方で、 ${f ILL}$ から ${f LL}$ への埋め込みにおいては、翻訳の定義がそれぞれ異なっている。

3本稿では以下の規則

$$\frac{\Gamma \vdash}{\Gamma \vdash A} Wr$$

は考慮していない。また、 \bot が IAL の言語に含まれていない場合には、IAL のシーケントの右辺には常に論理式が一つ含まれるため、この規則は定義されない。

4 ここでの議論は、部分空間、準古典相空間、アフィン相空間の定義を前提として行われる。部分空間および準古典相空間の定義については [4] を参照せよ。アフィン相空間の定義については [5, 8] を参照せよ。

文献

- V. M. Abrusci, Sequent calculus for intuitionistic linear propositional logic. *Mathematical Logic*, edited by P. P. Petkov, 1990, Plenum Press, 223–242.
- [2] N. Galatos and H. Ono, Glivenko theorems for substructural logics over FL. Journal of Symbolic Logic, 71(4), 2006, 1353–1384.
- [3] J. Y. Girard, Linear logic. Theoretical Computer Science, 50(1), 1987, 1–102.
- [4] M. Kanovich, M. Okada and K. Terui, Intuitionistic phase semantics is almost classical. *Mathematical Structures in Computer Science*, 16(1), 2006, 67–86.
- [5] Y. Lafont, The finite model property for various fragments of linear logic. *Journal of Symbolic Logic*, 62(4), 1997, 1202–1208.
- [6] M. Okada, Phase semantics for higher order completeness, cutelimination and normalization proofs. Special Issue on the Linear Logic 96, edited by J. Y. Girard, M. Okada, A. Scedrov, *Electronic Notes in Theoretical Computer Science*, 3, 1996.
- [7] M. Okada, A uniform semantic proof for cut-elimination and completeness of various first and higher order logics. *Theoretical Computer Sci*ence, 281, 2002, 471–498.
- [8] M. Okada and K. Terui, The finite model property for various fragments of intuitionistic linear logic. *Journal of Symbolic Logic*, 64(2), 1999, 790–802.
- [9] H. Schellinx, Some syntactical observations on linear logic. *Journal of Logic and Computation*, 1(4), 1991, 537–559.
- [10] K. Terui, Which structural rules admit cut elimination? an algebraic criterion. *Journal of Symbolic Logic*, 72(3), 2007, 738–754.

(慶應義塾大学)

Abstract

Do worlds other than this actual world – so called *possible worlds* – exist? If possible worlds exist, what metaphysical status do they have? That is to say, *metaphysically speaking as what* do possible worlds exist? David Lewis has famously presented the theory of *genuine modal realism*, which insists that possible worlds exist as *concrete* entities. This article circulates objections against genuine modal realism and presents a novel categorization based on three layers: *system*, *disappointment*, and *conversion*.

1 Introduction: metaphysical debates on possible worlds

Possible worlds (i.e., worlds including not only our actual one but also non-actual merely possible ones) appear in many areas of philosophy. Metaphysical natures of worlds per se attract philosophers (see [5]). In addition, philosophers have employed possible worlds as a powerful but handy analysis tool to give accounts of many kinds of modality, ubiquitous throughout philosophy (cf. [3]). Modality offers a more detailed description of truth; it describes how true something is rather than simply whether or not it is true. Modal notions (including possibility, necessity, counterfactuals, dispositions etc.) and reasoning on modal claims via modal logics (recall widespread semantics of Kripke [8] features possible worlds as its crucial components.) are apparent cases of possible world talk. Furthermore, non-metaphysical philosophers have also enjoyed possible world analysis or possible worlds talk for other kinds of modes, including epistemic (as Hintikka [7] initiated), temporal (e.g. transitions, process), and ethical issues.

Most philosophers seem to agree with the existence of possible worlds until we begin to give a definition. This trend is due to the combination of two dominating schemes; one employs possible worlds talk and the other holds Quine's [15]'s ontoloical commitment. The former is observed in how widely possible worlds are used in philosophical enterprises. The latter secures ontological existence from mentioning or being quantified in such possible worlds talk. Hence, the existence of possible worlds is no longer the central metaphyiscal problem. Rather, the current concern is a qualitative question about the ontological status of possible worlds.

The current paper gives direction towards a program of defending concrete modal realism. The next section digests concrete modal realism. The subsequent section presents a novel categorization of these attacks based on three layers: system, disappointment, and conversion. Upon this chart, I will pinpoint the problem of current discussion concerning genuine modal realism. In the appendix, I will sketch the new program of the concrete modal realism to come. That should be, for one aspect, generalization of genuine modal realism, which can and should be more general than what David Lewis presented. This generalization, however, is not done by compromising or weakening. Rather, it made possible to cast the hard vision of concrete modal realism by shrinking it to spatial entities.

2 Digest: concrete modal realism

Toward the *as what* question on possible worlds, there are two major options. One is *abstractionism*, a more popular, more modest, more intuitively believable position, which includes possible worlds as *abstract* entities. The other, less popular, less modest, and less believable one is *concretism*, which insists possible worlds exist as concrete objects [10]. Concretism is suggested and supported (almost solely) by David Lewis [9].

On top of their existence, Lewis further specified the ontological profile of possible worlds. A possible world, according to Lewisian concrete modal realism, is a maximal mereological sum of spatiotemporally (and causally) connected individuals. According to this clarification, our actual world then contains the farmost galaxies as well as the ancient Romans. This description entails, Lewis argues, the isolation among possible worlds. There cannot be any overlap between two possible worlds since they are characterized as entities spatiotemporally (hence causally) isolated from each other.

This view is, fairly speaking, unpopular (witness [17]). Lewis himself noticed that his courageous statement –whose list appears to violate a widely believed ontological virtue known as Ockham's razor, for containing talking donkies, and Wittgenstein's unborn daughter and too many merely possible entities– is hard to accept for most, meriting less a response than an *incredulous stare*. His strategy is to appeal its *theoretical utility*; concrete modal realism is *fruitful*. Theoretical benefit of expanding the list of concretes overweights, Lewis insists, the cost of incredulity.

What is the virtue? Firstly, as shown in the introduction, concrete modal

realism provides powerful but handy analysis of modality. However, this benefit is not solely promised for Lewisian gang but also modal realists camps in general. Two distinctive prises are prepared for concretism believers. The first one is reduction [17]. Supporters [11] argue that the only successful reduction is done by Lewis. Recall that possible worlds are employed to give an account of what modality is. However, possible worlds under abstractionists understanding need modal concepts to explain what possible worlds are. For instance, linguistic ersatz approach sees a possible world as a maximal consistent set of propositions. Once they move on to what is these propositions composing the set, it would end up with saying that a proposition which is possible. Such explanation relies on unanalyzed or primitive modal concept, which is exactly what we are supposed to give an account to.

The second benefit is *metaphysical economy*. Lewis' ontological system, at a first glance, seems to be further generous than others since it adds merely possible but not actual entities such as the dragon breathing fire, Wittgenstein's unborn daughter into the list of what concretely exists. Lewis argues that we should count the cardinality of kinds instead of simply and naively counting the cardinality of entities. Lewis' list has only one single category: everything exists is concrete. This is ontologically speaking huge advance compared to abstractionist's list which requires an *extra* category: concrete and abstracts. The similar argument holds for a counterargument to who approves possible worlds only as a way of speaking such as metaphor or useful fiction); Lewis (nor I) does not find any point of importing an extra kind of entities "entities of reality" and "entities only appears in the possible world story". Therefore, Lewis concludes his concrete version gains the same benefit out of more restricted metaphysical resource, which is a ontological reasonable deal.

3 The three layers over anti-concretism objections

Many objections have been made to Lewis' theory of modality. Contemporary discussions for and against Lewis' perspective have been executed in a piece by piece manner. Some critics picks up a particular problem of their (occaisionally unjustified and arbitrary) choice from Lewis' theory; others defend some objections (relatively easy ones) of their choice.

This might, preferably for Lewis' methodological preference (shown in [14,

Ch. 9], [13]), reflect that there is no single shot knocking down argument in philosophy. Nevertheless, this piecemeal situation harms another virtue of philosophical studies that blesses analysis of a theory as a whole, not particular segments of the theory. Moreover, being suffocated with each particular objection and its particular consequences might miss the broad picture of our strategy to defend concrete modal realism. This section presents three layers laying through these countless objections. The layers are: system, disappointment, and conversion.

3.1 System objections. The first kind of objection claims that Lewis' ontology is ill or unsound as a *system*. Let us call this line of disputes system objections since it attacks Lewis' belief or thought as if it were a formal system such as of logic. Once a formal system (e.g. classical propositional logic) is expressed in a formal syntactical manner as a set of propositions, we usually expect the set to satisfy two norms: consistency and maximality. Neither an inconsistent system, which contains contradictions nor a non-maximal system, which has some explanatory gap, is wanted.

The non-maximal failure hosts especially many famous charges. The most prominent one may be the *epistemic* objection. It disputes that Lewis does not or cannot explain how we get modal knowledge. How do we know, the objection questions, what is going on in such Lewisian multiverses isolated from us? A similar charge [18] aims at *irrelevance*. How do such Lewisian isolated multiverses, if any, have anything to do with modality? This irrelevance objection charges Lewis' shortening to give an account to how such concrete and isolated worlds contribute to our modal notions.

3.2 Disappointment objections. The second type of disagreements is based on *disappointment*. These are points where Lewis fails to achieve the goal he himself promised. As observed previously, Lewis advertises his modal realism with expectations of its theoretical benefits which he even calls *paradise* for philosophers. Objections of this kind demonstrate that it is unobtainable by believing Lewis.

<u>Fruitfulness?</u> The most general merit of Lewisian modal realism is its *fruitfulness*. No one refuses to admit the fruitfulness of possible worlds analysis. However, there is a huge space for arguing how much, if at all, possible worlds with Lewisian characteristics solely enable this practical purpose. It seems to suffice to observe two facts. Firstly, what possible world semantics

the majority enjoys is Kripke semantics [8], whose ontological characterization and commitment are very different from of Lewis. Secondly, some critics along this line provide a more creative counterargument. They (e.g. [2] [4]) try to build possible worlds semantics without relying on Lewisian ontological entities.

Reduction? More crucially, reduction was a distinctive deal for genuine modal realism. Lewis' is claimed to be the *only* option which succeed to eliminate primitive modality in explaining what is modality. Shalkowski [16] doubts that Lewis achieved this goal. Shalkowski points out that primitive modality still remains in Lewis' theory when considering *alethic* modality. Everything (appearing in his discourse, which exists due to Quinean ontological commitment) is possible in Lewisian framework. In other words, anything in Lewis' discourse should be *already modalized*. Lewis' argument then becomes *circular* i.e. *non-reductive* argument, which the whole program of concrete modal realism should avoid.

3.3 Conversion objections. The third and final variation is also the grumpiest. These are conversion objections, which attack at Lewis' strategy or methodology. Lewis' methodology (see [13]), they claim, cannot work for convincing non-Lewisians. The former two layers are not methodological in this sense. Rather, they simply complains that what Lewis describes is *not* the promised paradise. This "conversion" version, in contrast, criticizes the attitude itself. This attitude can be said to which merely describe one option among many, plausible perhaps but lacking the means (say, uniqueness of explanatory power) to convince the skeptical.

<u>Bad deal?</u> To begin with, genuine modal realism offers ontological shoppers a bad deal. Some cannot buy Lewis' theory due to its unacceptable consequences, which require them to revise wide range of their beliefs – not only metaphysics but also common sense, scientific knowledge, and epistemological system. To them, therefore, such huge workloads seem a bad ontological deal.

<u>Transcendental argument?</u> Moreover, Lewis does not provide sufficient arguments to discard the competition. Even those who admit that genuine modal realism delivers all the merits it promised have no reason to convert until they are convinced that no other tool could do the same job. Lewis,

from their perspective, demonstrates at most how his theory works as a possible (and hopefully plausible) account. The creative counterarguments such as [2] and [6], which suggest alternative theories which can do the same job (modeling modality) without the aid of possible worlds, follow this type of objection.

Begging the question? Another allegation is that parts of Lewisian ontology are only acceptable to those who are already Lewisians. This argument is here represented by Cameron [1] as the common fallacy of begging the question. Lewis only explains what would happen if we accept Lewis' assumptions. But whether or not we should take these assumption is the very question that we are supposed to be discussing.

Merely philosophical tastes? Finally, we can point out the many other unjustified preferences which ground Lewis' whole argument. We may have some preferences which we take as our common starting points or aims, which we have traditionally called *Moorean facts*. Conversion charges permit Lewis to set facts of his own choice. However, they question how he can (or whether he even intends to) convince(s) other philosophers with different starting points.

In total, the strategical issue is that Lewis's argument does not address everyone. To resist Lewis, they do not have to conquer Lewis. It is enough to declare that Lewis' theory does not provide sufficient reason for them to discard their current position.

4 Conclusion: usage of this categorization

What did I do by offering such an overview of objections? The majority, anti concrete modal realists, can make use of this layered list for further elaborating their counterarguments. However, this guide is more important for the minority, hard pressed to reply to so many objections and doubly challenged by concrete modal realism's widespread unpopularity. The merit of my chart is to help to cast a grand plan of effective defense strategy for concrete modal realists. We do not have to provide an ad hoc revision or interpretation to avert a particular issue. The new genuine modal realist to come does not provide a response to each piecemeal objection propped against genuine modal realism. Rather, she intends to suggest a "general" kind of modal realism as the next step.

Appendix: feature of concrete modal realism for everyone

The tragedy of genuine modal realism is that, due to its unpopularity, many assume genuine modal realism is exactly as Lewis explicates. The endless stream seems to force us to abandon concrete modal realism completely; we ought to not get pushed further downstream than is warranted. Granted, the three layers are strong enough to challenge a portion of Lewisian modal realism. Nevertheless, these layers do not and should not conclude the end of concrete modal realism in general (or *in genuine*). The possibility of a more general version of concrete modal realism which eliminates some problematic Lewisian constraints remains open.

The general modal realism should keep the heart of concretism, and so worlds (and their residing individuals) should all be *concrete*. To be concrete is, following the standard criterion, to be *spatio-temporal*, or to be placed in space(-time). This weakened and thin version only requires us to admit possible worlds as spatial entities while leaving open what properties space satisfies. Each philosopher may want to express properties or constraints over worlds and other ontological entities of their choice. This general concretism, for the time being, may host such disputes by providing a vast common ground or platform called *space*. They would express their stances and differences from others in terms of space. In fact, most common characterizations can be easily expressed through spatial terms. (Lewisian *isolated* universes, Kripke's *accessibility*, dimensions or areas for the "abstracts" and "concretes" distinction, etc.)

Our first step should be not to knock down but to tame attackers by offering the general version of concrete modal realism, which does not commit to some of Lewis' opinions (as [12]). Once we succeed in securing that thin but rich ground, we can start the next phase, where the original disputes may be conducted in the unified framework of a general concretism.

参考文献

- [1] Ross P. Cameron. Lewisian Realism: Methodology, Epistemology, and Circularity. *Synthese*, 156(1):143–159, 3 2007.
- [2] Charles S. Chihara. The Worlds of Possibility: Modal Realism and the Semantics of Modal Logic. Oxford University Press, 10 2001.

- [3] Louis deRosset. Possible Worlds I: Modal Realism. *Philosophy Compass*, 4(6):998–1008, 12 2009.
- [4] J. Divers. Possible-Worlds Semantics Without Possible Worlds: The Agnostic Approach. *Mind*, 115(458):187–226, 4 2006.
- [5] John Divers. Possible Worlds. Routledge, 2002.
- [6] John Divers. Agnosticism About Other Worlds: A New Antirealist Programme in Modality. *Philosophy and Phenomenological Research*, 69(3):660–685, 11 2004.
- [7] Jaakko Hintikka. Knowledge and Belief. Cornell University Press, Ithaca, 1962.
- [8] Saul A. Kripke. Semantical Considerations on Modal Logic. Acta Philosophica Fennica, 16:83–94, 1963.
- [9] David Lewis. On the Plurality of Worlds. Basil Blackwell, Oxford, 1986.
- [10] Christopher Menzel. Possible Worlds, 2013.
- [11] Richard B. Miller. Genuine Modal Realism: Still the only non-circular game in town. Australasian Journal of Philosophy, 71(2):159–160, 6 1993.
- [12] Richard B. Miller. Moderate modal realism. *Philosophia*, 28(1-4):3–38, 6 2001.
- [13] Daniel Nolan. Topics in the Philosophy of Possible Worlds. Routledge, 2011.
- [14] Daniel Patrick Nolan. *David Lewis*. McGill-Queen's University Press, 2005.
- [15] Willard V Quine. On What There Is. Review of Metaphysics, 2:21–39, 1948.
- [16] Scott A Shalkowski. The Ontological Ground of the Alethic Modality. The Philosophical Review, 103(4), 1994.
- [17] Theodore Sider. Reductive theories of modality. In Michael J. Loux and Dean W. Zimmerman, editors, The Oxford Handbook of Metaphysics, pages 180–208. 2003.

[18] Peter van Inwagen. Plantinga on Trans-World Identity. In Alvin Plantinga, pages 101–120. Springer Netherlands, Dordrecht, 1985.(Universiteit van Amsterdam)

グラウンディング概念の理解における実体の役割について On the role of substance in understanding the concept of grounding

後藤真理子

Abstract

This paper aims to counter the doubt of grounding discussed by Daly. Grounding is a noteworthy concept in recent metaphysics. However, there are doubts about grounding. Among them, I will focus on an objection to grounding by Daly and discuss it. First, I briefly outline the objection by Daly that grounding is unintelligible. Second, I accept Audi's idea that if we take into account substance, it may be possible to answer Daly's objections. Finally, I suggest that we may be able to answer the objection that grounding is unintelligible by incorporating the characterization of substance given by Lowe into Audi's idea.

(1) 研究テーマ

本稿は、グラウンディング概念に対するデイリーの批判に再反論する事を目的としている。具体的には、デイリーのグラウンディング概念は理解不可能であるとする批判に対し、実体概念を考えるならばデイリーの批判に返答可能であるかもしれないというアウディーのアイデアを出発点として、ロウの実体概念を取り入れる事でデイリーの批判を退ける可能性を示す。

(2) 研究の背景・先行研究

近年分析形而上学内部において、「アリストテレス的現代形而上学」と呼ばれる伝統的形而上学を復興させようと目論む動きが見られる。このアリストテレス的現代形而上学は、分析哲学における二つのメルクマールである言語主義と物理主義へのある種の反抗として受け取る事が可能である。特に後者への反抗は、「第一哲学としての形而上学」というアリストテレス的現代形而上学の特徴付けの一つと結び付けられる。クワイン的な物理主義は、一般に、物理学を「諸事象を包括的に扱うことができ、かつ極力単純であるような体系を構築しようとする学問」だと措定し、その上で物理学に「事実判断の最終的審判者」という特権的身分を与える1。このクワイン的な「自然化された」形而上学観とでも呼ばれるところのものでは、形而上学はその方法と目的に関して科学と連続的であるとされ、存在についての問いは「規格化された理論」であるところの物理学を適用することで答えられるべきであると考えら

れる。これに対してアリストテレス的現代形而上学では、形而上学を「第一哲学」として、言い換えるならば「世界に関するありとあらゆる哲学的・合理的な探求の中核をなし、開始点となる」学問として捉える。つまり、アリストテレス的分析形而上学者は、形而上学と諸科学との関連性を認めはするものの、形而上学を諸科学で用いられる科学的知識の前提となる「第一哲学」として位置づけようとするのである。

このクワイン的な方法と対立すると考えられるアリストテレス的形而上学者にとっての存在論の課題と方法をシャッファーは以下のように簡潔に纏める²。アリストテレス主義的な存在論の課題とは、何が何に基礎付けられているのかについて述べることであり、そしてアリストテレス主義的な存在論の方法論とは、グラウンディングのついての診断と共に何が基礎的であるかについての診断を展開することなのである、と。ここで重要となるのがグラウンディングである。グラウンディングは「形而上学的説明」として理解され得るものであり、概して事物の間の優先性を表現すると考えられている。グラウンディングという概念が目的としているところのものは、クワイン的な「何が存在するのか」という問いかけに基づく存在者のリストの作成ではなく、存在者間の優先性あるいは存在者間の依存関係に関する構造の解明であり、また「何がより基礎的な存在者であるのか」というアリストテレス的な問いかけへの解答なのである。

しかしながら、このグラウンディングという概念への批判は複数あげられ ているのだが、その中でも、デイリーはグラウンディングという概念それ自 体に懐疑を寄せている。デイリーは、グラウンディング論者の行う主張を以 下の三点にまとめている。すなわち、(1)「グラウンディング」という用語 は理解可能である、(2) その用語は原始的である、(3) その用語は有用であ る、という三点である³。(1)について、デイリーはより理解しやすい言葉 でグラウンディングを定義することによってそれを理解しようとするという 戦略をあげている。例えば、「[Fa]は[Gb]によって基礎付けられる」という命 題を、「[Fa]は[Gb]によって必要とされる」、または「[Fa]は[Gb]の持つ本質 的性質の構成要素である」、または「[Gb]は[Fa]にスーパーヴィーンする」と いうように定義するならば、グラウンディングという概念は理解可能である。 しかしながら、この戦略は主張(2)に抵触する。ある概念が原始的である とは、すなわちその概念が他の概念によって定義不可能であるという事を示 す。故に、主張(1)において採られた他の用語によってグラウンディング を定義するという戦略は採る事が出来ない。では、どのようにすればグラウ ンディングという概念を説明することが出来るのか。このような疑問を出発

点とし、デイリーはグラウンディングに関する批判を開始している。

デイリーによれば、グラウンディングの支持者は、(a) その論理的性質、(b) 他の用語との関連の提示、(c) 実例の提示という三点を示す事によってグラウンディングという概念の説明を試みる。しかし、以上の三点の説明は達成できていないとデイリーは主張している。本稿では(a) と(b) に焦点を当て、デイリーの論点を明確化したい。

まず、(a)について、デイリーは「説明する」という述語とグラウンディ ングとの比較を例として挙げ、グラウンディングの持つ論理的性質が、その 内容を確定しないと述べる。「[Fx]は[Gy]を説明する」という形式の文章に含 まれる「説明する」という用語の使用を考える時、「説明する」という語は先 に示した(1)から(3)の形式的特徴を持つとデイリーは述べる。しかし、 「説明する」と「グラウンディング」は、想定される外延とは異なっている。 例えば、因果関係を考える時、原因はその結果を基礎付けていないが、原因 は結果を説明していると述べる事が可能である。もし[Fa]が[Gb]を基礎付け るならば、その時[Gb]という命題は[Fa]という命題を伴うのだが、ある原因 がある結果を引き起こす場合、「ある原因が起こる」という文は、「ある結果 が起こる」という文を伴わない。これらの見解を採るならば、原因はその結 果を基礎付けないという事になるが、原因は結果を説明しているので、「説明 する」と「グラウンディング」という語の外延は異なるという事になる。以 上のように、デイリーは「説明する」という語との比較を通して、グラウン ディングの論理的性質がどのようなものであるのかをよりクリアにし、その 結果グラウンディングの持つ論理的性質はその内容を確定しないという結論 を下す。

またデイリーは(b)他の用語との関連の提示についても批判を行う。グラウンディングという用語を理解する一つの方法は、既に理解されている用語を用いてその分析的関係を辿ることであるとしたうえで、デイリーはグラウンディングと他の用語との関連を例示する。ここでグラウンディングに関連する語である「より高い実在性の程度(a greater degree of reality)」の定義が挙げられている。より高い実在性の程度という語は、以下のように定義される。

xはyよりも高い実在性の程度を持つ=df xはyを基礎付ける

この定義によれば、事実[Fa]が事実[Gb]よりも高い実在性の程度を有するかどうかを知る唯一の方法とは、[Fa]が[Gb]を基礎付けるかどうかを確立する

事であるとされているように思われる。しかし、この定義には問題があるとデイリーは主張する。この定義は「より高い実在性の程度を持つ」という述語をグラウンディングによって定義している。それ故、まずグラウンディングを理解することによって、我々に「より高い実在性の程度を持つ」という語への理解が与えられる事になってしまう。このように、グラウンディングとそれに関連する語との関係性を提示することでグラウンディング概念の説明を行おうという実際の試みは失敗しているとデイリーは示している。グラウンディングとそれに関連する語との関係性を提示したとしても、それを理解するためにはグラウンディングへの理解が必要不可欠なのである。

以上の点を挙げ、デイリーは論文中でグラウンディング概念そのものを 我々は理解できないのではないかという批判を行っている。本稿ではデイリ ーの批判に対して、アウディーのアイデアを援用しつつ、ロウの『形而上学 の可能性』にて提示されている実体概念に基づいて再反論を試みる。

(3) 筆者の主張

前節でみたデイリーの批判について、アウディーは実体概念を取り入れる事によって解決が可能であるかもしれないという事を示唆している4。アウディーは、デイリーの(b)他の用語との関連の提示を有望なアプローチであると述べ、実在性の程度という観点からの説明は注目に値するとする。アウディーによれば、実在性の程度というアプローチは、実体と関連させる事でデイリーの疑問に解答しうるアプローチであるとされる。つまり、実体についての先だった理解によって、我々は先に見た実在性の程度の定義に出てくるグラウンディング概念を理解する事が可能となる。例えばすべてのものが最終的に実体に依存しているというアリストテレスの思想とグラウンディングが結びついていると考える時、グラウンディングはこのヒエラルキー構造における依存関係を示す概念であるという説明が可能となりうるとアウディーは述べるのである。

このアウディーのアイデアに則るならば、実体を導入する事でグラウンディング概念の説明が可能になる。しかし、ここで一つの疑問が生じる。ではその実体とはどのような概念であるのか。実体概念が具体的にどのような内実を持つのかという事を明らかにしなければ、このアウディーのアイデアを額面通りに受け入れる事は難しい。何故ならば、実体についての先立った理解がなければ、実在性の程度の定義に現れたグラウンディング概念を理解する事が出来ないからである。さて、アリストテレスにとって実体概念とは何であるのかという事について、「本質」「形相」「現実態」といった複数の解答が考えられうる5。すなわち、アリストテレスの哲学においても実体の定義は

一義的にはなされておらず、様々な解釈が可能なのである。

現代において、実体を基礎的な存在者であるとするアリストテレス的現代 形而上学者のうちの一人がロウであろう。ロウは実体が何らかの意味での「独 立性」を持っているという点で、他のカテゴリーに対する特権的な位置にあ るという事を認め、それ故に実体を基礎的存在者であると見なす6。では彼は 実体をどのように定義するのか。以下では、『形而上学の可能性』および加地 (2008)を参照する事で、その概略を確認する。

ロウはまず、実体を暫定的に以下のように定義する。「xが実体であるのは、 xが個体であり、かつ、xと同定されず x がその存在を y に依存しているよ うな個体 y が存在しない時かつその時に限る。」この実体の定義において問 題となるのは存在論的依存とはどのようなものであるのかという事である。 その点について、ロウは存在論的依存を「x はその存在を y に依存している =df 必然的に、x の同一性は y の同一性に依存している」と定義している。 以上の実体の暫定的定義に存在論的依存の定義を加えたものが、実体のより 正確な定義となる。すなわち、「xが実体であるのは、xが個体であり、かつ、 xと同定されず x がその同一性を y の同一性に依存しているような個体 y が 存在しない時かつその時に限る。」さて、この定義に関してもまた、より子細 に分析されなくてはならない点がある。それは同一性への依存である。Xの 同一性がyの同一性に依存するとは果たしてどういう事であるのか。ロウは この点を明確にするために、「もしxの同一性がyの同一性に依存している ならば、必然的に、以下のような関数 F がある。つまり、x は必然的に y の Fと同定される。」という関数を導入する。これはどういうことか。ロウは結 婚という例を挙げて、この事を説明している。結婚の同一性は結婚している 二人の人間に依存する。それ故、もしxが結婚であり、yとzが問題におけ る二人の人間であれば、上記の関数の定義は、xがyとzとの結婚と必然的 に同一であるという事実のおかげで満たされる。この場合、必要な関数は「z との結婚」という事になる。しかし、注意しなくてはならないのは、これは 同一性の依存についての必要条件であっても十分条件ではないという事であ る。この点について加地は以下のように簡潔に纏めている。「例えば、xの同 一性は x を唯一の要素とする集合(シングルトン) {x} の同一性に依存しない にもかかわらず、「必然的に、x は、シングルトン {x} の唯一の要素と同一 である」という命題が成立してしまうからであるで。」このような場合を考え る時、少なくともこの関数の導入による同一性の依存の定義は不十分である。 故にロウは同一性の依存について、本質を導入する事で解決しようと試みる。

ロウは、有望な戦略として、{x}の唯一の要素であるという事が x の「本質」

の一部ではないという理由で、除外する事を提案する。この考えに従えば、yのFであるxの本質の一部ではないような関数Fを除外することができるとロウは述べ、以下の新たな定義を示す8。

 \mathbf{x} の同一性は \mathbf{y} の同一性に依存している= \mathbf{df} 必然的に、 \mathbf{x} が \mathbf{y} の \mathbf{F} であるという事が、 \mathbf{x} の本質の部分であるような関数 \mathbf{F} がある。

さて、この定義における関数Fは、xの属する種の同一性基準によって供給 されるものだと考えられると加地は述べる9。同一性条件とは、ある対象 a がある対象bと同一であるということを決定する条件であるのだが、ロウは、 一般的な同一性条件が「xと y が条件 $C_{\mathfrak{o}}$ を満たす時そしてその時に限り、 もしx とy が同じ種φであるならば、x はy と同定される」と述べる10。 こ こで言及されている条件 C。は、種φによって異なるものとなる。例えば、 もし個別者 a が彫像であり、個別者 b がブロンズ原子のメレオロジカルな 和であるとするならば、その時個別者 a は個別者 b と同定されることがで きない。何故ならば、彫像という種とブロンズ原子の和という種は異なる種 であるために、それぞれの同一性条件もまた異なるからである。このような 同一性基準によって関数Fは与えられる。加地によれば、例えば集合の場合 は、外延性の公理(Axiom of Extensionality)がその同一性基準を規定してい る。故に、a,b,cという三つの要素からなる集合 S は、(a,b,c)のみ)を要素 とするもの」であるということがその本質の一部であることになる。この事 から、集合 S は要素 a,b,c にその同一性を依存しているので実体とは言えな いという事になる。対してもし a,b,c それぞれが例えば生物的個体や人物で あるとすれば、それらは実体であることになるのである。ロウは個体が属す る種の本質がその個体の同一性基準を与えるとした上で、その本質を他の個 体に依存しないことをもって実体の条件としている、と加地は結論付ける。

以上の点を考慮する時、アウディーの示唆は以下のように解釈される事が可能であろう。実在性の程度というアプローチについて、実在性の最も高い実体概念と関連させる事でデイリーの疑問に解答しうるアプローチであるとアウディーは述べる。この時、実在性の最も高い程度を持つ実体はアリストテレスの述べるところの「それだけで存在しうる」ものでなければならないだろう。この事について、ロウの解釈するところの実体概念は、本質によって他の個体に依存しないという実体の条件を与えている点でアドバンテージがあるように思われる。ロウの考えに則るならば、実体はそれの持つ本質に関して他の個体に依存せずに存在するものである。故に、「それだけで存在し

うる」ものであり、ヒエラルキーの最も下に存在することのできるようなものとして実体を捉えている。アウディーはヒエラルキーの最下層に措定できるようなアリストテレス的な実体を考える時のみ、グラウンディングをそのヒエラルキーにおける形式的な依存関係を示すものとして説明可能であるとしている。ロウの実体についての理解をアウディーの提案に組み込むとき、グラウンディングについての理解は可能となる。すなわち、実体を基礎としたヒエラルキーを考えるならば、グラウンディングとはヒエラルキー構造内の依存関係であるという説明が可能となり、グラウンディングをグラウンディングによって定義する事になってしまうというデイリーの批判点は解決され得るのである。故に、実体を取り入れる事により、デイリーのグラウンディングは理解不可能であるという批判に応答可能となる。

しかしながら、本質に関して他の個別者に依存していない存在者こそが実体であるというロウの考えについて批判を行う事は可能である。実体と本質とが深く関係していると考える点に関して、そのような批判は、よりアリストテレス的なものであろう。この点において、アリストテレス的な探究を行う事を目的とするグラウンディングとロウの実体概念とは噛み合いやすいであるように思われる。だが、アリストテレスが実体について「それだけで存在しうる」と述べているのに対して、本質によって実体が個別化されるというロウの実体についての考え方は、実体が本質にある種依存しているというロウの実体についだろうか。この事に関しては、アリストラレス自身が『形而上学』 Z 巻において「実体とは何か」という問いの答えらて、本質である」という解答をしているという点が重要なヒントとなりるかもしれない。実体それ自身が本質であるならば、上記の問題点に関する解答の糸口が見出されるように思われるのである。

(4) 今後の展望

現代形而上学の内部においてグラウンディングに関して多々論じられている中、実体についての言及はしばしば見られる。しかしながら、グラウンディングについて論じる文脈の中で、実体がそもそもどのようなものであるのかという点について論じられる事は比較的少ないように見受けられる。実体について言及される時、その実体がどのようなものであるのかという事に目を向ける事で、グラウンディングに関する更なる深い理解が可能になるように思われる。今後の展望としては、以下の二点が挙げられる。まず、グラウンディング関係のみに目を向けるのではなく、グラウンディングという形而上学的説明の中で実体がどのような内実を持ち、またどのような役割を負っ

ているのかという事を明らかにしていく事、第二に実体に目を向ける事でグラウンディングについて更なる分析を行い、アリストテレス的なメタ存在論の輪郭を更にはっきりとさせた上で、現在支配的なクワイン的なメタ存在論とアリストテレス的なメタ存在論との対立について論じていく事である。

(5) 参考文献

- アリストテレス (出隆 (訳)).『形而上学』,アリストテレス全集第一巻,岩波書店
- Audi, P, 2012. "A clarification and defense of the notion of grounding", in Correia, F, Schnieder, B (2012), pp. 101-121.
- Chalmers, D.J., Manley, D, Wasserman, R (Eds.) 2009.

 metametaphysics: new essays on the foundations of ontology. Oxford:
 Oxford University Press.
- Correia, F, Schnieder, B (Eds.) 2012. *Metaphysical Grounding: Understanding the structure of reality*. Cambridge: Cambridge

 University Press
- Daly, C.2012. "Scepticism about grounding", in Correia, F, Schnieder, B (2012), pp. 81-100.
- 井頭昌彦. 2005.「クワインの物理主義と自然化された認識論」.『科学哲学』 38-2:pp. 109-122.
- 岩田圭一, 2015. 『アリストテレスの存在論 〈実体〉とは何か』. 早稲田大学 出版部
- 加地大介,2008.「現代的実体主義の諸相一一実体の独立性をめぐって」.『哲学の探究』35.pp. 37-49.
- Lowe, E.J. 1998. The Possibility of Metaphysics: Substance, Identity, and Time. Oxford: Oxford University Press.
- Lowe, E.J. 2009. More Kinds of Being: A Further Study of Individuation, Identity, and the Logic of Sortal Terms. Oxford; Wiley Blackwell
- Schaffer, J. 2009."On What Grounds What", in Chalmers, D.J., Manley, D, Wasserman, R (2009), pp. 347-383.

(九州大学)

註

- 1 井頭(2005), p. 110.
- ² Schaffer(2009), p. 351.
- ³ Daly(2012), p. 82.
- ⁴ Audi(2012), pp. 118-119.

- ⁵ 岩田(2015), p. i.
- 6 加地(2008), p. 37.
- ⁷ Ibid, p. 44.
- 8 Lowe(1998), p. 149.
- 9 加地(2008), p.44.
- ¹⁰ Lowe(2009), p. 16.

ダメット意味理論再考 —実践的能力としての言語理解を考える—

三上温湯

Abstract

This paper deals with Michael Dummett's theory of meaning. The theory is known as a theory of understanding, that is, the aim of the theory is to give an account of 'what it is that someone knows the language'. He requires to specify what we are taking as constituting a manifestation of understanding of the language, but this requirement has been taken to be impossible demand by many researchers. However, it is not clear what he consider as the understanding. In this paper we conclude that the understanding consist of a series of practices, which he calls the manifestations of our understandings. From this point of view, we suggest two approaches to specifying the the manifestations of our understandings.

1 研究テーマ

ダメット的な意味の理論とは何か

M・ダメットが、「意味の理論(theory of meaning)」と呼ばれる独自の哲 学的理論のプログラムを提唱したことはよく知られている。この彼のプログ ラムは、一方で、あまりにも理想主義的であるとか、ラディカルすぎるといっ た批判を受けてきたし、また現在に至るまで、文字通りのプログラムにとど まり続けており、ほとんど具体的な実現に至っていないという弱点を抱えて もいる。だがそれにもかかわらず、このプログラムはきわめて多くの現代哲 学者たちの関心を刺戟し、彼らの考えに重大な影響を及ぼしてきたと言って よいだろう。では、ダメットの考えるこの意味の理論とはどのようなもので あり、さらにそもそも、彼がこのような理論を構築しようとした目的はどこ 一見意外であるが、こうした疑問を検討してみると にあったのだろうか。 じきに気づかれるのは、まさにダメットの考えの独自の根底性ゆえに、本来 あまりにも基本的であるはずのこれらの疑問に答えることが、実は決して簡 単ではないということである。例えば、意味の理論とはどのような理論であ ろうか。次項で一度、一般に、意味の理論がどのように探求されるものであ るかに目を向け、ダメットの意味理論の特質をどのように捉えるべきか、考 察することとしよう。

2 研究の背景・先行研究

意味の理論のアイデア自体は、もちろんダメットに固有のものではなく、とりわけ近年では一つの共有概念となっており、現代の代表的な哲学事典(Stanford

Encyclopedia of Philosophy)ではこの語が独立した項目として立てられ、二つに細分化されて説明されている(https://plato.stanford.edu/entries/meaning/)。すなわち、意味の理論とは、(1) 第一には、ある言語の各表現について、それがどのような事柄を意味するか(表現するか)を特定しようとする、言語学的・形式意味論的な部門(「意味論的理論 semantic theory」と呼ばれる)であり、(2) 第二には、一般にある行為主体がある言語に習熟しそれを実践的に使用しうるのはいかにしてかを究明する、認知科学的・コミュニケーション論的な部門(「基礎づけ的意味論 foundational theory of meaning」)である。こうした意味の理論の試みがある中で、ダメットの考えはどこに位置づけられるであろうか。

(1) の意味論的理論が行うことは、より詳しく言えば、およそ次のようなも のだと考えられる。すなわち、ある言語 L にどのような表現タイプ(文結合 子、固有名、普通名詞、等々)がどのような結合可能性を伴って用意されてい るかを明らかにし、各表現タイプについて、その意味と見なしうる存在者の タイプは何であるかを特定した上で、各語に対し、まさにその意味(いわゆ る「意味論的値 semantic value」)であるものとして、対応するタイプ中の一 定の存在者を割り当てる、ということである。こう述べてみれば明らかであ るように、ここでの意味論的理論とは、基本的には、フレーゲ(言うまでも なく、ダメットの意味理論の考えに最も大きな影響を与えたのはフレーゲで ある)に端を発する「Bedeutung (意味、指示)の理論」そのもの、あるいは その現代的な展開形態であり1、したがってそれが、ダメットの考える意味の 理論と重なるものであること、より適切には、ダメットが自らの意味の理論 の一部を成すべきだとしていたものに他ならないことがわかる。だが同時に、 これもダメット自身が強調した通り、こうした「Bedeutung の理論」は、彼 の意味の理論にとっては「単なる一部」にすぎず、もっとはっきり言えば、彼 の意味の理論にとって不可欠ではあっても、最も枢要な部分なのではない。2 では、そうした最も枢要な部門とはどのようなものだろうか。それはまさに、 (2) の基礎づけ的な意味理論なのだろうか。おそらくある程度までそう言って よいと思われる。というのも、ダメットにとって意味の理論とは第一義的に は、ある言語の習熟話者が持つ理解内容を明らかにする理論、もっと踏み込 んで言えば、ある人がこの理論を習得し、理解していることが、直ちにその 人を当該言語の習熟話者とさせるような、そうした理論だからである。実際 彼は、What is a Theory of Meaning(I) において以下のように強調している。

意味の理論は、理解の理論である。すなわち、意味の理論が説明を 与えねばならないものは、ある人がある言語を知っているとき、つ まりその人が、当該言語の諸表現や諸文の意味を知っているときに、 知っていることとは何か、ということである。[WTM1, p.3]

しかし、ここで言われている「意味の理解」とは、より精確にはどのようなことだろうか。再び(2)の基礎づけ的意味理論の考えに戻ってみよう。基礎づけ的意味理論が行おうとするのは、基本的には、文字通り、認知科学的色彩の強い企て、すなわち、ある言語の習熟話者が、当該言語について標準的に保持している認知的内容—習熟話者が当該言語について標準的に知っている(あるいは少なくとも、信じている)事柄、すなわち例えば、そうした話者が記憶しているある語の定義(別の語による言い換え)、あるいはさらに、ある語に結びつけているイメージ、プロトタイプといったもの等—を明らかにすることであり、あるいはまたコミュニケーション理論的な企て、すなわち、習熟話者が言語を共有する他者との間で相互理解を果たすためにどのような技量を駆使するか、といった事柄を明らかにすることである(あるいは少なくとも、それ主要な部分として含む)と考えられる。では、ダメットの意味の理論もまたこうしたものだろうか。

3 筆者の主張

実践的能力としての概念と、その習得の説明という課題

確かに、ある程度まではその通りである。実際ダメットは、フレーゲから、 先に (1) との関係で述べた「Bedeutung の理論」のみならず、「Sinn (意義) の理論」を受け継いでおり、その際ダメットは、フレーゲにおける Sinn が、 まさに認知的なものであることを再三強調している。だがそれにもかかわら ず、ダメットの意味の理論を(2)の基礎づけ的意味理論と一般的に同一視して しまうことは、決定的に不十分だと考えられる。なぜなら、ダメットが問題 にする「習熟話者が獲得している意味理解」とは、話者の心的領域に蓄えら れた言語についての信念や知識自体でも、あるいはまた、言語を介して他者 の意思を察知したり自己の意思を告知したりできるコミュニケーション・ス キル自体なのでもなく、何よりもまず、当の語を適切に使用・適用し、そのこ とを通じて一定の行為(そこに含まれる目的)を達成することのできる実践 能力そのものだからである。例えば、「ならば」という接続詞(文結合子)の 意味の理解とはどのようなことだろうか。それはダメットによれば、何か辞 書的な定義を提示できるとか、「p ならば q」の真理表(古典論理であろうと なかろうと)を書いて見せることができるといったことではない(それらを 含んでもよいが、本質ではない)。そうではなく、ある話者が「ならば」の意 味を理解しているとは、一般にその話者が、仮定pからの帰結qの導出が確 立済みとなっている状況下に置かれたとき、そこからさらに仮定 p を撤回して「p ならば q」を結論することができる(「ならば」の除去則)という、そうした実践(この場合は推論実践)を行う能力(ability)を持つこと、そのような態勢(disposition)を身につけていること、に他ならない 3 。

「徹底した意味理論」という考えのポイント

こうした実践的能力・態勢としての意味理解ということを、おそらく (2) の基礎づけ的意味理論もまた、ある程度まで探究主題としていよう。だが、ダメットの意味の理論においては、こうした実践的能力・態勢としての意味理解そのものの記述・分析が中心的関心事であるため、一貫して次のような課題が重要な考察主題となる。その課題とは、一般に、ある語の意味を理解し、その語の適用能力を身につけた話者は、そのことによってまさに何を実践しうるようになるのか、言い換えれば、当該の意味理解(より一般的に言えば、そうした意味理解を介して達成される概念把握)により、いまや当該話者にとってどのような〈実践的可能性の空間〉が開かれるに至っているのかを特定する、という課題である。(2) の基礎づけ的意味理論にとっては、こうした課題は、その本来の射程を超え出る過大な要求となるだろう。まさにこの点で、それはダメットの意味理論と大きく異なると言えると思われる。

さて、いま述べた課題、意味理解がどのような実践的可能性の空間を切り開くかを特定するという課題に取り組む意味理論とは、ダメット自身の用語で言えば、(単なる「つつましい意味理論」と対比される)「徹底した意味理論」に他ならないと言えよう。と言うのも、徹底した意味論とは、一般に人のいかなる振る舞い(行為)が、ある語についてのその人の意味理解の表出、顕現(manifestation)として認められてよいかを定めようとするものであり、これは言い換えれば、まさに当該の意味理解(を介した概念把握)が、それを達成した者に対して何を可能とさせるかを特定することだと考えられるからである4。実際ダメットは、意味の理論が明らかにすべき内容を以下のように述べている。

我々が、何らかの実践能力を、命題的知識の形で表現することに関心がある場合、そしてとりわけ、その実践能力がまさに言語を習得しているということである場合、以下のことが我々に義務としてのしかかってくる。すなわち、もし我々の提起する考えが、説明力を持つものであるならば、[1] その人が、当の能力を持つために、知らなかればならないこと [内容] が何であるかを特定するだけではなく、 [2] そうした知識を持つということどのようなことであるのか、すなわち、われわれがどのようなことを、そうした命題たち

についての知識の顕現を構成しているとみなしているのか、ということをも特定せねばならいのである。もしこのことをし損なったなら、その理論的表現と、その理論が表現しようとするところの実践的能力とを結びつけることができないだろう。⁵ [WTM1, p.21]

冒頭でも触れたように、従来、このように意味理解の顕現化を特定せねばならないというダメットの徹底した意味理論の要求は過大なものであり、実現不能なものと目されてきた。6 実際、ダメットが実践能力ということで、具体的にはどのようなことを念頭に置いていたのかということも不明瞭であり、その眼目が十分明らかでない不可解なものとして扱われてきたように思われる。だが、ここまで述べてきたように、ダメットの意味理論が、言語習得を介して開かれた行為の可能性を特定することにあると理解するならば、意味理解の顕現化を特定することは、決して何か特殊で実現不能な要求ではなく、当然明らかにされるべき、ダメットの意味の理論の中核を成すものだと見るべきであることが理解されるのではないだろうか。

4 今後の展望

彼の理論をより適切に具体化するにあたってさしあたり取り組むべき問題は以下の二点であると考えられる。すなわち [1] 一般にある語に関する理解や知識の規準として、一体どれほどの言語使用を行えることが要求されるかという問題と、[2] 主体が行う振る舞い(言語使用)や、主体が持つそうした振る舞いの能力・傾向性を、そこで問題となっている語の意味理解の顕現化だと認めてよいのはどのような場合か、という問題である。

[理解のエビデンスの過度に厳格な規準をリベラル化するための論理の探究] 第一の問題について、ダメットがとっている立場は、非常に、あるいは過度に厳格なものであるように思える。この厳格さは、ダメットが、対象言語の分析に際して採用されるべき論理が、そしてその理解内容を記述するためのメタ言語(そこで採用される論理)も、一般的な意味の理論(真理条件意味論)で用いられる古典論理ではなく、ブラウワー以来の直観主義者たちが開発した直観主義論理であるべきだと主張している

ということに起因していると考えられる。ところで、この直観主義論理を採用する理由とは、(i) 第一に、我々人間的主体が持ちうる言語理解(それを介した世界についての知識と信念)は、本来的に局所的・未規定的・生成的なものでしかありえず、大局的・状況超越的な二値原理・排中律に立脚する古典論理によっては決して適切に記述しえないからであり、(ii) 第二に、他方でひ

とたび我々が適切に確立した理解や知識は、古典論理が許容するような決定 不能性や非構成性を決して含んではならず、徹頭徹尾構成的な正当化によっ て裏づけを与えうる純正なものであるはずだからである。だがこの考えのう ち、(i) については多くの哲学者が賛成するが、(ii) はある種の極端な認識礎 づけ主義としておそらく疑問視されることが多い考えであり、実際、ここで は詳論の余裕がないが、ある人がある語の意味理解を持つと言えるための規 準は、ダメットの考えに従うと厳格になりすぎてしまうという事実を指摘す ることができる。実は、情報の哲学の近年の展開が明らかにしつつある点の 一つとして、確かに直観主義論理は古典論理的な大局性・超越性を拒否しは するが、その本質は、まさに (ii) のような認識論的基礎づけ主義の理想を具 現することにあると見なすのが適当であり、他方で、(i) のような人間主体の 活動の局所性・未規定性・生成といったものを適切に扱うための論理は、そ れとは異なる多様な情報論理の諸体系のうちに見出されるものだということ である。8このような状況を踏まえ、情報論理の諸体系、特に幾何学的論理の 技術的改良と哲学的基礎の解明に努め、適正な仕方でリベラルである理解規 準を提案することが、現代において、ダメットの意味理論の着想の啓発性を 明らかにし、さらなる具体化を図る有効な試みになると考えられるのである。

[理解の顕現化についての懐疑論を退けるための、行為の合理化の構造の再分析 - reason 論の展開]

第二の問題は、言い換えれば、ある主体がある振る舞いをし、あるいは振 る舞いの能力を有しているとき、その振る舞いや能力が、まさに問題となっ ている語の意味(概念)の理解を具現していると我々が結論してよいのはど のような場合か、という問題である。この問題が困難であるのは、一つには、 すでに触れたウィトゲンシュタインの規則遵守のパラドクスなどにおけるよ うに、その振る舞いや能力に基づいて自然にある語の意味理解を帰属してよ いように思える主体が、実はその後の振る舞いによってそうした理解を持って いなかったことが露見する、といった懐疑論的問題があるからでもあるが(こ の点の精確な解明も、この研究の目的の一部としたい)、より重要で興味深い と考えられるのは、この問題がもっと一般的な合意の合理化(rationalization of action) にまつわる困難を示唆していると考えられるからである。行為の 合理化とは、やはりデイヴィドソンが指摘した、我々が日常的に互いの間で 行う重要な営みであり、すなわち、ある主体がある振る舞いを行った際に、そ の振る舞いの理由(reason)を特定すること、より詳しく言えば、そこでの 目的および手段―デイヴィウドソンにより忠実に言えば、(目的に対する) 賛 成的態度と、(目的実現のための手段に関する)信念のペアという、いわゆる

「主たる理由」―を帰属する、という営みに他ならない。この行為の合理化につきまとう一つの難問は、賛成的態度と信念の組み合わせが適切でさえあれば(つまり、当の信念に従う限り、確かにその行為が当の目的の実現手段だと見なしうる限り)、例えばその目的がどれほど荒唐無稽なものであっても合理化が行えてしまう、ということである。こうした困難を避けるためには、行為の合理化においては、行為が目的実現の手段となっている(と主体が信じている)だけでは不適切であり、ダメット的な観点から言えば、更に進んで、主体が当該の目的や手段に関する適切な概念理解を備えていることを前提せねばならないはずである。この意味で、行為の合理化においても概念理解の顕現化の問題は重要であり、まさにその前提となるものであるが、では顕現化を認めるための規準をどう詳細に与えるべきかは、ダメットによってほとんど論じられていない。

われわれが言語について理解している内容を、そしてその意味理解の顕現化とはどのようなものであるかを明らかにしようとするダメットの意味理論は、以上のような問題をその射程に含んでいる。これらの問題に対し、一定の見通しを与えることが、ダメット的な意味の理論の、とりわけその徹底した意味理論としてのあり方を具体的に実現する方策となっていると考えられる。本稿では、ダメットのいう意味の理解とはいかなることであるかという問題に力点を置きながら、実は彼の意味の理論が言語習得、とりわけ、意味理解を通じた概念の獲得という極めて一般的な重要性を持つ哲学的問題に取り組むものであり、しかもその際、概念の獲得がそれを果たした行為主体にとってどのような実践的可能性の空間を切り開くかを明らかにしようとする点で、確かにダメット自身によっては十分に展開されなかったとはいえ依然としてそこには更なる発展をもたらすべく我々が努力する価値が大きく認められることを見てきた。今後こうした問題をさらに続けて検討したい。

注

¹ フレーゲにどれほどまでに意味論値の概念があったかは議論の余地あるところであるが、少なくともダメットの解釈によれば、フレーゲの Bedeutung は現代における意味論値の概念の萌芽であることがわかる [OAP, p.53]

 2 ダメットは意味理論の構築の際、フレーゲの指示、意義、力の概念を採用しており、度々、指示の理論だけでは不十分であることを述べている。[WTM2, p.40,pp.84ff][OAP, pp.15ff]

³ここで詳論できないが、実際のところダメットは、検証という概念に依

拠して結合子の意味の説明を行う。ここで取り扱った例は、論理定項の意味をその導入則で定める検証主義的な意味論における標準的な結合子の説明に習った。ここではとりわけ、「ならば」の意味が、ダメットの言う意味理解ということを捉える上で重要であると考え、例としてあげたが、ダメット自身は'or' などについて、検証主義的な意味の説明を与えている。こうした見方は[WTM2, p.40-41][1990] その他多くの箇所で述べられており、とりわけ [LBM]では、検証主義的意味理論の基盤について総括的な説明がなされている。

4 徹底した意味理論とは、一般的には、主体にいかなる言語知識も前提とせずに、その主体がその理論を学べば、当の言語の習熟話者になることができるような理論であるが、ここでは、こうした理論の構築は、まさに言語習得を介して可能になる行為の特定を行うことで果たされると考えている。徹底した意味理論とはどのようなものと捉えられるべきか、ここでは詳論できないが、議論の余地あるところである。この要求についての批判や、徹底した意味理論についてのこれとは異なる解釈が多く存在するのでここではそのいくつかを挙げておく。[Mcdowell1987][Mcdowell1997],[Gaifman1996][金子 2006]

- 5本文中の番号、[]内の語は筆者による補足である。
- 6代表的なものは注5で挙げたものである。

⁷ 本稿では、直観主義論理の、局所性や、未確定性が、言語理解の本質的特性であるという側面を中心に据えて解釈したが、査読者からは、検証超越的な真理概念は本来的に理解不可能なものであるという、いわゆる意味論的反実在論から、直観主義論理の採用が動機づけられるべきだという指摘を受けた。本稿の考えと、この指摘が必ずしも対立するとは思わないが、ダメット自身の議論の流れも踏まえ、今後詳細な検討を行いたい。

⁸中でも、最も重要なのは、圏論の中心分野の一つであるトポス論で開発されてきた、いわゆる「幾何学的論理」であると考えられる。この点はさらなる調査と議論が必要であるが、近年の、トポス論における傾向性概念等に関わる成果 [Vickers1993]、計算機科学における様相概念の再検討 [Pfenning and Davies 2000] がこうした議論の裏付けになると考えられる。

文献

[FPL] Dummett,M(1973)."Frege: Philosophy of language" first published in 1973, second edition 1981:Duckworth

[TOE] Dummett,M(1978). "Truth and Other Enigmas":Harvard University Press.(藤田晋吾訳『真理という謎』:勁草書房 1986 年)

- [LBM] Dummett, M. (1991). "The Logical Basis of Metaphysics", Cambridge, Massachusetts: Harvard University Press.
- [SOL] Dummett,M(1993). "The Seas of Language":Duckworth
- [OAP] Dummett,M(1993). "Origins of Analytical Philosophy":Harvard University Press.(野本和幸ほか訳『分析哲学の起源』: 勁草書房 1998 年)
- [1973] Dummett.M(1973). .'The Philosophical Basis of Intuitionistic Logic'"Truth and Other Enigmas"
- [WTM1] Dummett,M(1974). 'What is a theory of meaning?(I)',Samuel Guttenplan.ed.Mind and Language:Oxford University press.reprinted in Dummett 1993a,1-33
- [WTM2] Dummett,M(1976). 'What is a theory of meaning?(II)', Gareth Evans and John McDowell.ed. Truth and Meaning:Oxford University press.reprinted in Dummett 1993a,34-93
- [1978] Dummett,M(1978). 'What do I Know when I Know a Language',Stockholm University .reprinted in Dummett 1993a,94-105
- [1990] Dummett,M(1990) 'Language and Truth'.reprinted in Dummett 1993a,p.117-146
- [Frege1884] Frege.G(1884) 'Die Grundlagen der Arithmetik'(野本和幸訳、「算術の基礎」, 野本和幸・土屋俊編『フレーゲ著作集2』所収:勁草書房 2001 年)
- [Frege1892] Frege.G(1892) Über Sinn und Bedeutung'(土屋俊訳、「意義と意味について」, 黒田亘・野本和幸編『フレーゲ著作集4』所収:勁草書房1999年)
- [Davidson1967] Davidson,D(1967)"Truth and Meaning". Reprinted in his Inquiries into Truth and Interpretation,2nd. ed. (2001), 1742. Oxford: Oxford University Press.
- [Mcdowell1987] Mcdowell, J (1987)" In Defence of Modesty". Reprinted in his Meaning, Knowledge, and Reality, 87107. Cambridge, Mass.: Harvard University Press (1998)
- [Mcdowell1997] Mcdowell, J (1997). "Another Plea for Modesty". Reprinted in his Meaning, Knowledge, and Reality, 108131. Cambridge, Mass.: Harvard University Press (1998).

[Vickers1993] Steven Vickers(1993)'Geometric Logic in Computer Science'.

[Gaifman1996] Gaifman.H(1996). 'Is 'Bottom-Up' Approach from the Theory of Meaning to Metaphysics Possible?'. Journal of Philosophy 93(8), 373407.

[Pfenning and Davies 2000] Frank Pfenning and Rowan Davies (2000)'A judgemental reconstruction of modal logic'

[金子 2006] 金子洋之 (2006)[『ダメットにたどりつくまで』: 勁草書房 (首都大学東京大学院)