

新進研究者 Research Notes

日本科学哲学会
科学基礎論学会

第6号 (2023年)

佐々木 健人	現象的意識とアクセス意識	1
坪井 祥吾	芸術の定義, グラウンド, 非循環性	10
久保埜 雄大	不完備アウェアネスゲームにおける均衡の論理的分析	19
谷川 綜太郎	理解についての内在主義と外在主義	28
岡部 幹伸	知覚のきめ細かさからの論証に対して概念主義を擁護する	35
太田 泰幹	判断と情動はどのように対立し、また調停されるのか	44
池原 優斗	アクターネットワーク理論の「人工主体」研究への適用——画像生成AIの分析を例に	53
水谷 亮介	総称文の表出主義的意味論	62
井上 朋彦	ヒルベルト前期形式主義の方法論	70
高山 馨	合理的なアクラシアはいかにして可能か	80

JAPANESE STUDENT RESEARCH NOTES OF PHILOSOPHY OF SCIENCE

Philosophy of Science Society, Japan

Japan Association for Philosophy of Science

No. 6

Contents

SASAKI Kento	Phenomenal Consciousness and Access Consciousness ·····	1
TSUBOI Shogo	Definition of Art, Ground, Non-Circularity ·····	10
KUBONO Yudai	Logical Analysis of an Equilibrium in Game with Incomplete Awareness ·····	19
TANIGAWA Sotaro	A critical study of internalist accounts of understanding ·····	28
OKABE Mikinobu	Defending Conceptualism against the Argument from Fineness of Grain of Perception ·····	35
OHTA Taiki	How Judgments and Emotions Conflict and are Mediated ·····	44
IKEHARA Yuto	Application of actor-network theory to the study of "artificial agents"--Analyzing AI image generators as an example ·····	53
MIZUTANI Ryosuke	An Expressivist Semantics for Generics ·····	62
INOUE Tomohiko	Hilbert's Methodology in his first period of formalism ·····	70
TAKAYAMA Kaoru	How Is Rational Akrasia Possible? ·····	80

現象的意識とアクセス意識

Phenomenal Consciousness and Access Consciousness

佐々木健人

Abstract

Ned Block classified the concept of consciousness into “phenomenal consciousness” and “access consciousness”. He claimed that the two types of consciousness are conceptually different and empirically distinct. Block also argued that Sperling’s experimental work supported his latter claim. However, it is ambiguous whether Sperling’s experiment supports this claim and if Block’s claim can be subjected to scientific validation. In this study, I will outline the controversy related to his claim and draw attention to the two distinct questions that are confused in the controversy, each of which, in my opinion, can be addressed in a different manner.

(1) 研究テーマ

本稿で扱う意識とは、私たちが何かを経験するときの主観的で質的なある感じであり、しばしば「クオリア」とも呼ばれる（例えばリンゴを見たときの独特な赤い感じ）。このような意識に関して様々な哲学的・科学的議論が存在するが、本稿で取り上げるのは Block の主張とそれにまつわる論争である。

Block (1995) は意識概念を「現象的意識」と「アクセス意識」に分類した上で、一部の意識研究はこれら二つの意識概念を混同してしまっていると指摘した（これが「概念的問題」である）。加えて彼は、このような二つの意識が概念的に分離されるだけでなく、経験的にも分離されうると主張した（これが「経験的問題」である）(Block 2005, 2011)。しかし、そのような経験的分離が本当にありうるのかどうかに関しては論争がある。本稿ではそのような論争を取り上げた上で、Block たちの論争が上記の概念的問題と経験的問題の区別を蔑ろにしてしまっていることを指摘する。

(2) 研究の背景・先行研究

Block (1995) の言う現象的意識とは、私たちが何かを感じる時の経験そのものであり、Nagel (1974) の言う「～であるとはどういうことか (what is it like to be ~?)」で定義されるような意識概念である (Block 1995, p. 230)。

リンゴを見たときに私たちが持つ独特な赤い感じはその典型例である。

これに対して Block の言うアクセス意識とは、推論、行為制御、報告などといった脳機能への利用可能性によって定義される意識概念である (Block 1995, p. 231)。すなわち、ある表象が脳内の機能的システムにとって利用可能であるとき、言い換えると、私たちがその表象に機能的にアクセスできるとき、その表象はアクセス意識的であるとされるのである。例えば、“目の前のリンゴは赤い”という報告（これはまさに機能的アクセスによって実現される）ができるとき、そのリンゴの赤さの表象はアクセス意識的であるとされる。つまり Block は、アクセス意識を、利用可能性という機能的性質を持つもの（もしくは性質そのもの）として定義しているのである。

このように二つの意識概念を区別した上で、Block (1995) は次のように主張する。すなわち、意識に対する既存の説明の一部は二つの意識概念を混同してしまっており、現象的意識を説明すると言いながら、アクセス意識についてしか説明を行っていない、と。

Block は 1995 年の論文では、上記のような概念的問題を指摘することに主に注力していた。しかし 2005 年の論文以降は、二つの意識にまつわる経験的問題に焦点を移し、アクセス意識と現象的意識が経験的に分離するという主張を一層強く行うようになる⁽¹⁾。このような経験的分離の存在を示唆すると思われる知見の一つが Sperling (1960) の実験結果である。以下、この実験の概要と、この実験を利用した Block の主張について概観しよう。

Sperling (1960) は計七つの実験を行っているが、それらは大きく二つに分類される。すなわち、被験者に対して全報告 (whole report) を求めるものと、部分報告 (partial report) を求めるものである。以下では佐藤 (2014) に倣って前者を「全報告パラダイム」、後者を「部分報告パラダイム」と呼ぶ。

全報告パラダイム実験において被験者は、文字 (アルファベットと数字) の配列 (例えば 3×3、3×4 などの配列) を 0.05 秒間提示され、その後、見えた文字をできる限り報告するように求められた。結果として被験者は、文字の配列に関わらず「提示された刺激のうち平均約 4.5 文字までしか報告することができなかった」(Sperling 1960, p. 6)⁽²⁾。

次に部分報告パラダイム実験においては、全報告パラダイム実験と同様の刺激が提示された直後、音によって報告する行が指示された (例えば、高い音ならば上の行、低い音ならば下の行というように) (Sperling 1960, p. 6)。この実験において報告を求められる文字は 4 文字以下であったが、結果として被験者は指示された行の文字をかなり正確に (しばしばその行の文字全てを正しく) 報告した (Sperling 1960, pp. 6-8)。

また、これら二つの実験結果以外に考慮されるべき要素として被験者の訴えがある。上記の実験で被験者は、結果としては3~4文字しか報告できなかったにもかかわらず、自分は報告できた以上の数の文字を見ることができたと訴えた (Sperling 1960, p. 26)。Sperling の実験を解釈するにあたっては、このような訴えもまた被説明項として考慮されなければならない。

以上を踏まえると、Sperling の実験を解釈するために考慮されるべき要素は大まかに以下の三つとなるだろう。

- (A) 被験者は全報告パラダイムにおいては配列全体の内3~4文字しか報告できなかった。
- (B) 被験者は部分報告パラダイムにおいては指示されたのがどの行であろうともほぼ正確に報告できた。
- (C) 被験者は配列全体に対する報告しえない経験を持っていたと感じた。

Block は以上の要素を、自身の主張を支持するものとして解釈する (Block 2011, p. 567)。その主張とは、現象的意識（もしくは知覚的意識）は機能的にアクセスしきれない程に豊かなのだというものである。彼の解釈によると、Sperling の実験において被験者は、配列全体に対する現象的に意識的な表象（これはそれぞれの文字を識別できる程度には豊かである）を持っているため、(C) のように感じ、(B) のように報告できた。しかし、機能的アクセスには容量制限があるため、全報告を求めると (A) のような結果が出たのである。

このような、現象的意識が機能的アクセスに収まらないほど豊かであるという事態を、Block (2011) は現象的意識の「オーバーフロー」と呼ぶ。このようなオーバーフローが実際に起きているのだとしたら、現象的意識とアクセス意識は異なる神経基盤⁽³⁾を持っているということになり、したがって両者は経験的に分離するということになるだろう。具体的には彼は、現象的意識の神経基盤を各感覚皮質での神経発火に、アクセス意識の神経基盤を前頭葉と頭頂葉を含む広域的神経発火に見出す (Block 2005, pp. 46-48)。

以上の Block の主張は妥当だろうか。Carruthers に倣うと、Block の主張は以下の二つのステップにまとめられる (Carruthers 2017, p. 66)。

- ステップ (1) 経験的証拠 (Sperling の実験結果) により、現象的意識が機能的アクセスをオーバーフローするというテーゼが擁護される。
- ステップ (2) ステップ (1) より、現象的意識とアクセス意識は脳内で別々に実現されているということが導かれる。

これが、Carruthersにより再構成されたBlockの「オーバーフロー論証」である。以下では、この論証に対する批判を、上記の二つのステップに対応させながら紹介していきたい。

まず紹介するのは、ステップ(1)に対するCarruthers(2017)による批判である。彼は手始めに、注意というものが限られたリソースであること、そして、注意にはある表象を意識的にする機能(つまり、その表象を実現している神経活動を強化する機能)があることを前提する⁽⁴⁾。その上で彼は、二つのパラダイム実験を以下のように解釈する。

一方で全報告パラダイム実験においては、刺激がある間は、画面上の全ての文字の表象がワーキングメモリに十全に保たれており、したがって全ての文字が現象的意識的かつアクセス意識的に知覚されている。このとき被験者の意識的表象は、刺激そのものによって支えられている。しかし刺激が無くなった後は、その表象を保つために大量の注意リソースが必要とされるため、被験者は全ての文字の表象を保つことができない。つまり、被験者は全ての文字に対する現象的意識的かつアクセス意識的な表象を持っていたが、それら表象は画面が消えた後にワーキングメモリから抜け落ちてしまったのである(したがって(A)のような結果が出る)(Carruthers 2017, pp. 68-69)。他方で部分報告パラダイム実験では、音による指示が注意リソースの割り方を修正する役割を果たしたため、被験者は(B)のように報告できたのである(Carruthers 2017, p. 69)。

このような実験解釈を踏まえてCarruthersは、Blockの主張する機能的アクセスに収まらないほど豊かな現象的意識は、実は機能的アクセスを受けていた可能性がある⁽⁵⁾と論じる。彼のこの批判の目的は明確だろう。それはすなわち、Sperlingの実験の新しい解釈を提案することによって、Block流の解釈の正当性を失わせることである。この批判の是非については、佐藤による分析に触れる箇所(本稿の5ページ19行目)で論じよう。

次に、オーバーフロー論証のステップ(2)に対する批判、すなわち、現象的意識とアクセス意識が脳内で別々に実現されるという主張に対する批判を見ていこう。その批判とは、機能的アクセスなしの現象的意識は反証不可能であり、科学的探究の外にあるというものである(Cohen & Dennet 2011)。

例えば赤いリンゴを見ている被験者について考えてみよう。Blockに従うならば、被験者がたとえ(脳の神経接続が遮断されたことなどによって)その赤さを報告できず、その赤の色合いが本来引き起こす筈の興奮感情を持たず、さらにはその感情によって引き起こされる筈の動機づけを失った状態に

あるとしても、Block の言う現象的意識の神経基盤（つまり感覚皮質）が活性化している限り、その被験者は確かに現象的意識を持っているということが成り立つ⁽⁶⁾。なぜなら、報告、感情、動機付けは全て機能的アクセスによって実現されるからである（Cohen & Dennet 2011, pp. 361-362）。

しかし、このような現象的意識の存在は一体どのようにして科学的に確かめられるのだろうか。すなわち、Block の言う現象的意識の神経基盤が、確かに現象的意識を実現していること（もしくは、まさに現象的意識そのものであること）をどうやって確かめるのだろうか。Cohen & Dennet (2011) は、このような現象的意識の存在は科学的には反証されえず、したがって科学的探究の範囲外にあると考える。その上で彼らは次のように主張する。すなわち、アクセス意識から分離されるような現象的意識というのは、意識に対する科学的探究を妨げるようなものであるため、私たちはそもそもこのような意識概念を受け入れるべきではないのだ、と（Cohen & Dennet 2011, p. 363）。彼らの主張を受け入れるならば、現象的意識とアクセス意識が脳内で別々に成り立つという Block の主張を受け入れる必要はない（なぜなら、そのような概念的区別をそもそもすべきではないから）ということになる。

しかし佐藤（2014）の分析に基づいて考えると、以上のような批判はオーバーフロー論証を論駁するには至らない。以下、佐藤の分析をまとめよう。

佐藤は、Sperling の実験の新しい解釈を提案するような批判（先ほど取り上げた Carruthers の批判もその一つである）はオーバーフロー論証を論駁するには至らないと考える。というのも、Block による実験解釈とその批判者による実験解釈は、それらが三人称的に観察可能なデータ（つまり(A)~(C)）を十全に説明している限りで実験解釈として等価であり、どちらかが優れているということはないからである（佐藤 2014, p. 98）。

加えて佐藤は、アクセス意識から分離されるような現象的意識は科学的探究の範囲外にあるという批判もまた、オーバーフロー論証を論駁するには至らないと述べる。なぜなら、Block は決して科学的に確かめられないような主張をしているわけではないからである。Block もまた、ある刺激に注意が向いた時にその刺激が報告可能になるということは認めるのであり、したがって現象的意識の神経基盤に対して「弱い報告可能性」を帰属している（佐藤 2014, p. 98）。つまり Block は、現象的意識の神経基盤は機能的アクセスを実際に（そして必然的に）受けている必要はなく、それを受ける可能性さえあればよいと考えているのである。しかし佐藤の見解では、このような Block の考えは経験的知見に基づいたものではなく、むしろ“現象的意識は機能的アクセスをオーバーフローするほどに豊かである”という自身の仮説

の概念理解に基づいたものである（佐藤 2014, p. 99）。

(3) 筆者の主張

筆者の分析では、Block と Cohen & Dennet の間の論争は、意識の概念的
問題と経験的問題が十分に区別されておらず、それによって非常に複雑にな
っている。例えば Cohen & Dennet は、科学的反証不可能性という経験的問
題に基づき、Block 流の意識概念を批判している。対して Block は、現象的
意識に対する自身の概念理解に基づき、現象的意識の神経基盤に対して機能
的アクセス可能性を帰属している。このように両者は、意識の概念的問題と
経験的問題を行き来しながら議論しているのである。

本稿の(2)節で、1995 年の論文と 2005 年以降の論文の間での Block の主
張の変化に言及したが、このような変化から見て取れるように、そもそも
Block は二つの意識にまつわる概念的問題と経験的問題を別物として扱って
いた。しかし論争の中で二つの問題の境界線は曖昧になっており、それによ
って論争が複雑化しているのである。

筆者の主張は、彼らの論争を解決するためには、概念的問題と経験的問題
を明確に区別し、それぞれに対して異なるアプローチを行う必要があるとい
うものである。以下では、Block と Cohen & Dennet の対立を二つの問題に
分解すると同時に、それら問題を解決するための方法を探る。

まずは概念的問題に目を向けよう。すでに見たように、Block は現象的意
識をせいぜい機能的アクセス可能性さえあれば成り立つような概念として理
解している。しかし Cohen & Dennet の側に立つならば、むしろ現象的意識
は機能的アクセスを必然的に含むような概念として理解されるべきかもしれ
ない。これが彼らの論争における概念的問題である。

この問題に対する解決法の一つとして、機能的アクセスをいくつかに区別
した上で、それぞれについて現象的意識の概念理解に（可能的にもしくは必
然的に）含まれるか否か考えるという方法が可能かもしれない。例えば、報
告を実現するアクセス、感情を実現するアクセス、動機付けを実現するアク
セスを区別した上で、それらアクセスが現象的意識の概念理解に可能的に含
まれるかそれとも必然的に含まれるかを考えることができる。このように考
えると、色などの知覚に伴う感情（赤を見たときの興奮感情など）を実現す
る機能的アクセスは、現象的意識の概念理解に必然的に含まれると直観的
には思われる。というのも感情は、その色の現象的意識の独特さを決定する要
素の一つだからである。このように、現象的意識の概念理解に含まれる機能
的アクセスは、直観的な分析によってある程度絞られうる。

次に経験的問題に目を向けよう。Block は、感覚皮質の活性化のみを現象的意識の神経基盤であると考えているが、対して Cohen & Dennet は、機能的アクセスを実現するようなより広域的な神経発火が現象的意識の神経基盤であると考えている。つまり両者は、現象的意識の神経基盤に何が含まれるのかに関してもまた対立しているのである。このような経験的問題は、すでに見た概念的問題とは区別されるようなものである。

筆者の考えでは、この問題への対処は今後の科学的探究に一任されるべきである。例えば、被験者の報告を必要としない意識研究パラダイムが開発されたならば⁽⁷⁾、報告を実現する神経発火が現象的意識の神経基盤に含まれるか否かが解明されうるだろう。このような科学的探究によって、現象的意識の神経基盤がどこまでかという問題は少なくとも部分的には解決される。

Block たちの論争は、上記の二つの問題に沿って分解され、そしてそれぞれに解決が目指されうる。また、本稿が提示する解決方法は決して Block と Cohen & Dennet のどちらかを支持するものではない。この解決方法は、機能的アクセスを逃れるような現象的意識があるかもしれないという Block の警鐘を考慮した上で、最低限どの機能的アクセスが現象的意識に（概念的にも経験的にも）含まれるかを検討するという、中立的な解決方法である。

(4) 今後の展望

以上で提案した方法によって現象的意識の概念理解や神経基盤がある程度絞られたならば、Block たちの論争は一定の進展を見せるだろう。例えば、現象的意識の概念理解がいくつかの機能的アクセスを必然的に含むならば、Block が想定するような現象的意識とアクセス意識の二分法自体が崩れるため、Sperling の実験に対するオーバーフロー論証に則った既存の解釈は再考されることになる。また、ある機能的アクセスの神経基盤が現象的意識の神経基盤から分離されると明らかになったならば、機能的アクセスなしの現象的意識は科学的探究の範囲外であるという Cohen & Dennet の主張は（少なくとも部分的には）否定されることになるだろう。したがって、Block たちの論争を二つの問題に分解した上で解決を目指すことは、彼らの論争に何かしら進展をもたらすはずである。

注

1. 各論文での Block の主張については、太田&山口（2009）が詳細なサーベイを行っている。
2. なお、刺激の提示時間を 0.15~0.5 秒に変更しても、正しく報告できる

文字の数は変化しなかった (Sperling 1960, p. 6)。

3. この神経基盤という言葉の本稿は「神経相関項 (neural correlates)」と同義で使用している。つまり、現象的意識の有無と関連した活動を示す最小の神経群を、本稿は現象的意識の神経基盤と呼ぶ (アクセス意識も同様)。

4. この前提は認知科学において広く受け入れられている (Carruthers 2017, p. 68)。なお、この前提には彼が支持する「グローバルワークスペース理論」が関係しているが、紙幅の関係上説明は省略する。

5. 報告しえないような豊かな現象的意識の存在がまさに報告されている以上、その現象的意識が同時にアクセス意識的でもあったと考えることは不自然ではない (Cohen & Dennet 2011, p. 360)。

6. 重要なのは、被験者を実際にこのような状態にできるかどうかでなく、もし被験者がこのような状態になったならば、Block の想定の下で彼／彼女は現象的意識を持っていることになるだろうということである。

7. 例えば Tsuchiya et al. (2015) がそのような試みをしている。

(5) 参考文献

- Block, N. (1995) "On a confusion about a function of consciousness", *Behavioral and Brain Science*, 18, pp. 227-287.
- Block, N. (2005) "Two neural correlates of consciousness", *Trends in Cognitive Sciences*, 9 (2), pp. 46-52.
- Block, N. (2011) "Perceptual consciousness overflows cognitive access", *Trends in Cognitive Sciences*, 15 (12), pp. 567-575.
- Carruthers, P. (2017) "Block's overflow argument", *Pacific Philosophical Quarterly*, 98, pp. 65-70.
- Cohen, M. & Dennet, D. (2011) "Consciousness cannot be separated from Function", *Trends in Cognitive Sciences*, 15 (8), pp. 358-364.
- Nagel, T. (1974) "What is it like to be a bat?", *The Philosophical Review*, 83, pp. 435-450.
- Sperling, G. (1960) "The information available in brief visual Presentations", *Psychological Monographs*, 74, pp. 1-29.
- Tsuchiya, N.; Wilke, M.; Frässle, S. & Lamme, V. A. (2015) "No-report Paradigms: Extracting the true neural correlates of consciousness", *Trends in Cognitive Sciences*, 19 (12), pp. 757-770.
- 太田紘史 & 山口尚 (2009) 「反機能主義者であるとはどのようなことか」, *Contemporary and Applied Philosophy*, 2, pp. 1001-1017.

佐藤亮司 (2014) 「視覚意識の神経基盤論争——かい離説の是非と知覚経験の見かけの豊かさを中心に」, 『シリーズ新・心の哲学Ⅱ 意識篇』, 信原幸弘・太田紘史編, pp. 81-130, 勁草書房.

(新潟大学)

芸術の定義, グラウンド, 非循環性
Definition of Art, Ground, and Non-Circularity

坪井祥吾

Abstract

What *form* should the definition of art take? This is the main question in this paper. In particular, I focus on the property of non-circularity as a constraint that an adequate definition of art must satisfy. Until now, this property has not been satisfactorily discussed in the philosophy of art. I propose to analyze it in terms of *meta-physical ground*. As a result, the concept of non-circularity will be clarified, and some definitions that have been erroneously criticized as circular will be shown not to be circular at all.

1 研究テーマ

芸術とは何か? この問いに答えることは、**芸術の定義**を与えることである。芸術の定義論にはすでに豊富な研究があり、たとえばある者は〈芸術作品とは、美的によいものを作ろうという意図を持って作られたものことだ〉という定義を提案し、またある者は〈芸術作品とは、ある特別な社会的文脈のもとで作られたものことだ〉という定義を提案する (cf. Davies[2], Stecker[11])。こうした定義を提案・検討することは、〈芸術とは何か?〉という問いに答えようとする点で、一階の研究だと言える。

一階の研究において数多くの定義が提案されると、それらの定義を評価・比較する必要が自ずと生じる。その結果、近年、〈適切な芸術の定義はどのような制約を満たすべきか?〉という二階の関心が高まっている (cf. Gaut[7], Lopes[8])。この種の研究領域は**芸術のメタ定義論**と呼ぶことができるだろう。私の関心もここにある。本論の狙いは、特に、芸術の定義に課せられる〈定義は**循環してはならない**〉という制約に対してある分析を与えることである。すなわち、定義の非循環性は**グラウンド** [ground] が持つ性質から派生するものだ、という分析である。この分析は、芸術の定義論に対して二つの重要な含意を持つ。第一に、非循環性という概念が明確化される。第二に、定義が循環的かどうかを評価する際にグラウンドの観点からそうすべきだと分かる。とりわけ後者は重要であろう、というのも、4節で見る通り、非循環性の理解不足により不当な評価を受けていた芸術の定義が存在するだろうからである。

2 研究の背景・先行研究

本節では、(1) 芸術を定義するとはどういうことか、(2) 芸術の定義にはどのような制約が課せられるべきか、という二点それぞれについて、哲学者の間での共通見解を確認する。

一点目、芸術を定義するとはどういうことか。これの明確化には以下の二つの区別が役立つ。第一の区別は**実在的定義** [real definition] と**名目的定義** [nominal definition] の区別である。実在的定義は「*X* とは何か？」という問いに答えるための定義であり、名目的定義は「語『*X*』はどのように用いられているか？」という問いに答えるための定義だと言える。たとえば、(自然種としての) 水の実在的定義は〈水とは、 H_2O のことだ〉となるだろう。一方で、語「水」の名目的定義は〈「水」は、無色透明で無味無臭で... であるものに適用される〉となるだろう。そして、芸術の「定義」と言うときに意図されているのは実在的定義のことだと考えるのが自然である。というのも、芸術の定義論における根本的な問いは「芸術とは何か？」であって「『芸術』という語はどのように使われているか？」ではないからである。換言すれば、私たちの関心は芸術という種にあるのであって、「芸術」という語や概念にあるのではないのである。従って以下では、断りなく「定義」と言うとき、私は実在的定義のことを意味している。

第二の区別は、「芸術」という語の**分類的使用法** [classificatory usage] と**評価的使用法** [evaluative usage] の区別である (cf. Dickie[3])。分類的使用法は、「『モナ・リザ』は芸術だ」や「『4:33』は芸術だ」といった言明を典型例とするような用法で、与えられた対象が芸術作品というカテゴリーに属する/属さないことを述べる際の用法である。他方で評価的使用法は、芸術作品であるかどうかにかかわらず、何か並外れた特徴を持つ対象を評価する際に用いられる用法である。たとえば、「前田智徳のバッティングフォームはもはや芸術だ」のような言明において、「芸術」という語は評価的に用いられている。もちろん、芸術の定義論においては純粋に評価的な意味での「芸術」は無視すべきだし、実際にそうされてきた。というのも、再び、私たちの関心は芸術作品という種にあるからである。以下ではこの点を明確にするため、芸術の定義の被定義項を「芸術作品」として議論を進める。

以上より、芸術を定義するとは、〈**分類的な意味で**芸術作品であること**の実在的定義**を与えること〉だと言える。ただし以下では煩雑さを避けるため、「分類的な意味で芸術作品であること**の実在的定義**」のかわりに「**芸術の定義**」とすることにする。

二点目、芸術の定義にはどのような制約が課せられるべきか。本論では、〈**外延的妥当性**を満たせ〉という制約と、〈**非循環性**を満たせ〉という制約を取り上げる。まず、芸術の定義が外延的に妥当であるとは、語「芸術作品」の外延をその定義が過不足なく捉えている、ということである。芸術の定義は芸術作品であるものを**適切に選び出す**ものであってほしいことを考えると、定義

は外延的妥当性を満たすべきだと言えよう。実際、Stecker[11, p.14] いわく、外延的妥当性を満たせという制約は芸術の哲学者に広く共有されているものである。外延的妥当性については上の特徴づけによって十分に把握されるものとし、本論ではこの制約の適切さはこれ以上は問わない¹。

次に、芸術の定義が非循環的であるとはどういうことか。3節でも見るが、これがまさに本論の主題であり、未だ満足な議論が行われていない点である。よって、非循環性とはどのような性質であるかを、現段階では明確化することができない。だが、芸術の定義は芸術とは何であるかを説明するものであってほしいことを踏まえると、定義は何らかの意味で非循環的である必要があるだろう。というのも、一般に循環的な説明は悪い説明だとされるからである。実際、非循環性が定義にとって望ましい性質であることは芸術の定義論における共通見解になっているように思われる。詳しくは4節で見るが、たとえば Dickie 流の制度的定義にとって、それが循環的であることは大きな欠点であるとしばしば言われる (cf. Davies[2], Stecker[11])。これが欠点であるためには、非循環性の制約が受け入れられていなければならないだろう。

以上から、芸術のメタ定義論の目標を、**外延的妥当性と非循環性を満たすような仕方**で次の図式を埋めることとして設定することができる。

定義図式 芸術作品とは ϕ であるようなものことだ iff ... 芸術作品 ... ϕ ...

つまり、図式の右辺の「...」部の埋め方によって、外延的妥当性と非循環性を満たすようにするのである。ここで、最も単純なのは定義図式を次のように埋めることだろう。

純粋に外延的なメタ定義 芸術作品とは ϕ であるようなものことだ iff $\forall x(x$ は芸術作品である $\leftrightarrow \phi(x))$ 。

こうすると、たしかに外延的妥当性は満たされる。だが、非循環性は満たされない。というのも、たとえば $\forall x(x$ は芸術作品である $\leftrightarrow x$ は芸術作品である) は真であるため、〈芸術作品とは芸術作品であるようなものことだ〉という定義も許容されてしまうからである。これは循環的定義の最たるものだろう。従って、定義図式を埋める際に他に何か条件を追加しなければならない。しかしどのような条件を？ これが次節で取り組む問いである。

3 筆者の主張

前節で確認した通り、定義図式を埋める際に、非循環性に対応するような条件を追加する必要がある。しかし、そのような条件を特定するためには、まず定義が非循環的であるとはどういうことかを明らかにしなければならない。

ただし、残念ながら、これまでの芸術の定義論においてこのことについての十分な検討は行われていない。たとえば、Dickie 流の制度的定義に対するお決まりの批判は〈それは循環している〉というものであり、それに対する制度主義者からのお決まりの応答は〈それは循環だが悪い循環ではない〉というものである (cf. Dickie[4], Davies[2], Stecker[11])。こうした議論の応酬は、特定の定義が循環的かそうでないかを争うものであって、循環という概念**そのもの**を問題視するものではない。この応酬からは非循環性そのものの明確化は得られないだろう。

そこで私は、非循環性は**グラウンド**の概念によって解明されるべきだ、と提案する。これは、近年提案されている**実在的定義一般**の分析に基づいている (Rosen[10], Fine[6], Correia[1])。興味深いことに、それらの分析は、実在的定義の非循環性はグラウンドの持つ性質に由来する、とする点で一致している。たとえば、Rosen と Fine はグラウンドと本質 [essence] という概念を用いて実在的定義を分析する。その結果、実在的定義の非循環性はグラウンドに由来する性質だとされる²。Correia[1] は、グラウンドと一般化された同一性 [generalized identity] という概念を用いて実在的定義を分析する。ここでもやはり、実在的定義の非循環性はグラウンドに由来するものとみなされる。この〈実在的定義の非循環性はグラウンドの持つ性質に由来する〉というテーゼを、**グラウンド-非循環性テーゼ** (略して **GN テーゼ**) と呼ぼう³。

本論の狙いは、GN テーゼが真であると論証することにはない。そうではなく、それが真だと仮定したときに、芸術の定義論に対しどのような含意があるかを検討することにある。そこで、本節の残りでは、まずグラウンドの概念を導入し (3.1 節)、次にグラウンドによって非循環性の概念を分析する (3.2 節)。その際、GN テーゼがさしあたりもっともらしいことは強調しておく。続いて 4 節で、GN テーゼの持つ芸術の定義論への含意を確かめる。

3.1 グラウンド

まずはグラウンドの概念を導入しよう。グラウンドは**説明的な**概念である (cf. Fine[5])。それは特に、非因果的で**決定的** [determinative] な説明である。たとえば、「なぜこの窓ガラスは割れたのか？」と問われれば、「その窓ガラスが脆いからだ」と答えることができる。だが、この説明は因果的ではない。というのも、窓ガラスが割れたことを引き起こしたのは、たとえばボールがぶつかったことなどであって、その窓ガラスの脆さ自体ではないからだ。これはむしろ、より**基礎的な**性質 (脆さ) に言及して、より**派生的な**性質 (割れたこと) の実現がいかにして決定されているのかを説明しているのである。グラウンドとはこの種の説明のことである。今の例では、〈この窓ガラスが脆い

ことが、それが割れたことをグラウンドする〉のように言う。

また、標準的には、説明的概念であるがゆえに、グラウンドは**非反射的**、**非対称的**、**推移的**であるとされる (cf. Raven[9])。まず、どんなものも自身を説明することはないように、どんなものも自身をグラウンドすることはない (非反射性)。つまり、 A は A をグラウンドしない。次に、説明の順序を逆転できないように、グラウンドの順序も逆転できない (非対称性)。つまり、 A が B をグラウンドするなら、 B は A をグラウンドしない。最後に、説明が鎖をなすように、グラウンドも鎖をなす (推移性)。つまり、 A が B をグラウンドし、 B が C をグラウンドするなら、 A は C をグラウンドする。

さらに、表記法も導入しておこう。本論では Fine[5] に倣い、グラウンドを文オペレーター \langle として表すことにする⁴。たとえば、「 p が q をグラウンドする」は「 $p \langle q$ 」と書く。また、「 a が F であることが a が G であることをグラウンドする」は「 $Fa \langle Ga$ 」と書く。「任意の x について、 x が F であることが x が G であることをグラウンドする」は「 $\forall x(Fx \langle Gx)$ 」と書く。

3.2 非循環性の分析

続いて、Rosen[10], Fine[6], Correia[1] に従い、定義の非循環性を次のように分析する⁵。これは、上で見た GN テーゼを形式化したものである。

GN テーゼ $\langle F$ とは ϕ であるようなものことだ〉という定義が**非循環的**である iff $\forall x(\phi(x) \langle Fx)$ 。

すると、定義図式は次のように埋められることになる。

GN メタ定義 芸術作品とは ϕ であるようなものことだ iff $\forall x((x$ は芸術作品である $\leftrightarrow \phi(x)) \wedge (\phi(x) \langle x$ は芸術作品である))。

図式の右辺の二つの連言肢が、それぞれ外延的妥当性と非循環性を反映している。よって、これは 2 節で設定したメタ定義論の目標を達成している。

本節の残りで、この GN メタ定義のさしあたりのもっともらしさを強調しておこう。そのために、私は以下の二点を指摘する。第一に、2 節でも見たように、定義が循環してはならないと考えたくなる理由の一つは、循環的な**説明**が禁じられるべきだということである。すると、定義の非循環性の分析にグラウンドという**説明的**概念を用いることは、ごく自然なアイデアであろう。第二に、次の明らかに循環的な二つの定義を考えよう。GN テーゼに従うと、これらの定義を適切に循環的であるとみなせる。この意味で、GN テーゼを組み込んだ GN メタ定義は私たちの直観的判断ともよく合致していると言える。

- \langle 芸術作品とは、芸術作品であるようなものことだ〉という定義。**直観的には**、この定義は循環的である。そして GN テーゼに従うと、これは

正しく循環的ということになる。というのも、グラウンドの**非反射性**により、任意の x について、 $\langle x$ は芸術作品である $\langle x$ は芸術作品である \rangle は偽となるからである。よって $\langle \forall x(x$ は芸術作品である $\langle x$ は芸術作品である $\rangle)$ は偽で、従ってこの定義は循環的である。

- \langle 芸術作品とは、時点 t より前で芸術作品であるか時点 t 以後で芸術作品であるようなものことだ \rangle という定義。直観的には、この定義は循環的である。やはり GN テーゼに従うと、これも（ある仮定のもとで）正しく循環的だということになる。というのも、（複雑な性質は単純な性質にグラウンドされるという原理を仮定すれば、）任意の x について、 $\langle x$ は芸術作品である $\langle x$ は時点 t より前で芸術作品である $\langle x$ は時点 t より前で芸術作品であるか時点 t 以後で芸術作品である \rangle というグラウンドの列が存在するだろうからである⁶。すると、グラウンドの**推移性**より、 $\langle x$ は芸術作品である $\langle x$ は時点 t より前で芸術作品であるか時点 t 以後で芸術作品である \rangle となる。それゆえ、グラウンドの**非対称性**より、 $\langle x$ は時点 t より前で芸術作品であるか時点 t 以後で芸術作品である $\langle x$ は芸術作品である \rangle は偽となる。よって $\langle \forall x(x$ は時点 t より前で芸術作品であるか時点 t 以後で芸術作品である $\langle x$ は芸術作品である $\rangle)$ は偽であり、従ってこの定義は循環的である。

以上より、GN メタ定義は直観的判断と一致する。一方で、GN テーゼを組み込んでいない純粋に外延的なメタ定義（2節）は、これらの直観的判断と一致しない。というのも、上の二例ではいずれも、定義項と被定義項が明らかに共外延的だからである。純粋に外延的なメタ定義に従うと、定義項と被定義項が共外延的でありさえすればそれは適切な定義だとみなされるため、上の二例も適切な定義だということになってしまうのである。よって、他によい候補がなければ、GN メタ定義の方を採用するのが合理的だろう。

4 今後の展望

以上を踏まえ、GN テーゼが真だと仮定したときに、芸術の定義論に対してどのような含意があるのかを明らかにして、本論を終えよう。含意は少なくとも二つある。一つ目は、非循環性についてより明確な理解が得られる、というものである。これはすでに3節で見た通りである。つまり、定義の非循環性はグラウンドが持つ性質から派生する性質だ、という理解である。一般に、議論で利用されている概念を明確化することは重要であろう。

二つ目の含意は、これまでに提案されてきた芸術の定義のうち、非循環性についての理解不足のせいで、不当に低い/高い評価を受けていたものがある

と分かることである。つまり、誤って循環的であるとみなされて却下された定義や、逆にその循環性に気付かれずに流通している定義があるのである。この点に留意することは、芸術の定義論においてフェアな議論を行うために重要だろう。

ここで、不当に低い評価をされた定義の一例として、Dickieの制度的定義を挙げておこう。これを検討することは、GNテーゼによる非循環性の明確化の利点を鮮やかに教えてくれるだろう。まず、その定義は次の通りである：芸術作品とは、アートワールドの観衆に提示されるために制作された種類の人工物のことだ (Dickie[4, p.80])。そしてこれに対するお決まりの批判は、〈Dickieは芸術作品の概念をアートワールドの概念によって定義するが、アートワールドの概念を芸術作品の概念に訴えることなく定義することは不可能であり、従ってこの定義は循環的だ〉というものである (Davies[2, pp.109-12], Stecker[11, pp.70-2])。しかし、GNテーゼに従えば、この批判は誤りということになる。というのも、たしかに任意の芸術作品 a の存在はアートワールドの存在にグラウンドされ、同時にアートワールドの存在はある芸術作品 b, c, \dots の存在にグラウンドされるだろうが、しかし a は b, c, \dots のうちには含まれないだろうからである。なぜなら、 b, c, \dots は a よりも時間的に先立って存在するはずだからだ。よって、ここにグラウンドの循環はないのである。

さて、上の批判者の誤りからは、次の一般的な教訓が導ける。批判者は、芸術作品やアートワールドという概念に関して——つまり**タイプレベル**で——循環があることから、与えられた芸術の定義が循環していると判断してしまった。しかしながら、GNテーゼに従えば、定義が循環的であるかどうかを確かめるには、任意の個別的な作品 a に関して——つまり**トークンレベル**で——グラウンドの循環があるかどうかを調べるべきだったのである。

以上の検討から明らかなように、提案された定義をGNテーゼのもとで見直すことは実りのある作業である。私たちは過去の定義からまだ学ぶことができるかもしれないし、逆に現在受け入れられている定義を修正する必要があるかもしれない。さらには、その個別的な定義の検討から、芸術の定義についての一般的な洞察が得られるかもしれないのである。それゆえ、この作業は間違いなく生産的だろうと私は思う。本論の意義は、芸術の定義をめぐる有意義な議論のためにはこの作業が不可欠だということを示した点にある。

注

¹ ただし、ここで想定されている〈定義に先立って外延が定まっておき、それを定義が捉える〉という描像は不正確かもしれない。というのも、芸術作

品は人工物であるため、どのような定義を人々（作者・鑑賞者）が受け入れているかが芸術作品の外延を部分的に決定しているかもしれないからである。たとえば、〈芸術作品とは、美しく作られた物のことだ〉という定義を受け入れるコミュニティにおいては、芸術作品の外延に美しくない作品（e.g., ダダイストの作品）は含まれないかもしれない。この**外延の心への依存性**を加味して、外延的妥当性の制約はさらなる明確化が必要だろう。この論点を指摘してくださった匿名の査読者に感謝する。

² ただし、Rosen と Fine の分析はやや異なる。Fine の分析のもとでは、実在的定義の非循環性はグラウンドだけでなく本質の持つ性質にも由来する。

³ 実は、Rosen, Fine, Correia らは、定義の非循環性だけでなく外延的妥当性（に相当する論点）についても論じており、その上で、外延的妥当性よりも**強い制約**が必要であると主張する。本質 (Rosen[10], Fine[6]) や一般化された同一性 (Correia[1]) の概念が、そのような強い制約の設定に用いられるのである。本論では紙幅の都合上この点を論じられないため、以下では外延的妥当性の制約が適切であるという仮定のもとで議論を進める。

⁴ 正確には、<は**完全な** [full] グラウンドのオペレーターである。これと**部分的な** [partial] グラウンドは区別される (cf. Fine[5])。ただし本論では、議論の単純さのためにこの区別は無視する。

⁵ ただし Fine[6] は、グラウンドの文脈に現れる変数の束縛は、全称量化的に解釈すべきではないと論じている。紙面の都合上、この論点は無視する。

⁶ この列の最初の構成要素は、正確には〈 x は芸術作品である、 x は時点 t より前に存在する < x は時点 t より前で芸術作品である〉だろう。ただしこの場合、Fine[5] に従って、グラウンドのオペレーターの左辺には単一の文ではなく複数の文が現れることを許容していることに注意せよ。

文献

- [1] Correia, F. (2017). Real Definitions. *Philosophical Issues*, 27(1): 52–73.
- [2] Davies, S. (1991). *Definitions of Art*, Ithaca: Cornell University Press.
- [3] Dickie, G. (1969). Defining Art. *American Philosophical Quarterly*, 6(3): 253–256.
- [4] Dickie, G. (1984). *The Art Circle*, New York: Haven.
- [5] Fine, K. (2012). Guide to Ground. In Correia, F. and Schnieder, B. eds., *Metaphysical Grounding: Understanding the Structure of Reality*.

Cambridge: Cambridge University Press. 37–80.

- [6] Fine, K. (2015). Unified Foundations for Essence and Ground. *Journal of the American Philosophical Association*, 1(2): 296–311.
- [7] Gaut, B. (2000). The Cluster Account of Art. In Carroll, N. ed., *Theories of Art Today*, Madison: University of Wisconsin Press, 25–45.
- [8] Lopes, D. M. (2008). Nobody Needs a Theory of Art. *The Journal of Philosophy*, 105(3): 109–127.
- [9] Raven, J. (2015). Ground. *Philosophy Compass*, 10(5): 322–333.
- [10] Rosen, G. (2015). Real Definition. *Analytic Philosophy*, 56(3): 189–209.
- [11] Stecker, R. (1997). *Artworks: Definition, Meaning, Value*, University Park, PA: Pennsylvania State University Press.

(一橋大学)

不完備アウェアネスゲームにおける均衡の論理的分析
Logical Analysis of an Equilibrium in Game with Incomplete Awareness

久保 雄大

Abstract

Game situations with incomplete awareness/unawareness among the players have been explored in the context of the analysis of bounded rationality. Besides, logic has been long used to describe and analyze agents' epistemic assumptions, including rationality in several fields. In this paper, we introduce logic of awareness to analyze rationality in a game with incomplete awareness/unawareness. As for the main achievement, we prove that an equilibrium in the games always satisfies the logical formula characterizing rationality. This study clarifies rationality in a game with bounded rationality, intending to contribute epistemic foundation programs in game theory.

1 研究テーマ

本研究では、認識論理の一種であるアウェアネスの論理 (logic of awareness) を使用し、不完備アウェアネスゲーム (game with incomplete awareness) の均衡点が保持する合理性に関する性質を公理的に定式化する。

2 研究の背景・先行研究

ゲーム理論は特定の状況における合理的行動の解明を目的としており、ゲームの前提と複数のプレイヤー間の推測に基づくゲームの結果、つまりゲームの解の分析は主要な研究軸である。ナッシュによって戦略型ゲームの解として均衡点が示されてから、展開型ゲームや不完全情報ゲーム¹などの、より複雑なゲームの均衡点が提示されてきた。さらにその存在性や精緻化などの分析が行われてきた。その中で、プレイヤーに仮定されている合理性 (rationality) の現実とのギャップの指摘は古くから存在する [1]。このような限定合理性 (bounded rationality) に関する研究は、進化ゲーム理論や行動ゲーム理論といった領域を構成するに至った。これらでは実験研究を通して推論能力の限界や互惠性 (reciprocity) などの従来の合理性とのギャップを発見し、新たなモデルを構築している。一方で、均衡点や推論が必要とする合理性の分析という、いわば実験を通じた研究とは逆向きの方向で限定合理性にアプローチする研究が存在する [2, 3, 4, 5]。これらは合理性に関するどの理想化が解概念の成立に影響を与えるのかを明らかにするものであり、その意味で限定合理性の理論へ貢献する。また本稿は上記の中でも、一般の戦略型ゲームにおける被支配戦略の逐次消去法 (iterated deletion of strictly dominated strategies)² を特徴づける合理性を提示した Bonanno [4] の道具立てを多く使用する。

この文脈の中で、本稿では各種ゲームの中でも不完備アウェアネスゲームにおける均衡点が保持する合理性について、アウェアネスの論理を使って公理的に定式化する。不完備アウェアネスゲームは他のプレイヤーの存在やプレイヤーのとり得る選択などのゲームのルールと呼ばれる前提について、プレイヤーのアウェアネスが欠けているようなゲームを指す。ここではアウェアネスには「対象に気づく」や「対象へ意識を向ける」といった解釈を当て、その対象はプレイヤーの思考に使用できる情報のことを指す。例えば将棋や囲碁などがこのゲームの例といえ、これは相手のとる戦略全てを常に意識し続けているわけではないからである。つまり不完備アウェアネスゲームとは、プレイヤーは自身や他のプレイヤーがとり得る選択などを常に意識できるわけではないという限定合理性を仮定したゲームであり、企業的意思決定など多くのプラクティカルな具体例が考えられる重要なゲームといえる。同じ不完備アウェアネスゲームであっても、不完備情報も同時に扱えるものなど、より一般化された形式化も存在する [6, 7]。しかし今回は手始めに完全完備情報不完備アウェアネスゲームを取り扱った Feinberg [8] の形式化を使用し、さらに選択に関する気づきにのみに焦点を当てる。また [8] は各行動の選択に対する確率分布自体を戦略として採用する混合戦略 (mixed strategy) を用いているが、同様の理由から純粋戦略 (pure strategy) に限定し議論を進めることとする。このような単純化を行った不完備アウェアネスゲーム Γ^U は戦略型ゲーム Γ とアウェアネス/アンアウェアネス構成 (awareness/unawareness construction) U の組として定義される。 Γ は組 $\langle I, \{S_i\}_{i \in I}, \{\pi_i\}_{i \in I} \rangle$ であり、それぞれ I はプレイヤーの有限集合、 S_i はプレイヤー i の戦略の集合、 $\pi_i : S \rightarrow \mathbb{R}$ は i の利得関数である。ただし $S := \prod_{i \in I} S_i$ である。 U は全プレイヤーの戦略の集合 $\alpha := \bigcup_{i \in I} S_i$ の部分集合として、各 θ に依存し与えられる気づきの状態にある戦略の集合の族 $\{\alpha_\theta\}_{\theta \in \Theta}$ である。ただし $\Theta := \bigcup_{i=0}^n (I)^i$ とし、かつ以下の二つの条件を満たす。なお $\alpha_{\{\emptyset\}} := \alpha$ とする。

1. $\theta \supseteq \theta'$ ならば $\alpha_\theta \subseteq \alpha_{\theta'}$ である。ただし $\theta \supseteq \theta'$ は、 θ と θ' をそれぞれ (i_1, \dots, i_n) と $(i_{k_1}, \dots, i_{k_m})$ としたとき、 $1 \leq k_1 < \dots < k_m \leq n$ とする。
2. $i_k = i_{k+1}$ となる $\theta = (i_1, \dots, i_k, i_{k+1}, \dots, i_n)$ と $\theta' = (i_1, \dots, i_{k-1}, i_{k+1}, \dots, i_n)$ について、 $\alpha_\theta = \alpha_{\theta'}$ である。

例えば $\theta = (1, 2, 3)$ のとき $\alpha_{(1,2,3)}$ は、「プレイヤー 1 が気づいている「プレイヤー 2 が気づいている「プレイヤー 3 が気づいている全プレイヤーの戦略の集合」」を表す。また θ はゲームの制限に使用することができ、 Γ_θ および U^θ が定義できる。 Γ_θ は $\langle I, \{S_i^\theta\}_{i \in I}, \{\pi_i\}_{i \in I} \rangle$ で、 $S_i^\theta := S_i \cap \alpha_\theta$ である。 U^θ

は $\{\alpha_{\theta, \theta'}\}_{\theta, \theta' \in \Theta}$ で、なお $\theta \cdot \theta'$ は θ と θ' の接続である。また [8] はゲームの解概念として拡張ナッシュ均衡 (extended Nash equilibrium) が提示した。これは上記の単純化の元で、各 $\theta = (i_1, \dots, i_n)$ について、以下の二つの条件を満たす純粋戦略 $s_{i_n}^\theta \in S_{i_n}^\theta$ を割り当てる。

1. $s_{i_n}^\theta$ が $\{s_j^{\theta, j}\}_{j \in \mathcal{I} \setminus \{i_n\}}$ に対する最適反応である。
2. 最後のプレイヤーが一致する θ, θ' について、もし $\Gamma_\theta^{\mathcal{U}^\theta} = \Gamma_{\theta'}^{\mathcal{U}^{\theta'}}$ ならば、 $s_i^\theta = s_i^{\theta'}$ である。

なおプレイヤー i のある戦略 s_i が他のプレイヤーの戦略 $\{s_j\}_{j \in \mathcal{I} \setminus \{i\}}$ に対する最適反応であるとは、全ての $s'_i \in S_i$ について、 $\pi_i(s_i, (s_j)_{j \in \mathcal{I} \setminus \{i\}}) \geq \pi_i(s'_i, (s_j)_{j \in \mathcal{I} \setminus \{i\}})$ が成立することをいう。条件 2 はプレイヤーの気づきの状態が同じであるならば、拡張ナッシュ均衡である戦略も同じであると解釈される。戦略組としてこの概念を表現するならば、各 $\theta \cdot i$ について、拡張ナッシュ均衡として割り当てられる戦略の組 $(s_i^{\theta, i})_{i \in \mathcal{I}}$ を、 $\Gamma_\theta^{\mathcal{U}^\theta}$ の拡張ナッシュ均衡点と呼ぶことができる。

認識論理の一種であるアウェアネスの論理は、論理的全能性 (logical omniscience)³ を排除した現実の思考に近いような推論の表現を目的として、コンピュータサイエンス・経済学・哲学の各領域に跨って発展してきた [9]。不完備アウェアネスゲームともアウェアネスの概念の記述という点で交わってきた [10]。また論理を使用した定式化は、今後の拡張性において利点である。すなわち均衡点のように静的なものだけでなく、[5] で示されたように Max-Min 戦略⁴ などの特定の性質をもつ戦略を選ぶという意思決定に必要な合理性を射程に入れる際に、推論規則という形で意思決定手続きを取り扱うことができる。アウェアネスの論理にはゲームの形式化と同様に複数の論理が存在するが、今回は Awareness Logic with Partition (ALP) [11] を使用する。これは情報に対する気づきだけでなく、その気づきに応じて変化するエージェントの視点を考慮した論理である。マルチエージェント間における推論を正しく扱うためのもので、ゲーム理論のような他人の思考を推測し、均衡点を見つけることを目的とするものの分析に適している。しかしながらゲーム理論への応用は主目的としてないため、三階以上の高階の気づき (higher order awareness)、すなわち「プレイヤー 1 が気づいている「プレイヤー 2 が気づいている「プレイヤー 3 が気づいている戦略」」といった気づきの表現ができない。この拡張は今後の課題とし、今回は手始めに二階の気づきに限定した不完備アウェアネスゲームにおける均衡点の性質としての合理性の公理的な定式化を提示する。

3 筆者の主張

論理体系 \mathcal{ACP} と [4] の戦略の論理的表現をベースとし、前章の不完備アウェアネスゲームに依存した新たな論理 \mathcal{ACPG} を導入する。 \mathcal{P} を原子命題の集合とし、 \mathcal{ACPG} の言語 $\mathcal{L}_{\mathcal{P}}$ は以下の文法

$$\mathcal{L}_{\mathcal{P}} \ni \varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid A_j^i \varphi \mid [\equiv]_j^i \varphi \mid C_j^i \varphi \mid K_j^i \varphi \mid CK^i \varphi$$

によって生成される。ただし $p \in \mathcal{P}$ であり、さらに \mathcal{P} は特別な命題の集合

$$\bigcup_{i \in \mathcal{I}} S_i \cup \bigcup_{i \in \mathcal{I}} \{s_i \succeq_i s'_i \mid s_i, s'_i \in S_i\}$$

である。これらの特殊な原子命題はそれぞれ「 s_i を選択する」と「 i にとって s_i が s'_i よりも利得が少なくとも同程度に高い」を意味する。また $s_i \succ_i s'_i$ は $s'_i \not\succeq_i s_i \wedge s_i \succeq_i s'_i$ の略記とする。それぞれ $A_j^i \varphi$ は「 i の視点で j が φ に気づいている」、 $K_j^i \varphi$ は「 i の視点で j が φ に気づいていて、かつ知っている」、 $CK^i \varphi$ は「 i の視点で φ は意識下にある共有知識である」と読む。共有知識とは、「ある命題を全てのエージェントが知っており、かつそのことも全てのエージェントが知っており…」が無限回続くことで表現される知識のことである。これは「ある命題を全員が知っている」だけでは不十分であり、そのことに対する相互認識が必要とされる。他のオペレータ $[\equiv]_j^i$ と C_j^i は上記 K_j^i オペレータの定義に使用する。他の論理演算子について $\varphi \vee \psi$ と $\varphi \rightarrow \psi$ は、それぞれ $\neg(\neg\varphi \wedge \neg\psi)$ と $\varphi \wedge \neg\psi$ の略記とする。

次にセマンティクスを与える認識モデル M を定義する。 M は組

$$\langle W, \{R_i\}_{i \in \mathcal{I}}, \{\mathcal{A}_j^i\}_{i, j \in \mathcal{I}}, \{\equiv_j^i\}_{i, j \in \mathcal{I}}, \{\sigma_j^i\}_{i, j \in \mathcal{I}}, \rangle$$

であり、それぞれ以下のように定義される。

- W は空でない可能世界の集合。
- $R_i \subseteq W \times W$ は同値関係。
- \mathcal{A}_j^i は空でない原子命題の集合。ただし \mathcal{U} 内の (i, j) に対応する要素 $\alpha_{(i, j)}$ について、 $\alpha_{(i, j)} \subseteq \mathcal{A}_j^i$ を満たす。
- $(w, v) \in \equiv_j^i$ となるのは、全ての $p \in \mathcal{A}_j^i$ に関して $w \in V(p)$ のとき、またそのときに限り $v \in V(p)$ が成立するときである。
- $\sigma_i : W \rightarrow S_i$ は写像であり、 $(w, w') \in (R_i \circ \equiv_i^i)^+$ について $\sigma_i(w) = \sigma_i(w')$ を満たす。ただし関係 R^+ は R の推移閉包とし、 $R_i \circ \equiv_i^i$ は関係 \equiv_i^i と R_i の合成とする。

なお関係 R の推移閉包 R^+ とは、 R を含み推移性をもつ関係の最小の集合であり、関係 R' と R の合成は $\{(x, y) \mid \text{ある } z \text{ が存在して、}(x, z) \in R' \text{ かつ } (z, y) \in R\}$ である。 R_i は全ての命題に気づいている理想化されたエージェント i が考え得る可能性を示す。 \mathcal{A}_j^i は i 視点の j に関するアウェアネス集合 (awareness set) であり、気づいている原子命題の集合を意味する。 \equiv_j^i は気づいている命題に関しては区別がつかない二つの可能世界の間にかれる同値関係である。 R_i と組み合わせることで、気づきの欠けが原因で思考が制限されることを反映する。 $\sigma_i(w)$ は w で選択される戦略を示し、定義後半の条件はエージェントは自身が選んだ戦略を常に知っていることを表す。

$\text{At}(\varphi)$ を φ に登場する原子命題の集合、関係 R の反射推移閉包を R^* とする。なお反射推移閉包は R を含み、反射性かつ推移性をもつ関係の最小の集合である。このとき充足関係は以下のように定義される。

- $M, w \models s_i \Leftrightarrow s_i = \sigma_i(w)$.
- $M, w \models s_i \succeq_i s'_i \Leftrightarrow \pi_i(s_i, (\sigma_j(w))_{j \in \mathcal{I} \setminus \{i\}}) \geq \pi_i(s'_i, (\sigma_j(w))_{j \in \mathcal{I} \setminus \{i\}})$.
- $M, w \models \neg\varphi \Leftrightarrow M, w \not\models \varphi$.
- $M, w \models \varphi \wedge \psi \Leftrightarrow M, w \models \varphi \text{ かつ } M, w \models \psi$.
- $M, w \models A_j^i \varphi \Leftrightarrow \text{At}(\varphi) \subseteq \mathcal{A}_j^i$.
- $M, w \models [\equiv_j^i] \varphi \Leftrightarrow (w, v) \in \equiv_j^i$ となる全ての v について、 $M, v \models \varphi$.
- $M, w \models C_j^i \varphi \Leftrightarrow (w, v) \in (R_j \circ \equiv_j^i)^+$ となる全ての v について、 $M, v \models \varphi$.
- $M, w \models K_j^i \varphi \Leftrightarrow M, w \models A_j^i \varphi \text{ かつ } M, w \models C_j^i \varphi$.
- $M, w \models \text{CK}^i \varphi \Leftrightarrow M, w \models A_i^i \varphi \text{ かつ } (w, v) \in (\bigcup_{j \in \mathcal{I}} (R_j \circ \equiv_j^i))^+ \text{ となる全ての } v \text{ について、} M, v \models \varphi$.

特殊な原子命題 s_i について、前述のとおり「 s_i を選択する」という意味であるため、 $M, w \models s_i$ で「 w で s_i が選択される」という解釈がされる。関係 R_j と \equiv_j^i の合成の推移閉包をとった $(R_j \circ \equiv_j^i)^+$ は同値関係になるため、可能世界群の新たな分割を与える。これはエージェントが気づいていないために区別ができない可能世界群が、同値類として潰されることを示す。このように分割を粗くすることで、「エージェント 1 は「エージェント 2 が p を知っている」ことを知っている」のような推測を、気づきによる思考への影響を加味した上で正しく扱うことができる。 CK^i オペレータは反射推移閉包を取る対象を $(R_j \circ \equiv_j^i)^+$ としたもので、気づきによって制限されるエージェントの思

考による共有知識に対応するオペレータとなっており、通常のもの自然な拡張となっている。

続いて、[4]で提案された公理系を参考に、戦略に関するいくつかの前提を表現する論理式

$$\begin{aligned} \mathbf{G}_1 &: s_{i_1} \vee \cdots \vee s_{i_m}, & \mathbf{G}_2 &: \neg(s_i \wedge s'_i), \\ \mathbf{G}_3 &: s_i \rightarrow K_i^i s_i, & \mathbf{G}_4 &: s_i \succeq_i s'_i \vee s'_i \succeq_i s_i \end{aligned}$$

を妥当な論理式として導入する。本稿では証明論についてはふれないため、詳細な証明はしないが、写像 σ_i および利得関数の定義から、任意の M について恒真な論理式となる。それぞれ \mathbf{G}_1 は「プレイヤーは必ず選択をする」、 \mathbf{G}_2 は「プレイヤーは同時に二つの戦略を選択できない」、 \mathbf{G}_3 は「プレイヤーは自身が選択した戦略を知っている」、 \mathbf{G}_4 は「プレイヤーの選好関係について連結的である」を意味する。

合理性の共有知識とは、それぞれのプレイヤーの選択が合理的であることの共有知識をいう。合理性の形式化はその強度に合わせていくつか存在する [4] が、これらをベースに「プレイヤー i がある戦略 s_i をとるとき、 i は他のプレイヤーの戦略に対して、 s_i よりも常に優れている他の戦略 s'_i を知らない」を合理性として考えれば、 j の視点で戦略 s_i が合理的であることは以下の論理式で形式化できる。

$$\mathbf{R}: \bigwedge_{k \in \mathcal{I}} (s_k \wedge A_k^j s_k) \rightarrow \neg K_i^j s'_i \succ_i s_i$$

これは「他のプレイヤーが実際にとることになる戦略がプレイヤー j の意識下にあるという条件の下で、同じく j が気づいている i の他の戦略について、 i が実際にとることになる戦略よりも真に優れていることを j 視点での i は知らない」という意味になる。戦略 s_i が合理的であることが j の視点での共有知識であることは、以下の論理式で形式化される。

$$\mathbf{CR}: \bigwedge_{k \in \mathcal{I}} (s_k \wedge A_k^j s_k) \rightarrow \neg CK^j s'_i \succ_i s_i$$

前章で述べたように、本稿では二階の気づきに限定した不完備アウェアネスゲームについて分析を行う。これは任意の $\theta = (i_1, \dots, i_n)$ について $\alpha_\theta = \alpha_{(i_1, i_n)}$ の条件を加えることで表現できる。これは例えば「プレイヤー 1 視点での「プレイヤー 2 視点での「プレイヤー 3 の気づきの状態」」が、「プレイヤー 1 視点での「プレイヤー 3 の気づきの状態」と一致するという状況への制限を意味する。

命題. 二階の気づきをもつ不完備アウェアネスゲーム Γ^U について、 $(\sigma_i(w))_{i \in \mathcal{I}}$ が拡張ナッシュ均衡であるならば、 $M, w \models \mathbf{CR}$ である。

証明. 前提より、各 $\theta = (i_1, \dots, i_n)$ について $\sigma_{i_n}(w) \in S_{i_n}^{(i_1, \dots, i_n)} = S_{i_n} \cap \mathcal{A}_{i_n}^{i_1}$ が $\{s_j^{(i_1, \dots, i_n) \cdot j}\}_{j \in \mathcal{I} \setminus \{i_n\}}$ に対して最適反応である。ゲームは二階の気づきに制限されているため、 $s_j^{(i_1, \dots, i_n) \cdot j} = s_j^{(i_1, j)} = \sigma_j(w)$ 。よって全ての $s'_{i_n} \in S_{i_n}$ について $\pi_{i_n}(\sigma_{i_n}(w), (\sigma_j(w))_{j \in \mathcal{I} \setminus \{i_n\}}) \geq \pi_{i_n}(s'_{i_n}, (\sigma_j(w))_{j \in \mathcal{I} \setminus \{i_n\}})$ である。**CR** の前件である $M, w \models \bigwedge_{k \in \mathcal{I}} (s_k \wedge A_k^{i_1} s_k)$ と s_{i_n} の充足関係の定義から、 $M, w \models s_{i_n} \succeq_{i_n} s'_{i_n}$ 。次に $M, w \models \text{CK}^{i_1} s'_{i_n} \succ_{i_n} s_{i_n}$ を仮定する。よって $(w, v) \in (\bigcup_{k \in \mathcal{I}} (R_k \circ \equiv_k^{i_1})^+)^*$ となる全ての v について、 $M, v \models s'_{i_n} \succeq_{i_n} s_{i_n} \wedge s_{i_n} \not\prec_{i_n} s'_{i_n}$ となる。しかし $(\bigcup_{k \in \mathcal{I}} (R_k \circ \equiv_k^{i_1})^+)^*$ は反射的であるため、 $M, w \models s'_{i_n} \succeq_{i_n} s_{i_n} \wedge s_{i_n} \not\prec_{i_n} s'_{i_n}$ 。これは $M, w \models s_{i_n} \succeq_{i_n} s'_{i_n}$ と矛盾する。したがって、 $M, w \models \neg \text{CK}^{i_1} s'_{i_n} \succ_{i_n} s_{i_n}$ 。

この命題は **R** で特徴づけられる合理性の共有知識 **CR** の成立が、拡張ナッシュ均衡点の成立の必要条件となっていることを示す。

4 今後の展望

本稿では特殊な状況における不完備アウェアネスゲームの均衡点の認識論的前提を示した。しかしながら本文中でもふれたように、いくつかの課題が残されている。具体的には、

- 拡張ナッシュ均衡点の必要十分条件となる合理性に対応する公理の発見
- 完全性定理を満たす証明体系の構築
- 二階の気づきに限定しないゲームの分析
- 混合戦略への拡張

などが考えられる。また長期的な展望としては、合理性以外の認識論的前提の定式化や、特定の性質をもつ戦略を選ぶ意思決定段階において必要な前提の分析が挙げられる。

均衡点や戦略の決定プロセスが、どの程度の認識論的前提に基づいているのかを形式的に示すことで、プレイヤーの認識構造などのゲームの前提を明らかにすることができる。この意味でゲーム理論への貢献ができると考えられる。

注

¹ プレイヤーが同時に意思決定をしない動学ゲームにおいて、過去に行われた少なくとも一つの意味決定の内容がプレイヤー全員に共有されない場合、不完全情報ゲームと呼ばれる。似た概念に不完備情報ゲームがあるが、これはゲームの前提となるプレイヤーの数やとり得る戦略などの情報がプレイヤー全員に共有されないゲームを指す。

² ゲームから被支配戦略 (他のプレイヤーがとり得る全ての戦略の組み合わせに対して、ある戦略よりも真に劣っている戦略のこと) を取り除くことで、ゲームを徐々に小さくしていく手続きのことである。このようにして得られた戦略組は支配戦略均衡となることが知られている。支配戦略均衡である戦略組はナッシュ均衡点であるので、ナッシュ均衡を求める手続きとされることもある。

³ 公理 K: $K(\varphi \rightarrow \psi) \rightarrow K\varphi \rightarrow K\psi$ や規則 RN: $\vdash \varphi \Rightarrow \vdash K\varphi$ などで特徴づけられる推論能力に関する性質。「知っている情報と論理的な推論から得られる全ての情報も知っている」などの理想化された推論能力のことを指す。

⁴ そのプレイヤーのとり得る戦略について、それぞれ最も利得が低くなるようなケースを考えたとき、その中では最も利得が高くなる戦略のこと。

文献

- [1] Simon, H. A. (1957). *Models of man; social and rational*. Wiley.
- [2] Bernheim, D. (1984). Rationalizable strategic behavior. *Econometrica: Journal of the Econometric Society*, 52(4):1002-1028.
- [3] Pearce, D. (1984). Rationalizable strategic behavior and the problem of perfection. *Econometrica: Journal of the Econometric Society*, 52(4):1029-1050.
- [4] Bonanno, G. (2008). A syntactic approach to rationality in games with ordinal payoffs. In *Logic and the Foundations of Game and Decision Theory (LOFT 7)*, pages 59-86.
- [5] Kaneko, M. and Suzuki, N.-Y. (2002). Bounded interpersonal inferences and decision making. *Economic Theory*, 19(1):63-103.
- [6] Copic, J. and Galeotti, A. (2006) Awareness as an equilibrium notion: Normal-form games. Working Paper.
- [7] Perea, A. (2022). Common belief in rationality in games with unawareness *Mathematical Social Sciences*, 119:11-30.

- [8] Feinberg, Y. (2005). Games with incomplete awareness. Technical report, Technical Report Research Paper Series# 1894, Stanford Graduate School of Business.
- [9] Schipper, B. C. (2015). Awareness. In van Ditmarsch, H., van Der Hoek, W., Halpern, J. Y. and Kooi, B.(Eds.), *Handbook of epistemic logic*, pages 77-146. College Publications.
- [10] Halpern, J. Y. (2001). Alternative semantics for unawareness. *Games and Economic Behavior*, 37(2):321-339.
- [11] Kubono, Y., Racharak, T. and Tojo, S. (2023). Logic of Awareness in Agent's Reasoning. In *Proceedings of the 15th International Conference on Agents and Artificial Intelligence*. 1:207-216.

(静岡大学)

理解についての内在主義と外在主義
Internalism and Externalism about Understanding

谷川 綜太郎

Abstract

The internalism-externalism debate about knowledge is traditional in epistemology. There is a similar debate about understanding, although understanding tends to be considered an internalist notion. This paper clarifies internalist and externalist conceptions of understanding and reviews their disagreement. This disagreement lies in the necessity of meta-cognition, that is, whether it is possible to understand without understanding that one understands. From an externalist standpoint, Grimm argues that infants and animals without meta-cognitive abilities can understand. This paper also argues that Grimm's argument fails to account for the special value of understanding.

1 研究テーマ

本研究では、理解についての内在主義と外在主義、およびそれらの対立点を明らかにしたうえで、適切な解釈を検討する。

2 研究の背景・先行研究

近年、認識論では理解に対する関心が高まっている。その理由の一つは、理解の認識的な価値の高さにある。たとえば、ある定理を理解している状態は、単にその定理を知っている状態と比べて、認識的に価値が高いように思われる。もちろん、これは単なる直観であり、この直観を正当化するためには、理解のどの部分が知識にはない価値を生み出しているのかを具体的に特定する必要がある。この作業を行うためには、理解を厳密に定式化しておく必要があるものの、現状、その定式化に関しては内在主義と外在主義が対立している。本節では、この対立について概観する。

2.1 認識論における内在主義と外在主義

本題に入る前に、まずは認識論における内在主義と外在主義について簡単に説明しておきたい。多くの場合、これらは知識についての立場である。知識についての内在主義によると、何かを知っていると言うためには、それに関して真なる命題を信じていることに加えて、認識主体が反省のみによってその信念を認識的に正当化できる必要がある。たとえば、日本が議院内閣制を採用していることを知っていると言うためには、「日本は議院内閣制を採用している」という命題が実際に真であり、また、認識主体はその命題が真であると信じていることに加えて、反省のみによってその信念を認識的に正当

化できる（たとえば、教科書にそう書いてあったからだと主張できる）必要がある。

それに対して、知識についての外在主義によると、最後の条件は必要ない。たとえば、信頼性主義という外在主義の立場によると、知識とは信頼できるプロセス（すなわち、実際に真なる信念を産出する傾向のあるプロセス）により産出された真なる信念に他ならない。したがって、認識主体が反省のみによってその信念を認識的に正当化できるか否かは関係ない。たとえば、教科書を見るというプロセスが信頼できるプロセスであるとするなら、教科書を見ることにより形成された真なる信念はすべて知識である。もし仮に、認識主体がその信念をどのように形成したのかを忘れてしまっていたとしても（すなわち、反省のみによって認識的に正当化できなかつたとしても）、それが知識であることに変わりはない。

知識についての内在主義と外在主義の対立には、KK テーゼが関連すると考えられている。KK テーゼとは、「認識主体 S が命題 P を知っているなら、S は命題「S は P を知っている」も知っている」というテーゼである¹。一般に、内在主義ではこのテーゼを支持する傾向があり、外在主義では支持しない傾向がある。たとえば、(外在主義である) 信頼性主義では、あるひよこ鑑定士がひよこの性別を知っていると言えるのは、そのひよこ鑑定士が信頼できる仕方でひよこの性別に関して真なる信念を形成しているときであり、かつそのときに限る。この条件を満たすために、そのひよこ鑑定士が自分がひよこの性別を知っているという事実（すなわち、自分が信頼できる仕方でひよこの性別に関して真なる信念を形成しているという事実）を知っている必要はない。他方、内在主義では、そのひよこ鑑定士がひよこの性別を知っていると言えるのは、自分の信念を反省のみによって認識的に正当化できるときである。この条件を満たしているとき、そのひよこ鑑定士が自分がひよこの性別を知っているという事実（すなわち、自分がひよこの性別に関して認識的に正当化された真なる信念を持っているという事実）を知らないとは考えづらい。このように、知識についての内在主義では KK テーゼを支持することが自然であり、外在主義では支持しないことが自然である。したがって、本稿ではさしあたり、KK テーゼを支持する立場を知識についての内在主義、支持しない立場を知識についての外在主義だと考えることにする。

2.2 理解についての内在主義

以上のことを踏まえて、議論の焦点を理解に移したい。まずは理解の基本的な成立条件を明確にしておこう。たとえば、ピタゴラスの定理を理解していると言うためには何が必要だろうか。少なくとも、知識の場合と同様、定理

に関して真なる命題を信じている必要があるだろう。そこには、「直角三角形における斜辺の長さの平方は他2辺の長さの平方の和に等しい」という命題や、定理の証明に関する命題が含まれているかもしれない。そのうえで、理解していると言うためには、それらの命題を個別に信じているのではなく、何らかの仕方に関連付けて信じている必要があると言いたくなるかもしれない。認識論者たちも同様に考える傾向があり、たとえばクヴァンヴィグは次のように述べている。

理解は、巨大で包括的な情報の集合の中で、説明的な関係や、その他の整合性を生み出す諸関係をつかむこと (grasping) を要求する。関連付けられていない多くの情報の断片を知ることができるが、理解は当の主体によって個々の情報が繋ぎ合わせられたときにのみ達成される。 [3, p. 192]

なるほど、たとえばある地域の歴史を理解するためには、その地域の歴史に関して個々の事実を信じているだけでなく、それらの事実を整合的になるような仕方でも結び付けている必要がある。さしあたり、このような要素が理解を本質的に特徴付けていると考えることにしよう。すなわち、関連する情報の中で整合性を生み出すような諸関係を「つかんでいる」という要素が、理解を本質的に特徴付けていると考えることにしよう。

さて、(上記のクヴァンヴィグの見解にも見られるように、) 多くの認識論者は関連する情報を単に関連付けられるだけでなく、内在的に (すなわち、反省のみによって) 関連付けられる必要があると考えている。この考えに従うなら、たとえばピタゴラスの定理を理解していると言うためには、単に証明が定理を正当化するという関連付けを行えるだけでなく、反省のみによってそれを行える必要がある。この点に同意する立場を、理解についての内在主義と呼ぶ。

理解についての内在主義をさらに明確化するために、本稿ではさしあたり次のテーゼを支持する立場だと考えることにする。すなわち、「認識主体 S が対象 O (あるいは命題 P) を理解しているなら、S は命題「S は O (あるいは P) を理解している」も理解している」というテーゼ (すなわち、KK テーゼの理解版 = UU テーゼ) である。このテーゼが正しければ、たとえば自分がピタゴラスの定理を理解しているという事実 (すなわち、自分がピタゴラスの定理に関して真なる命題を信じており、かつそれらの命題がどのように繋がっているのかをつかんでいるという事実) を理解しないままピタゴラスの定理を理解することは不可能である。このことは、ザグゼブスキの以下の記述にも現れている。

理解は〔中略〕ある種の意識的な透明性 (transparency) によって構成されている状態である。自分が知っていることを知らずに知ることが可能かもしれないが、自分が理解していることを理解せずに理解することは不可能である。 [6, p. 246]

このように、理解についての内在主義者は UU テーゼに同意しつつ (すなわち、理解についての内在主義を支持しつつ)、KK テーゼに同意しない (すなわち、知識についての外在主義を支持する) ことで、理解と知識を区別する傾向がある。

2.3 理解についての外在主義

理解についての内在主義に対する主要な批判は、グリム [1] によって提出されている。グリムによると、メタ認知能力 (すなわち、自分の認知状態を認知する能力) を持たない乳児や動物などの存在者も十分に理解できることから、UU テーゼは誤りである。

グリムの議論を順に見ていこう。グリムによると、何かを理解していると言うためには、それに関して正確な「因果マップ (causal map)」を持っている必要がある [1, p. 214f.]. 正確な因果マップとは、実際に成立している因果関係や依存関係を正確に表現したものであり、ある人が正確な因果マップを持っているか否かは、その人が実際に正確な因果推論を行えるか否かに等しい。たとえば、ある議会の仕組みを理解していると言うためには、その議会の仕組みに関して正確な因果マップを持っている必要がある。この場合の正確な因果マップとは、その議会の仕組みを構成する様々な要素 (たとえば、法案の提出や委員会の設置など) の複雑な依存関係を正確に表現したものであり、ある人がそれを持っているか否かは、その人がその議会での手続きに関して実際に正確な因果推論 (たとえば、その議会において法律が成立するまでにどのような手続きが必要であるかについての推論) を行えるか否かに等しい。

ここで重要なことは、理解していると言うためには、実際に正確な因果推論を行えていれば十分であり、自分が正確な因果推論を行えているという事実を理解している必要はないということである。グリムがその根拠として挙げているのは、メタ認知能力を持たないが、実際に正確な因果推論を行える乳児や動物などの存在者に対して、十分に理解を帰属できるという直観である。グリムが挙げている事例の一つ見てみよう [1, p. 221]. ある実験では、生後数ヶ月の乳児に対して衝立の後ろからお手玉を投げ、衝立を取り払ったときに人がいる場合といない場合とで乳児の反応の違いを観察した。結果、人がいない場合にのみ、乳児は驚いたような反応を見せた (すなわち、衝立が

あった場所を有意に長い時間眺めていた)。実験者の分析によると、乳児は既に簡単な因果マップを形成しており、人の存在がお手玉の出現を引き起こすという簡単な因果推論を行っていたため、その予想が裏切られたことにより驚いたような反応を見せたのである。もちろん乳児は自分が理解していること（すなわち、実際に因果推論を行えていること）を理解していないだろうが、グリムはこのような事例を真正な理解の事例だと考えることで、UU テーゼを拒否し、理解についての外在主義的な定式化が可能だと結論している。

3 筆者の主張

最初に述べたように、理解が注目に値する理由は、理解には特別な価値（すなわち、知識にはない価値）があると考えられているからである。しかし、グリムによる定式化を採用した場合、理解にそのような価値があることをうまく説明できないだろう。

この点を説明するために、まずはグリムによる定式化を採用した場合に、理解の特別な価値をどのように説明できるのかという点を明らかにしておきたい。グリムは次のように述べている。

理解の特別な価値は、世界がどのように展開するのかを予測し、場合によっては制御することに対するわれわれの自然な関心に由来する。 [1, p. 226.]

グリムによる定式化を採用した場合、理解している人は少なくとも実際に正確な因果推論を行える。グリムはその能力を用いることで、「様相空間や反事実的状况において物事がどうなっているのか」を「見通せる」と考えており [1, p. 225]、さらにその能力を用いることで、世界の展開を予測したり制御したりできると考えている。グリムによると、それが理解の特別な価値に他ならない。

しかし、世界の展開を予測したり制御したりできるということは、本当に理解の特別な価値だと言えるだろうか。この点は明らかではないと主張したい。なぜなら、そのような能力は知識の集積によって十分に獲得可能だと考えられるからである。なお、ここで想定している知識とは、事実についての知識（すなわち、knowing-that 型の知識）や方法についての知識（すなわち、knowing-how 型の知識）を含む、これまでに議論されてきた種類の知識である。たとえば、明日の天気を予測する能力は、今日から明日にかけての天気を理解するまでもなく、現在の天気の状態についての知識や、それに基づいて天気を予測する一般的な方法についての知識の集積によって十分に獲得可能だと考えられる。また、計算機を制御する能力は、計算機を理解するまで

もなく、計算機を制御する方法についての知識の集積によって十分に獲得可能だと考えられる。したがって、予測したり制御したりする能力が理解することによって獲得可能だとしても、そのような能力は（これまでに議論されてきた種類の）知識の集積によって十分に獲得可能だと考えられるため、それによって理解の特別な価値を説明することはできないと考えられる。このような理由から、グリムによる理解の定式化を採用する限り、理解の特別な価値は明らかではないと主張する。

以上のことを踏まえて、改めて理解の特別な価値は何かと考えるなら、それはやはり、情報がどのように関連しているのかを反省のみによって説明できることだと思われる。すなわち、天気を理解している人に価値があるのは、天気に関する様々な事実がどのように関連しているのかを反省のみによって説明できるからであり、計算機を理解している人に価値があるのは、計算機に関する様々な事実がどのように関連しているのかを反省のみによって説明できるからである。反対に、メタ認知能力を持たない乳児や動物がいくら因果推論を行えたとしても、そのような能力を持ち得ないことは明らかである。したがって、理解が本質的にそのような価値を持つと考えるなら、UU テーゼを適用することもまた妥当である。以上の理由から、本稿では理解についての内在主義を支持する。

4 今後の展望

理解の価値を他の観点から説明する試みもある。たとえば、クヴァンヴィグ [4] は理解が「好奇心 (curiosity)」を充足するという観点から、理解には「最終的な価値 (final value)」があると主張している。また、プリチャード [5] は理解が本質的に「認知的達成 (cognitive achievement)」を伴うという観点から、理解には最終的な価値があると主張している。これらの試みを検討することは今後の課題である。

注

¹KK テーゼはヒンティッカ [2] により定式化されている。

文献

- [1] Stephen R. Grimm. Understanding and transparency. In Stephen R. Grimm, Christoph Baumberger, and Sabine Ammon editors, *Explaining Understanding: New Perspectives From Epistemology and Philosophy of Science*, pages 212–229, Routledge, New York, 2017.

- [2] Jaakko Hintikka. *Knowledge and Belief: An Introduction to the Logic of the Two Notions*, Cornell University Press, Ithaca, 1962.
- [3] Jonathan L. Kvanvig. *The Value of Knowledge and the Pursuit of Understanding*, Cambridge University Press, New York, 2003.
- [4] Jonathan L. Kvanvig. Curiosity and the response-dependent special value of understanding. In Tim Henning and David P. Schweikard editors, *Knowledge, Virtue and Action: Putting Epistemic Virtues to Work*, pages 151–174, Routledge, New York, 2013.
- [5] Duncan Pritchard. Knowledge and understanding. In Duncan Pritchard, Alan Millar, and Adrian Haddock editors, *The Nature and Value of Knowledge: Three Investigations*, pages 1–88, Oxford University Press, New York, 2010.
- [6] Linda Zagzebski. Recovering understanding. In Matthias Steup editor, *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue*, pages 235–252, Oxford University Press, New York, 2001.

(千葉大学)

知覚のきめ細かさからの論証に対して概念主義を擁護する
Defending Conceptualism against the Argument from Fineness of
Grain of Perception

岡部幹伸

Abstract

Is perceptual content conceptual? This problem is one of the most-focused issues in philosophy of perception. Non-conceptualism, according to which perceptual content is non-conceptual, holds that the fine-grained nature of perception yields a good argument against conceptualism. Non-conceptualists claim that perception is not conceptual because it is too fine-grained to be captured by concepts. Conceptualists, however, reply that “demonstrative concepts” is possible and such concepts could object-dependently capture perceptual content. This paper aims to refute the argument from fineness of grain of perception and then to show that demonstrative concepts could play an essential role in McDowell’s therapeutic philosophy.

(1) 研究テーマ

知覚経験の内容は概念的であると主張する立場は、一般に概念主義と呼ばれている。もし知覚が概念的でないとすると、非概念的な知覚が概念的な信念をいかにして正当化できるかという問題が生じるように思われる。私は赤いリンゴがそこにあるという知覚をして、赤いリンゴがそこにあるという信念を獲得する。このとき、この信念を知識たらしめるのは私がしたリンゴの知覚である。概念主義をとる利点の一つは、知覚と信念のこのような正当化関係を問題のないものにできることにある。

しかしながら、概念主義に対してはいくつかの異論が提出されている。その一つが知覚のきめ細かさからの論証であり、知覚は概念によってはとらえきれないほどきめが細かく、知覚の内容と概念的な内容一般にはギャップがあるため、知覚は概念的ではないとするものである。本稿は、この論証およびそれに対する概念主義者からの応答（直示的概念による応答）を検討し、知覚のきめ細かさからの論証に反論することを目指す。その後、知覚の哲学の局所的な論争テーマにすぎないと思われがちな直示的概念が、マクダウェルの治療としての哲学において、重要な役割を担うことを示す。

(2) 研究の背景・先行研究

おそらく、経験の内容をすべて捕捉できるほど多くの概念を我々が持っているとは信じられないという直観が、知覚の非概念性を主張する一番大きな動機となっている。「赤」や「ピンク」といった色概念しか使うことができないのなら、我々が経験する微妙な色の移り変わりを概念が捉えることなどありそうもないように思われる。たとえば、赤₁₉と赤₂₁を私は現象学的に異なったものとして経験する。同じ赤でも色合いが違い、私はそれに気づいているのである。しかし、両者を区別して記憶しておくことはできないということはあるだろう。そのようなとき、ペンキ屋に行って、特定の赤色を選び出すことはできないはずだ(Tye 2000, 11; 61)。というのは、私は両者を区別して記憶できていない、すなわち再認することができないからである。これは赤という概念はもっていても、それ以上細かい概念をもっていないということではないのか。このような現象学的な論点はエヴァンズによって「我々が感覚的に識別できる色合いと同じほど多くの色概念を我々がもつという提案を、我々は本当に理解するだろうか？」(Evans 1982, 229)と明確に表現され、概念主義論争での主要な争点となった。

ヘックはこの困難を「時間が欠けているのではなく、記述的資源が欠けている」(Heck 2000, 490)と表現する。細かすぎるために、人間の一生が終わるまでには、あるいは太陽が膨張して地球を飲み込むまでには経験内容の記述が終わらないということではない。そもそも概念によって記述し、経験の内容を特定しようとしてもそれが叶わないということなのである。その理由こそ、我々のもつような概念が、我々が識別できる色合いよりもきめが粗いというものである。

経験の内容が概念よりもきめが細かいという主張に対して、概念主義者はどのように応答できるだろうか。『心と世界』においてマクダウェルは、直示的概念というものを認めることで、経験の内容を概念が捕捉できるようになると言う。

主体の概念能力を超えているとされる種類の経験——仮定により適切な見本を与えてくれる経験——のただなかにあつて、主体は「あの色合い」といった句を発話することによって、当の経験と正確に同じきめ細かさをもつ概念に言語的な表現を与えることができる。「あの色合い」のなかで、直示語は見本の現前を利用する。(McDowell 1996, 56-57 [邦訳、102頁])

直示的概念を用いる能力とは、短時間にすぎなくても、色見本を提示されたなら当の色なのかどうか答えることができるような再認の能力である。ポイントは、経験をする前から予め概念をもっている必要がなくなるということである。このような能力は、当の経験を離れてどういう能力なのか言うことはできず、直示的概念もその経験を離れて成立するようなものではない。このような概念を見落としたことに対して、マクダウェルはエヴァンズを批判する。

直示的概念を認めれば、経験内容のきめ細かさからの論証をかわすことができる。しかし、直示的概念に対しては、いくつかの異論が提出されている。

ピーコックは直示的概念はきめが細かすぎると批判する。ピーコックによれば、形を知覚するとき、三つの異なるレベルがあるという (Peacocke 1998, 381)。

(i) 形そのもの

(ii) 経験において知覚された形

(iii) 「あの形」「あの正方形」といった発話で捉えられる直示的に概念化された形⁽¹⁾

マクダウェルは(iii)によって(ii)を説明しているとピーコックは考える。色の例がわかりやすいのでそれを用いる。「あの色合い」「あの赤」「あの緋色」といった直示的概念はすべて異なる概念であろう。だが、これらの概念のどれもが経験において知覚された同じ色を指示するということがありえる。二人のひとが、ひとり「あの赤」で、もうひとは「あの緋色」で同じ対象の同じ色を指すことがありえるのである。「あの赤」と「あの緋色」は異なった概念だが、直示的概念として捕捉すべきは同じ対象の同じ色という程度までの識別可能性である。二人がそれぞれ「あの赤」「あの緋色」によって対象の色を捉えるのは不必要なきめの細かさを用いているのである。

この二人が同じ対象の同じ色を知覚しているという直観を退け、経験の内容、および経験の内容の捕捉が「あの赤」「あの緋色」というレベルで個別化されると考えるのはどうだろうか (Kelly 2001b; Ablondi 2002)。画家や室内装飾家という色の専門家の経験の内容が、一般人の経験の内容とは異なるというののもっともらしい。それなら、それぞれの経験が同じ対象の色についてのものであったとしても、経験の内容が異なるのだから、直示的概念が不必要にきめが細かいということにはならないと思われるかもしれない。しかし、ピーコックが指摘するように、「あの赤」「あの緋色」でそれぞれの経験は個

別化されるとしても、なおもそれぞれの経験に共通する要素が残り続けるのである (Peacocke 2001, 610)。いわば指示のレベルをこの方針は取り逃がすことになる。

直示的概念を用いる際に「赤」や「緋色」といった一般概念を使うから問題が生じるのなら、すべて「それ」で押し通し、自分を中心にした空間の位置関係を補足することで問題を回避できないだろうか。しかし、「それ」だけでは「その表面」なのか「その色合い」なのか「その質感」なのか「その形」なのか決定できないだろう。経験の内容を推論にもたらずなら、直示的概念がどのような種類についてのものなのかはわかっているなければならない。それゆえ、一般概念は必要で、この方向で問題を回避することはできないとピーコックは指摘する (Peacocke 1998, 382)。このように、たんに「それ」がすべてを問題なく指示しうると言うことはできず、この方向では解決ができない。ピーコック自身は後に、色や形といった諸性質の中のどれが与えられるにせよ、一つの与えられ方によって与えられる性質は一つだけなのだから、与えられ方が固定されれば、「色」や「形」で補足されなくても済み、「それ」だけでよいと主張する (Peacocke 2001, 610-11)。だが、経験の内容の与えられ方を固定するためには、対象の色が与えられているのか形が与えられているのか決定しなければならないので、非決定性は回避できないはずである。どのみち「それ」がすべてを問題なく指示しうると言うことはできない。

(3) 筆者の主張

結局のところこの問題に対しては、ケリーが示唆した、「その色合い」というレベルで同じものを指す可能性を探るべきだろう⁽²⁾ (Kelly 2001b, 226-7)。実はピーコックは、「あの赤」「あの緋色」という直示的概念を用いる二人について、どちらも「色合い」という一般概念を欠いているという不当な制限を課していた。しかし、この制限を取り払うなら、「その色合い」という表現を用いることによって、経験の内容とちょうど同じきめ細かさで経験の内容が概念的であると言えるだろう⁽³⁾。

ところがタイは、「その色合い」にしたところで、色に注意を向けてそのような概念を形成したりすることなく、経験を享受することが可能なのだから、直示的概念は役に立たないと主張する (Tye 2000, 61; 75)。だが、このタイの指摘は妥当なものではない。直示的概念のポイントは、問題となる知覚経験の前にあらかじめ概念を主体が持っている必要がないということにあるからである。そして、マクダウェルによれば、再認能力が持続している限り、その経験が終わったあとでも概念には明確な表現が与えられうる (McDowell 1996, 171 [邦訳、277])。タイの批判は直示的概念の狙いを理解していない

ものでしかない。

ヘックは、直示的概念の指示対象の固定の問題を議論している(Heck 2000, 494-498)⁽⁴⁾。「その色」と言って私の机の一部の色を指すとする。このとき、知覚が真正なものならば、「その色」の指示は現実の机の色によって固定されるだろう。しかし、私が誤知覚をしているならばどうだろうか。たとえば、目の不調によって、実際の茶色よりも赤みがかかった茶色を経験することもありえるだろう。この場合、「その色」によって現実の机の色が指示されるなら、経験の内容を「その色」という直示的概念によって特定することはできないことになる。なぜなら、誤知覚の場合の経験の内容は、現実の机が持つ茶色ではなく赤みがかかった茶色だからである。このようにして、直示的概念の指示が現実の見本によって固定され、それによって経験の内容を特定するということには抵抗したくなってくる。

それならば、直示的概念の指示は、現実の机が持つ色ではなく、私に見える色によって固定されると考えるのはどうだろうか。ヘックは、このようにすれば直示的概念の指示の問題は解決すると考える。しかし、直示的概念の指示は世界の側の色見本によって決まると主張するマクダウェルにとって、この方針は受け入れがたい。ヘックはマクダウェルの反応は次のようなものになるだろうと言う。すなわち、知覚の哲学において選言説をとるマクダウェルは、ヘックのこの議論に対して、この議論は錯覚論法と同様であり、誤知覚と真正な知覚のあいだに共通要素はなく、誤知覚の場合を持ち出して真正な知覚経験の内容を直示的概念によって特定することに反対するのは間違っていると言う、と。

これに対してヘックが論じるのは、内容の理論について、選言的なものは見込みがないということである。それが真なら雪は白いという内容を持ち、それが偽なら雪はピンクという内容を持つ、選言的な信念を考えよう。このような信念は悪循環に陥る。どちらの内容なのか特定するには、この信念が真か偽か決定しなければならない。しかし、真偽を決定するには内容が確定していなければならない。前もって真偽と内容のどちらも知ることができないので、ここから抜け出せないのである。

同様のことが知覚の内容についても言えるとヘックは主張する。選言的な説明では、知覚状態が真正なものであるかが、知覚状態の内容が何であるかに先立つことになる。しかし内容がわからなければ真実かどうか決められないので、このような理論は見込みがないというのである。

これに対して何が言えるだろうか。ヘックのこの議論は、内容についての選言的な説明に対しての批判としては成功していない。選言説は、主体が A

または B という内容の経験を享受すると主張する立場ではない。主体が経験するのは A だけ（または B だけ）なのである。そして、A と B が主体からは区別できないというにすぎない。ヘックの議論が当てはまるのは主体が A または B という内容の経験を享受すると想定した場合だけである。主体の内的な観点からは経験が真正なものかはわからないが、内容は客観的に決まっている。主体が経験するのは選言の片方のみだということがわかれば、内容の非決定性も生じないのである。

ヘックのような反応はおそらく、心の領域を主体にとって不可謬なものにしたいというデカルト以来の認識論に特徴的な願望に由来している。「完全なデカルト的描像においては、内的生〔筆者注：心〕は自立的な領域で起こり、それは主体の内観的気づきに透明である」(McDowell 1986, 236)。デカルト的描像最大の問題は、不可謬なものだけからなる心という領域に閉じ込めて安定した足場を確保しようとしたあげく、世界を喪失することにある。直示的概念の指示が現実に机が持つ色ではなく、私に見える色によって固定されるなら問題がなくなると考えたヘックの考えは、デカルト的描像そのものである。幻覚や錯覚の場合にも直示的概念を機能させようとして、真正な知覚と共通するとされる主体に対する見えに後退し、この見えが直示的概念の指示を決定すると、おそらくは彼なりに最大限好意的なかたちで論敵を解釈する。しかしこれはマクダウェルが拒否しようとする描像そのものであるから、マクダウェルによって拒否されるのは当然である。にもかかわらず、ヘックはデカルト的描像を克服するという背景に目を向けないまま、選言的な説明を斥けようとするのである。

我々は、直示的概念は真正な知覚においてのみ有効な概念だと結論すべきである。その指示対象が存在すると誤って思われている場合などは、そもそも直示的概念を使った知覚内容の同定に失敗している。直示的概念を導入する当初の眼目が、真正な知覚内容を概念的に捉える方法を明らかにすることにあっただから、幻覚や錯覚の場合にも内容の特定に成功するような説明は、初めから求められていないのである。我々は「その色」と言って指示に失敗することもあるが、指示の失敗はありふれたことなのである。

ここでより広い文脈に位置づけるなら、直示的概念は対象依存的な思考の一種だということが重要である。マクダウェルによれば、見知りによる知識というラッセルの概念は主体にとって不可謬な領域に制限される必要はない(McDowell 1986)。見知りによる知識によって構成されるラッセル的思考は、世界の中の対象そのものをその内容の一要素として持つ。その意味で、そのような思考は対象依存的である。見知りによる知識が（それについて誤りう

る) 日常的な対象にまで拡張されるなら、知覚によって世界の事物を認識するとき、その内容は世界にどんな対象が存在するかによって決まるという常識的な考え方を問題のないものにすることができる。そのような対象が存在しなかったなら、そのような思考はたんに成立しないだけであり、直示的概念も同様なのである。

(4) 今後の展望

以上のように、直示的概念は「その色合い」というレベルで個別化されるべき対象依存的な思考である。知覚のきめ細かさからの論証に対して、概念主義の側が直示的概念によって応答するという道は閉ざされていないし、その背景にあるデカルト的描像を克服するという動機に照らせば、正当なものであるだろう。

直示的概念による応答それ自体は、あくまでも論敵に対する反論であり、概念主義を積極的に論証するものではないと言われることがある(Coliva, 2003; Bermúdez and Cahen, 2020)。しかし、本稿のようにマクダウェルによる心と世界の関係に関する治療の細部を埋めるものとして、直示的概念が整備されるならば、この概念は概念主義の積極的な擁護に役立つだろう。

最後に、概念主義を擁護する別の方法として、知覚のきめ細かさからの論証そのものを単に拒絶することもできるかもしれないという選択肢に触れておく。本稿の冒頭において、概念主義を採ることで、知覚と信念の正当化関係を問題のないものにすることができると言った。それに対して、これまで論じてきたように、知覚のきめ細かさが重大な反対根拠になりうると思われるのである。しかし、ここで検討する方法によれば、知覚と信念の正当化関係にとって、知覚のきめ細かさはそもそも反対根拠になるようなものではない。ひとが正当化できるのは、きめの粗いレベルまででしかなく、きめの細かい内容は初めから概念的に捕捉する必要がないということになる。そのため、知覚のきめ細かさからの論証は問題にならなくなる。

しかしながら、その立場が仮に整合的だったとしても、それはもはや概念主義と呼べるものであるかは定かではない。概念的でないが正当化にも寄与しないような知覚内容というものを認めることは、知覚経験の内容は全て概念的であるわけではないとする非概念主義にむしろ近くなっているようにも思われる⁽⁵⁾。概念主義という立場を取って押し出すのであれば、本稿で展開したような知覚のきめの細かいレベルまで概念的だという概念主義の方がその名にふさわしいだろう。そのため、非概念主義者が概念主義に対する反論として知覚のきめ細かさからの論証を持ち出したのは正しいが、それに対しては直示的概念によって応答可能だということを本稿は示した。そしてそれを

踏まえれば、概念主義は依然として有効な立場であり続けるのである。

注

- (1) この三つのレベルの表現について、ピーコックそのままではなく、Kelly (2001b)を参考にしている。
- (2) ケリー自身は知覚の状況依存性という概念主義を取らない理由をもっているので、示唆しただけにとどまる。
- (3) ところで、ひとつ確認すべき修正点がある。マクダウェルが、ピーコックらの批判を承けて、『心と世界』において「その色合い(that shade)」の代わりに「そのように色づいている(is colored thus)」という表現を使うべきだった、と言っているのである(McDowell 1998, 415, 416-7)。さらに、そのように言うことで、『心と世界』においては経験の内容が命題的なものであることが明確になるだろう。ただし、我々の論点には影響してこないなので、引き続き「その色合い」を用いる。
- (4) 同様のことにケリーも懸念を表明している Kelly(2001a, 398,n.2)。
- (5) 概念主義が知覚の内容は全て概念的であるという立場だとすれば、非概念主義は知覚の内容は全て概念的であるわけではないという全称の否定になる。よって、知覚内容の一部が概念的であっても非概念的内容を認める立場は非概念主義として分類される。Schmidt(2015, ch.3)を参照。

(5) 参考文献

- Ablondi, F. (2002) “Kelly and McDowell on perceptual content”, *Electronic Journal of Analytic Philosophy*, Issue 7, URL=<<https://ejap.louisiana.edu/EJAP/2002/Ablondi.html>>
- Bermúdez, J and Cahen, A. (2020) “Nonconceptual Mental Content”, *The Stanford Encyclopedia of Philosophy*, URL=<<https://plato.stanford.edu/entries/content-nonconceptual/>>
- Coliva, A. (2003) “The argument from the finer-grained content of colour experiences: A redefinition of its role within the debate between McDowell and non-conceptual theorists”, *Dialectica: International Journal of Philosophy of Knowledge*, 57(1), pp. 57-70.
- Evans, G. (1982) *The Varieties of Reference*, Oxford: Clarendon Press.
- Heck, R.G. (2000) “Nonconceptual Content and the "Space of Reasons"”, *The Philosophical Review*, Vol. 109, No. 4, pp.483-523.
- Kelly, S. (2001a) “Demonstrative concepts and experience”, *The*

- Philosophical Review*, Vol. 110, No. 3, pp 397-420.
- (2001b) “The Nonconceptual Content of Perceptual Experience: Situation Dependence and Fineness of Grain”, reprinted in Y.H. Gunther (ed.) *Essays on Nonconceptual Content*, Cambridge, Mass: The MIT Press, 2003.
- McDowell, J. (1986) “Singular Thought and the Extent of Inner Space”, reprinted in his *Meaning, Knowledge, and Reality*, Cambridge, Mass: Harvard University Press.
- (1996) *Mind and world : with a new introduction*, Cambridge, Mass: Harvard University Press. (『心と世界』、神崎繁・河田健太郎・荒畑靖宏・村井忠康訳、勁草書房、2012年)
- (1998) “Reply to Commentators”, *Philosophy and Phenomenological Research*, Vol. 58, No. 2, pp.403-431.
- Peacocke, C. (1998) “Nonconceptual Content Defended”, *Philosophy and Phenomenological Research*, Vol. 58, No. 2, pp.381-388.
- (2001) “Phenomenology and nonconceptual content”, *Philosophy and Phenomenological Research*, Vol. 62, No. 3, pp.609-615.
- Schmidt, E. (2015) *Modestnonconceptualism: epistemology, phenomenology, and content*, Cham: Springer.
- Tye, M. (2000) *Consciousness, Color, and Content*, Cambridge, Mass: The MIT Press.

(慶應義塾大学)

判断と情動はどのように対立し、また調停されるのか
How Judgments and Emotions Conflict and are Mediated

太田泰幹

Abstract

If judgment and emotion conflict, it would seem that we should make decisions or act according to judgment and not according to emotion. However, the episode of Huckleberry Finn's dilemma introduced by Jonathan Bennett (1974) suggests that it is possible to be praiseworthy or rational to make decisions or act according to emotion rather than judgment. In this paper, I will show that to resolve the issues surrounding Huckleberry Finn's dilemma, it is necessary to elucidate how judgments and emotions conflict and are mediated. Then I will present a model that explains its mechanism.

(1) 研究テーマ

情動とは何か、というテーマは論争的であるが、現代の標準的な見解によれば、情動は表象として説明される (Scarantino 2021)。たとえば、足下にヘビが這い出てきたのを見つけ、怖くて飛び退るとき、恐怖は私にとってヘビが危険であることを表象している。また、行為の観点からみれば、恐怖は飛び退るという行為に理由を与えている (Döring 2013)。つまり、ヘビが危険であると評価したから、私はヘビから逃げたのである。このように、情動は主体と世界の関係を表象し、また行為に理由を与えるという役割をもつ。

ただし、そのような役割は情動だけでなく、判断も果たしうる。そして、情動の表象する内容や与える理由と判断の内容や理由はしばしば食い違う。たとえば、私はヘビに詳しく、這い出てきたヘビがアオダイショウで、こちらから刺激しない限り襲ってくる可能性は低く、万一噛まれたとしても無毒なため、危険はないと直ちに判断できたとしよう。このとき、〈このヘビは危険である〉という情動の内容と〈このヘビは危険でない〉という判断の内容は対立するが、両者を比較すると判断の方が正しく、またヘビから逃げる理由もないように思われる。このような考えを一般化すると、次のような主知主義的 (intellectualist) 見解になるだろう。すなわち、情動と判断が対立するときは、判断の方が内容や理由において適切なため、判断に従って意思決定や行為をするべきであり、情動に従って意思決定や行為をするべきではな

い。

〈ハックルベリー・フィンのジレンマ〉は、このような主知主義的見解に対する反例とみなすことができる。Bennett (1974) による導入以降、このエピソードは後述の通り多くの哲学者によって論じられてきた。本稿は、ハックルベリー・フィンのジレンマを巡る問題を整理したのち、それらの問題を解決するためには、〈判断と情動はどのように対立し、また調停されるのか〉を解明しなければならないと指摘する。その上で、判断と情動の関係を説明するモデルを提示したい。

(2) 研究の背景・先行研究

Bennett が導入したハックルベリー・フィンのジレンマとは、以下のようなエピソードである。小説『ハックルベリー・フィンの冒険』(以下 AHF) において、主人公のハックルベリー (以下ハック) は逃亡奴隷のジムと共に筏で旅をする。AHF 第 16 章で、ハックは逃亡奴隷を探す男たちと出会い、筏に乗っているのは逃亡奴隷ではないかと尋ねられる。作中において逃亡奴隷を匿うことは (道徳的に) 悪いとされていたため、ハックは己の良心に従ってジムを通報しようとしたが、結局嘘を吐いてジムを助ける。このエピソードにおいて、ハックは良心に基づく判断とジムに対する共感の間でジレンマに陥っている、と Bennett は指摘した。

ハックのジレンマが興味深いのは、このエピソードが〈情動と判断が対立するとき、情動に従って意思決定や行為をするべきではない〉という主知主義的見解に対する反例になっているからである。ハックは、判断に基づいて行為することに失敗し、結局情動に基づいて行為したように思われるが、ジムを救うという帰結が道徳的に間違っているとはいえない。このエピソードを、先行研究は大きく二つの観点から論じてきた。第一の観点は、〈ハックの行為は称賛に値するか (praiseworthy)〉である。これは、ある行為に道徳的な価値があるとはどういうことかを巡る問いであり、代表的な仕事として Arpaly (2002) が挙げられる。第二の観点は、〈ハックの行為は合理的か〉である。これは意志の弱さ、あるいはアクラシア (akrasia) に基づいた行為は合理的でありうるかを巡る問いであり、McIntyre (1993) の影響が大きい。

いずれのテーマも重要かつ論争的であり、本稿において直接答えることは適わない。本稿の関心は、むしろ両テーマの前提にある。先行研究は、ハックの良心に基づく判断とジムに対する共感の対立を前提とし、ハックの行為と道徳的価値や合理性の関係を主題的に論じているが、本稿は、そもそもハックの良心に基づく判断とジムに対する共感が対立するとはどういうことか、

という問いに焦点を当てる。これは〈判断と情動が対立するとき、(道徳的に
であれ合理的にであれ)どちらに従うべきか〉という規範的な問いではなく、
〈判断と情動はどのように対立し、また調停されるのか〉という記述的な問
いである。本稿は、判断と情動の対立と調停を説明するモデルを提示するこ
とで、この問いに答えたい。

(3) 筆者の主張

本節では、ハックの判断と情動の関係を説明するモデルを提示する。まず、
ハックのジレンマではハックの判断と情動が対立しているとされているが、
そもそもなぜこのような対立が可能なのかは説明を要するだろう。比較のた
め、判断同士の対立を考えてみたい。たとえば、私が足下のヘビをアオダイ
ショウと見て〈このヘビは危険でない〉と判断したにもかかわらず、なぜか
同時に〈このヘビは危険である〉と判断したとする。このとき、両判断の内
容は対立しているというより、矛盾している。ハックの判断と情動の対立も
同様の矛盾なのであれば、そのような心的過程が行為の道徳的価値に関係す
るとは思われないし、導かれた行為の合理性を問うこともできないだろう。

しかし、Döring に従えば、ハックの判断と情動の対立は矛盾なき対立
(*conflict without contradiction*)として合理性の枠内で説明できる。Döring
によれば、判断と情動では内容に対するコミットメントの有無が異なる
(Döring 2013: 182-4)。すなわち、判断はその内容、たとえば〈このヘビは
危険でない〉に対するコミットメントがあるのに対し、情動はその内容、た
とえば〈このヘビは危険である〉に対するコミットメントをもたない。たし
かに、私がどれだけ平静を装って〈このヘビは危険でない〉と判断したとこ
ろで、怖いものは怖いのであり、その現れは否定しがたいだろう。しかし、
ヘビが怖いからといって、私は〈このヘビは危険である〉という内容にコミ
ットしているわけではないので、判断と情動は矛盾なく対立しうるのである。

判断と情動が矛盾なく対立しうることを確認できたところで、ハックのジ
レンマにおける判断と情動の内容を整理しておこう。ハックのジレンマとは、
ハックの良心に基づく判断とジムに対する共感の間における葛藤であったか
ら、両者の内容を特定すればいい。先に、共感の内容を確認しよう。ハック
のジレンマを導入した Bennett によれば、共感とは仲間意識の感じ (*fellow-
feeling*) である (Bennett 1974: 124)。そこで、本稿はハックの共感を〈ジ
ムは仲間である〉という内容をもつ情動として扱う。次に、ハックの共感と
対立する判断の内容とは何か。ハックは己の良心に従ってジムを通報しよう
としていたので、その内容は〈ジムを通報すべきである〉となるだろう。一

見、〈ジムを通報すべきである〉という内容と〈ジムは仲間である〉という内容は対立しないように思われるが、〈ジムを通報すべきである〉という内容は〈ジムは仲間でない〉という内容を前提していると考えられる。したがって、〈ジムを通報すべきである〉という判断は、その前提において、すなわちジムが仲間か否かを巡って、共感と対立しているといえるだろう。

また、ハックの判断と行為の関係を考えるためには、判断が与える理由についての整理も重要である。ここで、2種類の判断を区別しておきたい。まず、ハックのジレンマにおいて、ハックは〈ジムを通報する〉か〈ジムを助ける〉という行為のみが可能であったとしよう。すると、〈ジムを通報すべきである〉という判断は、〈逃亡奴隷を匿うことは悪い〉や〈ジムは持ち主の元に戻る気がない〉といった事情を考慮して、〈ジムを助ける〉ではなく〈ジムを通報する〉という行為を評価していることになる。ここで、〈ジムを通報すべきである〉という判断がすべての事情を考慮した (all things considered) 判断なのか、それとも一部の事情のみを考慮した一応の (prima facie) 判断なのかを区別するために、前者を ATC 判断、後者を PF 判断とよぼう。そして、〈ハックの行為は合理的か〉を問う先行研究は〈ジムを通報すべきである〉という判断を ATC 判断とみなしている (McIntyre 1993; Döring 2013)、本稿もそのように仮定して議論を進める。

以上の整理より、ハックルベリー・フィンのジレンマを巡る問題は以下のように特定化できる。まず、ハックは〈ジムを通報すべきである〉という ATC 判断を下す。しかし、逃亡奴隷を探す男たちを前にして、その判断と〈ジムは仲間である〉という共感が対立する。結局、ハックは ATC 判断に従わず、ジムを助けた。はたして、ハックの行為は称賛に値する／合理的であるといえるだろうか。

この問題に対し、本稿はその前提に焦点を当てる。すなわち、ハックの ATC 判断と共感が対立するとはどういうことだろうか。そもそも、これは不可解な対立である。〈ジムは仲間である〉というハックの共感は、〈ジムを通報する〉ではなく〈ジムを助ける〉という行為に理由を与える。しかし、〈ジムを通報すべきである〉という判断が ATC 判断であるならば、〈ジムは仲間である〉という事情も考慮した上で、ハックは〈ジムを助ける〉ではなく〈ジムを通報する〉という行為を評価したはずである。それにもかかわらず、なぜハックはジムを助けたのだろうか。このように、ジムを助けるというハックの応答が出力されるまでのメカニズムは不可解であり、〈ハックの行為は称賛に値するか／合理的か〉という問いを考えるためにもその解明が求められる。そこで、まずは一般に判断と情動はどのように対立し、また調停される

のかを説明するモデルを作り、その上で ATC 判断と情動が対立するとはどういうことかという問いに答えよう。代表的な情動理論を提唱し、判断との関係についても示唆的なプリンツ（2004）を本稿は参照する。

まず、判断と情動は矛盾なく対立しうるのであった。それでは、判断と情動が対立したとき、その対立はどのように調停されるのか、すなわち、主体がどちらの内容にコミットするようになるかはどのように決定されるのか。調停の基準として、まず情動の強度が挙げられるだろう。判断と対立する情動が無視できるほど弱ければ、主体は判断の内容にコミットし続ければいいし、情動が極めて強ければ、主体は判断の内容にコミットし続けることが難しくなるだろう。一方で、判断と対立する情動が無視できるほど弱いわけでも極端に強いわけでもなければ、もはや対立を調停する基準はないように思われる。しかし、後述する通り、プリンツによれば情動は判断によって〈再較正 (recalibration)〉されうる、すなわち、判断を通じて引き起こされるようになりうるという。もし判断が情動に影響を与えうるのであれば、その関係によって判断と情動の対立が調停されるのかもしれない。そこで、プリンツの議論を確認してから、判断と情動の対立が調停されるメカニズムを検討しよう。

判断を通じて情動が引き起こされるようになることがあるとはどういうことか。例として、AHFにおける〈へびの皮に対する恐怖〉を取り上げる。AHF第10章において、なぜかジムはへびの皮に触れることは縁起が悪いと信じ、恐れている。そんなジムの迷信をハックは馬鹿にしていたが、たしかにハックがへびの皮に触れたのち、ジムがへびに噛まれたり、筏が予定通りに進まなかったり、挙句の果てには蒸気船に轢かれたりしたために、やがてハックもへびの皮を恐れるようになった（AHF第16章）。このエピソードは、以下で示す通り、〈へびの皮に触れることは危険である〉という判断によって恐怖が再較正された、と解釈することができる。まず、プリンツによれば、一般に情動は特定の状況によって信頼のおける仕方で引き起こされ、その状況に関する何かしらの情報を担う機能をもつ（Prinz 2004: 52-4）。しかし、恐怖はへびの皮に触れることによって信頼のおける仕方で引き起こされたりはしないし、それが危険であるという情報を担う機能もないだろうから、へびの皮に対する恐怖はそのような意味での情動とはいえないだろう。ただし、プリンツは、ある判断が何かしらの情動的な感じを信頼のおける仕方で引き起こすまでに強く情動と結びつくとき、情動は判断によって再較正される、とも主張している。つまり、情動は判断によって、元々引き起こされていた状況とは異なる状況でも引き起こされるように調整されうるのである（Prinz 2004:

99)。この考えをさきほどのエピソードに当てはめると以下のようになる。ハックがヘビの皮に触れたあと、ジムがヘビに噛まれたことで、ハックは〈ヘビの皮に触ることは危険である〉と判断し、また恐怖の感じを覚えた。さらに、その後も繰り返し危険が訪れたことで、〈ヘビの皮に触ることは危険である〉というハックの判断は、やがて恐怖の感じを信頼のおける仕方で引き起こすほど強く恐怖と結びつき、恐怖を再較正したのである。

この〈再較正〉という仕組みを参照することで、判断と情動の対立が調停されるメカニズムを説明できると本稿は主張する。調停のパターンは、大きく2通りに分けることができるだろう。まず、主体は判断を修正するだけで対立を調停できるかもしれない。数学に苦手意識をもっている、文学専攻の学生を例に考えてみよう。その学生は文学を学ぶつもりなので、数学の能力向上は自分の成長と関係がないと判断している。しかし、単位の都合で数学の授業を履修したところ、その勉強に喜びを感じていることに気がついた。ここで、やはり数学の能力向上は自分の成長と関係があると判断を修正するならば、数学の勉強に対する喜びとの対立は調停される。このように調停できる対立を〈浅い対立〉とよぼう。

しかし、主体が判断を修正しても、その判断とさらに対立する情動があるかもしれない。授業の内容が少し難しくなっただけでまったく分からなくなってしまえば、数学への苦手意識が蘇り、嫌悪感が生じるだろう。このような場合、判断を修正しようがしまいが対立する情動が存在するので、情動に変化が生じない限り、もはや対立は調停されなくなってしまう。しかし、判断は情動を再較正しうるのであった。理解できなかった数式が理解できるようになったり、数学と文学に関係性を見出したりすることで、数学の能力向上は自分の成長と関係があるという判断は強く喜びと結びつくようになるだろう。そして、判断によって再較正されたこの喜びが、判断と対立する嫌悪よりも強ければ、その対立は情動の強度という基準によって調停可能になると考えられる。このようにしか調停されない対立を〈深い対立〉とよぼう。

以上のような、判断と情動の対立と調停のフローは、図1のように表せる。これを〈判断と情動の循環モデル〉とよぼう。

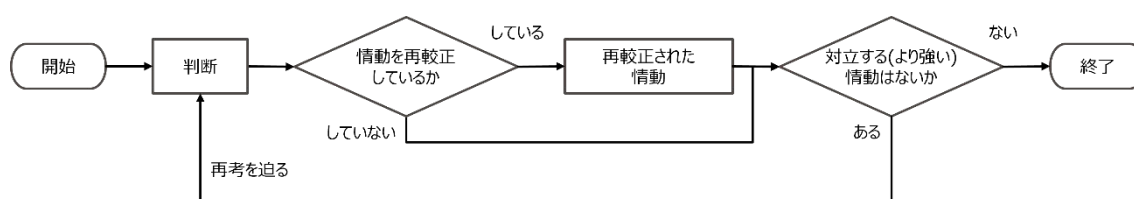


図1 判断と情動の循環モデル

これで、判断と情動はどのように対立し、また調停されるのかを一般的な仕方で説明するモデルが得られた。そして、PF判断やATC判断も判断の一種であることから、循環モデルは両者にも適用可能である。まず、PF判断と対立する情動があるということは、その情動が与える理由を主体が考慮する余地が残されているということである。そのため、ある評価的な判断がすべての事情を考慮したといえるためには、浅い対立であれ深い対立であれ、すべての対立が調停されていなければならない。したがって、ある判断がATC判断であるといえるためには、その判断と対立する情動は存在しないか、存在するとしても、その情動は判断が引き起こす再較正された情動よりも弱くなければならない。以上の整理から、ATC判断と情動が対立するとはどういうことか、という問いに説明を与えることができる。すなわち、情動と対立する判断がATC判断であるといえるためには、その情動よりも強い再較正された情動を判断が引き起こさなければならない。

最後に、循環モデルによってハックのジレンマを説明しよう。まず、〈ジムを通報すべきである〉というハックの判断はATC判断であると仮定する。この判断は、その前提において〈ジムは仲間である〉という共感と対立する。対立が調停されるためには、ハックが判断を修正するか、共感よりも強い再較正された情動が引き起こされなければならない。しかし、ハックは判断を修正したわけでも、共感より強い再較正された情動、たとえば正義感などが引き起こされたわけでもない。したがって、ハックの判断と共感の対立は調停されていない。しかし、情動と対立する判断がATC判断であるといえるためには、その情動よりも強い再較正された情動を判断が引き起こさなければならないのであった。よって、帰結は仮定と矛盾する。以上より、本稿は〈ジムを通報すべきである〉というハックの判断はATC判断ではなく、PF判断に過ぎない、と結論する。ハックはジムに対する共感が与える理由を考慮したわけではなく、ただ無視していたのである。

ただし、これは作中においてハックがATC判断を下さなかったということの意味しない。AHFは、もっと豊かな読みを許容する作品であるように思われる。第31章で、とうとうジムが捕まってしまい、再びハックはジムを助けるべきか否かで葛藤する。しかし、ハックは“All right, then, I’ll go to hell”と、奴隷を逃がす〈罪〉を受け入れ、ジムを助けると決意した。本稿は、この第31章でハックのATC判断が下されたと考える。このとき問題になっているハックの判断は、第16章時点の判断を修正した〈ジムを助けるべきである〉という判断である。しかし、ハックは判断を修正するだけで対立を

調停できたわけではない。なぜなら、その判断はさらに、そのようなことをすれば地獄に落ちる、という非常に強い恐怖と対立したからである。ここで、ハックのジレンマは浅い対立ではなく、深い対立であったことが分かる。それにもかかわらず、以下の解釈からハックの判断と恐怖の対立は調停されたといっているだろう。長い旅路の中で、ハックはいつしか〈ジムは仲間である〉という内容にコミットするようになった。しかも、その判断は共感の感じを信頼のおける仕方で引き起こすほど強く共感と結びつき、やがて地獄行きの恐怖さえ上回るほど強い再較正された共感を引き起こしたのである。以上の解釈より、第 31 章におけるハックの〈ジムを助けるべきである〉という判断は、地獄に落ちるといふ帰結まで考慮して、深い対立を乗り越えた ATC 判断であると本稿は主張する。

(4) 今後の展望

本稿は、判断と情動はどのように対立し、また調停されるのかを説明するモデルを提示した。一方で、本稿は（AHF 第 16 章における）ジムを助けたハックの行為は称賛に値するか／合理的かという、先行研究が焦点を当ててきた問題には答えられていない。

ただし、ハックの行為は合理的かという問いに対しては、そもそもハックの〈ジムを通報すべきである〉という判断は ATC 判断ではなかった、という積極的な含意を本稿はもつ。しかし、これは問いの重要性をまったく減じないだろう。私たちは、時間や情報など、ATC 判断を下せるだけの十分なリソースがない中で意思決定や行為を求められることがよくある。したがって、本稿はむしろそのような状況下における実践的合理性を追究する重要性を示唆している。

また、ハックの行為は称賛に値するかという問いに対しても本稿は示唆的である。行為の理由は道徳的に正しいか、その理由を与える判断は行為が称賛に値するために必要かといった道徳的価値を巡る問い (Arpaly 2002) は、まさに循環モデルとの関係において特定化されるだろう。

(5) 参考文献

- Arpaly, N., 2002, "Moral Worth," *The Journal of Philosophy*, 99(5): 223-45.
- Bennett, J., 1974, "The Conscience of Huckleberry Finn," *Philosophy*, 49(188): 123-34.
- Döring, S.A., 2013, "Emotion, Autonomy, and Weakness of Will," Kühler,

- M. & Jelinek, N. eds., *Autonomy and the Self*, Springer, 173-90.
- McIntyre, A., 1993, "Is Akrotic Action Always Irrational?," Flanagan, O. & Rorty, A.O. eds., *Identity, Character, and Morality: Essays in Moral Psychology*, MIT Press, 379-400.
- Prinz, J.J., 2004, *Gut Reactions: A Perceptual Theory of Emotion*, Oxford University Press. (源河亨訳、2016、『はらわたが煮えくりかえる——情動の身体知覚説』勁草書房)
- Scarantino, A. & de Sousa, R., 2021, "Emotion," *The Stanford Encyclopedia of Philosophy*, (Summer 2021 Edition, <https://plato.stanford.edu/archives/sum2021/entries/emotion/>).

(一橋大学)

アクターネットワーク理論の「人工主体」研究への適用

--画像生成 AI の分析を例に

Application of actor-network theory to the study of "artificial agents"

--Analyzing AI image generators as an example

池原優斗

Abstract

This research proposes Actor-Network Theory (ANT) as a methodology for organizing and analyzing problems involving "artificial agents," which are artifacts that induce humans to experience them as a certain kind of agent, such as AI and robots. ANT makes it possible to handle the boundaries of the categories of human, "artificial agent" and "other artifacts" in a dynamic and flexible way. This enables effective organization and analysis of problems involving "artificial agents." After presenting an overview and discussion of ANT, this paper will present how ANT can be used concretely to describe and analyze problems involving "artificial agents" through an analysis that applies ANT to the problem of AI image generators.

(1) 研究テーマ

現在、AI やロボットの社会実装が進んでいる。それと同時に AI やロボットという新しい存在の登場は、例えば、AI が法的責任を持つことができるかというような、人間とモノの境界を揺るがす新たな問題を生み出している。本論では、AI やロボットのような人間がそれをある種の主体として経験することを誘発するような人工物を「人工主体」と呼ぶ¹²。「人工主体」は AI やロボットに限定されるものではなく、定義を満たすのであれば、人工生命や自動運転車など様々な人工物を含めることができる。本研究では、「人工主体」および、それが関わる問題について記述・分析する方法論としてアクターネットワーク理論 (ANT) を提示し、それを用いた「人工主体」が関わる問題の記述・分析を行う。

(2) 研究の背景・先行研究

ロボット倫理学の分野では、ロボットを主体として扱うべきか、道具として扱うべきか、という議論が行われている。例えば、久木田 [2020 : 18] は

ロボット倫理学の課題として、次のような問題を挙げている。ロボットが人に危害を与えた際の責任の帰属はどのようにすればよいのか、人間にそっくりで人間のように振る舞うロボットに人格や人権は認められるべきか、ロボットが戦場で戦うとき人を殺傷する権限を与えるべきか、ロボットが人間の仕事を奪うようになったらどのように対応すべきか、といった事柄である。こうした論点はどれもロボットという人工物を人間と同じような主体として扱うべきなのか、という問題に行き着くものである。また、呉羽 [2021: 63] もロボット倫理学におけるロボットを道具以上のものとして扱うべきかという問題を踏まえて、社会ごとの文化による差についてどのような態度で考えていくべきかという観点からの議論を展開している。

人工物の主体性に関する問題は、ロボットだけでなく、人工知能 (AI) についても同様に存在する。例えば、2022年6月、グーグルのAI開発部門の社員が、グーグルの自然言語生成AIである「LaMDA」について、意識や感覚を持っているとし、グーグルに対して実験に使用する前に同意を求めべきであると主張した出来事があり、議論を呼んだ [ヴァランス 2022]。人間がAIに対して主体性を感じるような現象は、自然言語生成AIだけではなく画像生成AIでも報告されている。Daras & Dimakis [2022] は、画像生成AIのDALI-2は生成した画像に登場した意味をなさない文字列が、AIにとって特定の対象を指示する単語である可能性を指摘している。Daras & Dimakis [2022] で報告された現象はAIが意識を持っている根拠とはなり得ないが、人間が画像生成AIのこのような振る舞いから主体性を経験することは十分にあり得るだろう。

このような、人間がそれをある種の主体として経験することを誘発するような人工物全てを含む概念が、「人工主体」である。「人工主体」という概念を用いることで、AIやロボットなどの私たちが主体性を経験する現代の技術的存在について、統一的に議論していくことができるようになる。「人工主体」の概念は、直観的であり、議論のきっかけとする上で有効なものだと言えるだろう。

しかし、「人工主体」についての分析を進めようとする、何が「人工主体」なのかを定める境界線が曖昧で、問題の整理や分析が困難になるという問題がある。「人工主体」という概念を導入した際、人間か人工物として表現されていた存在は、人間／「人工主体」／「その他の人工物」に分けられることになる。このとき、どこからが「人工主体」でどこからが「その他の人工物」であるかをはっきりと決めることは難しい。さらに、AI等の技術がより発達し社会実装が進んでいけば、人間と「人工主体」の境界線も徐々に曖昧なも

のようになっていこう。本論は、この問題の解決を試みるものである。

(3) 筆者の主張

3-1 アクターネットワーク理論

ANT はブルーノ・ラトゥール [ラトゥール 2019] やミシェル・カロン、ジョン・ロー [カロン・ロー 1999] らによって発展した理論である。ラトゥール [2019: 24-6] による、ある研究が ANT であるかの暫定的基準を基に、概要を端的に示すならば、第一に、「非人間にはっきりとした役割が与えられている」こと、第二に、社会的なものを最初から安定させたまま、社会的なものによって、ある物事を説明しないこと、第三に、脱構築ではなく、「社会的なものを組み直すことを目指している」ことが ANT の特徴と言える。また、存在物を異種混合の集団の結果として捉え、またその行為を集団的な性質であるとする点も ANT の特徴である [カロン・ロー 1999: 249-54]。

本論では ANT の議論のうち、特に次の二つを取り上げたい。一つは、人間だけでなく、生物やモノである非人間もアクターとして捉えるという議論。つまり、ANT は人間と非人間を対称的に扱うのである。もう一つは、エージェンシーについての議論である。ANT においてエージェンシーは、人間および非人間であるアクターの中にあるというよりも、アクターが織りなす布置から生まれてくる [青山 2008: 177]。したがって、ANT は分析の対象となる事柄について、それを人間と非人間が含まれるアクターが関係し合っているネットワークとして扱い、そして、エージェンシーはそのネットワークの布置連関から生まれてくると捉えるのである。

ANT は事象を記述し、分析することに長けたフレームワークであるが、ある対象に対して批判的態度を持ち続けるにはどのようにすればよいのか、といった規範的な議論への応用も検討されている。以下で取り上げる土橋 [2015] は、ケータイ的な情報環境における「批判」を ANT によって考察したものである。土橋 [2015] は、ケータイという可動的で可変的なメディアと人間はどのように関係を取り結び、人間はそうした技術に対して批判的な姿勢をどのようにすれば持つことができるのか、という問題に対して、ANT の視点から分析している。

土橋 [2015: 19-22] は、可動的かつ可変的なケータイが常にそばにある情報環境を、デスクトップパソコンのような固定的な空間に設置されるメディアによる情報環境と対置する。そして、アドホックなニーズに柔軟に応じるため利用者の生活スタイルや価値観と摩擦を起こさずに利用者の文脈に適合するケータイを、人間と対峙する固定的メディアと異なり、人間と協働する

メディアであるとする [土橋 2015: 21-2]。ANT を用いて人間とケータイの協働を捉えれば、人間とケータイはハイブリッド・コレクティブであり、エージェントは人間とケータイと空間の関係的な布置連関から、関係的な効果として立ち上がってくるのだとする [土橋 2015: 22-4]。

ただ、土橋はケータイとの関係が協働であったとしても、ケータイ的な情報環境への批判的な構えを持ちづける必要があるとする [土橋 2015: 31]。しかし、ユビキタス化とアプリケーション化によって特徴付けられるケータイ的な情報環境においては、技術に対する批判的構えを持つための距離が存在せず、また、その都度異なるエージェントが立ち現れてくるため、一貫した論理によって批判するための根拠が存在しない [土橋 2015: 30-1]。土橋 [2015: 32-3] は批判のための「距離」と「根拠」の不在に対して、外部ではなく、内部から目の前にあるケータイをめぐる布置連関に思考や実践を「付加」し、その布置連関を組み替えていくような批判のあり方を提案している。

土橋 [2015] が扱ったケータイ的な情報環境において起きている問題は、正にケータイ（スマートフォン）を含む移動するコンピュータのアプリケーション上に現れ、ユビキタス的に存在する「人工主体」にも共通するものだろう。更に言えば、土橋 [2015] において問題とされているのは、情報技術が否応なく実装され浸透される世界で、どのように批判的態度を持ち、批判的な受容が可能であるかということであり、より広範な問題に対して、ネットワークの内部から思考や実践を付加するという批判のあり方は有効であるかもしれない。いずれにせよ、内部から布置連関を組み替えることによる批判に着目することは、ANT の記述的分析を「人工主体」の具体的問題に応用し、実際性のある議論を展開するための一つの有効な方法となりうる。

本論では、人間と非人間を対称的に扱う ANT を、「人工主体」が関わる問題の整理・分析のための方法論として提案する。つまり、本論の主張は、ANT を用いることで、人間／「人工主体」／「その他の人工物」という区分そのものを問い直す形で、境界線の曖昧さという問題を乗り越え「人工主体」が関わる問題を整理し分析することが可能となるということである。

次の節では、ANT を用いて画像生成 AI について記述・分析する。この画像生成 AI についての考察では、主に「人工主体」と「その他の人工物」の境界の曖昧さに焦点を当てる。また、土橋 [2015] が扱った、情報技術が否応なく実装され浸透する世界で、どのように批判的態度と受容が可能であるかという観点からも ANT を用いて記述・分析する。

3-2 画像生成 AI について ANT を用いて記述・分析する

2022 年は複数の影響力がある画像生成 AI が公開された。2022 年 4 月 6 日に DALL-E-2 が発表された [OpenAI n.d.] のを皮切りに、7 月 12 日に Midjourney のオープンベータが開始された [@midjourney 2022]。そして、8 月 22 日に Stable Diffusion が一般に公開された [Stability AI 2022]。画像生成 AI は、従来は人間にしかできないと思われていた創造的な絵を描くことができる。そのような画像生成 AI の登場は作家や、創造性を人間が犯されることのない領域だと考えていた人々に衝撃を与え、イラストレーターの仕事が AI によって奪われるのではないかという不安を抱かせた。Stable Diffusion を開発した Stability AI の CEO がキャリアを心配する若いアーティストに対して、「イラストやデザインの仕事は、とても退屈な仕事だ。芸術的であるかどうかではなく、あなたは道具なのだ」、「画像生成 AI は今後大きく発展していく。お金を稼ぎたいなら、新しい技術からチャンスを見つける方がずっと楽しくなるはずだ」というメッセージを発して物議を醸した [松浦 2022]。また、日本においては、画像生成 AI サービスの mimic について、一部の作家から悪意のある第三者が勝手に画像を使用することを止められないという反発があり、サービスを一時停止したという出来事があった [高木 2022]。

ANT の枠組みでこの状況を考えるならば、2022 年現在は、画像生成 AI という新しいアクターの登場によって、従来のイラストを書くことに関するネットワークが組み替えられようとしている最中であると言える。また、自由にパソコンやスマートフォンから使える画像生成 AI は、アプリケーション的かつユビキタス的であり、上に述べた「人工主体」および土橋 [2015] で論じられている「ケータイ」の特徴と同様である。したがって、画像生成 AI も批判のための「距離」と「根拠」が不在であるという問題を抱えていると言えよう。

画像生成 AI の登場は、イラストを描くことに関する既に安定したネットワークに存在していたイラストレーターにとって、不安を抱かせ、著作権や職種そのものに対する脅威として存在するようになった。このとき、画像生成 AI は正にイラストレーターに脅威をもたらす「人工主体」として存在していたと言えるだろう。

しかし、一部のアクターは画像生成 AI と「対峙」して、拒絶するのではなく、別の関係のとり結び方を模索し始めた。例えば、Midjourney を用いたイラスト背景の作成方法を複数の作家が研究していたことが記録されている [@golden_haniwa 2022]。また、画像生成 AI を用いて漫画を書く実践は池田・柳澤 [2022] の例など複数行われている。そうした漫画の制作者が、熟

練した作家が絵を描くようには自由に描けないという苦勞を語っているケースもあり、このような実践を通して、徐々に画像生成 AI が差し迫った脅威にはならないことが明らかになっていった。

このような「協働」は、めまぐるしく変化する画像生成 AI が含まれるイラストを描くことに関するネットワークのただ中で行われた、「批判的実践」である。そしてこの、「批判的実践」の「付加」は画像生成 AI に関する布置連関を組み替え、画像生成 AI という存在をイラストレーターの職に関する差し迫った脅威から、創作活動での活用の可能性があるツールのような存在に変化させた。これは土橋 [2015] が提案した、ネットワークの内部から批判を付加する実践の一つであると言えるだろう。

さらに、この画像生成 AI という存在の変化は、「人工主体」と「その他の人工物」の区別を固定したものとして捉えない ANT の視座によって描きやすくなった変化である。画像生成 AI は、当初はイラストレーターにとって、「すべてのイラストレーターの仕事を奪っていく万能なイラストレーター」というような「人工主体」としての存在であった。しかし、それが徐々にデジタルイラストソフトのツールのような「その他の人工物」に近い存在へと変化していった。

3-3 流動的な「人工主体」の存在のしかたを捉える

上の例では、画像生成 AI が「人工主体」として存在するかは流動的であり、それらが存在するネットワークによって存在のしかたが変化した。これは人間／「人工主体」／「その他の人工物」という区分で対象を固定的に捉えていては記述することのできないものである。この「人工主体」の流動的な存在のしかたを捉えられることこそが ANT を方法論として導入する利点であると考えられる。

(4) 今後の展望

本研究の課題について二つ述べておきたい。まず、本論では ANT の中でも人間と非人間を対称的に扱うことと、エージェンシーに関する議論を中心に用いて記述・分析を行った。しかし、ANT にはより多くの理論的な要素があり、それらを十分に活用できているとは言いがたい。したがって、ANT の持つ各理論的要素を俯瞰し、総合的に「人工主体」について考察していくことが課題となる。もう一つは、人間は技術に対して批判的であるべきかどうかそれ自体を ANT によって説明する必要があるかを検討し、そうであるならば、それを ANT で説明することである。本論では、土橋 [2015] で提示

された技術に対して「批判的な構え」を持つべきであるという規範を、前提として議論した。しかし、よりラディカルに考えるならば、なぜ人間というアクターが、別の特定のアクターである技術（を活用したメディア等）に対して、「批判的な構え」を持つ必要があるのかという問題が生じる。「批判的な構え」を持つ必要性自体も、アクターのネットワークの関係性の中から説明される必要があるかもしれない。この点についてより研究を進めることも本研究の持つ課題の一つである。

注

1. 「人工主体」についての研究はまだ始まったばかりだが、既に複数のプロジェクトが実施されている。2020年度には戦略的創造研究推進事業のプロジェクト企画調査として「人工主体の創出に伴う倫理的諸問題を分析・討議するプラットフォームの構築に向けた企画調査」が行われた[科学技術振興機構 2022]。また、2022年度からはトヨタ財団の助成対象プロジェクト「人間と人工主体の共存のあるべき姿を学際的に問うための新たな枠組み『人工主体学』の構築に向けて」が開始されている[トヨタ財団 日付不詳]。
2. 本論の「人工主体」の定義は、科学基礎論学会 2022年度秋の研究例会ワークショップ「AIに主体性を帰属させること：学際的アプローチの中間報告」の要旨である宮原[2022]における定義を参考にした。

(5) 参考文献

- 青山征彦 2008「アクターネットワーク理論が可視/不可視にするもの--エージェンシーをめぐって」『駿河台大学論叢』35: 175-185.
- 池田佳恋・柳澤あゆみ 2022「AIが絵を描く？ 進化する画像生成 AIの最前線」『NHK』10月8日 <https://www3.nhk.or.jp/news/html/20221008/k10013851401000.html> 2022年11月4日閲覧.
- ヴァランス, C. 2022「AIに『感情がある可能性』 グーグルのエンジニアが主張」『BBC ニュース』6月14日 <https://www.bbc.com/japanese/61793898> 2022年12月8日閲覧.
- 科学技術振興機構 2022「JSTプロジェクトデータベース — 人工主体の創出に伴う倫理的諸問題を分析・討議するプラットフォームの構築に向けた企画調査」<https://projectdb.jst.go.jp/grant/JST-PROJECT-20343952/> 2022年12月11日閲覧.
- カロン, M. & J. ロー 1999 (1997)「個と社会を超えて--集団性についての

- 科学技術社会論からの視座」林隆之訳 岡田猛ほか編『科学を考える--人工知能からカルチュラル・スタディーズまで 14 の視点』北大路書房 pp. 238-257.
- 久木田水生 2020「ロボットの倫理--友達ロボットから殺人ロボットまで」『日本ロボット学会誌』38(1): 18-22.
- 呉羽真 2021「日本人とロボット--テクノアニミズム論への批判」『Contemporary and Applied Philosophy』13: 62-82.
- ゴールデンハニワ@Eve ghost enemies 発売! (@golden_haniwa) Together 投稿 2022年8月14日「キャラを自分で描ける人によって『Midjourney』にマンガやイラストの背景を描かせる研究が着々と進んでいる模様」<https://together.com/li/1930520> 2022年11月4日閲覧.
- 高木克聡 2022「イラスト作画AI、作家の敵か味方か 画風学習し新たな絵 盗用を恐れ炎上」『産経新聞』9月13日 <https://www.sankei.com/article/20220913-64SABJB6PBPLPFPPRKNKD2GJ64/> 2022年10月9日閲覧.
- 土橋臣吾 2015「移動するモノ、設計される経験--ケータイの可動性と可変性をめぐって」『マス・コミュニケーション研究』87: 17-35.
- トヨタ財団 日付不詳「助成対象詳細 | 公益財団法人トヨタ財団」<https://toyotafound.secure.force.com/psearch/JoseiDetail?name=D21-ST-0012> 2022年12月10日閲覧.
- 松浦立樹 2022「「イラストやデザインの仕事は退屈」--Stable Diffusion 開発元の代表インタビュー記事が話題」<https://www.itmedia.co.jp/news/articles/2210/26/news183.html> 2022年11月4日閲覧.
- 宮原克典 2022「AIに主体性を帰属させること--学際的アプローチの中間報告」https://phsc.jp/dat/rsm/20221020_01.pdf 2022年12月10日閲覧.
- ラトゥール, B. 2019 (2005)『社会的なものを組み直す--アクターネットワーク理論入門』伊藤嘉高訳 法政大学出版局.
- Daras, G. & A. G. Dimaki. 2022. Discovering the Hidden Vocabulary of DALLÉ-2. arXiv preprint arXiv:2206.00169.
- Midjourney(@midjourney) Twitter 投稿 2022年7月13日 午後3:41 <https://twitter.com/midjourney/status/1547108864788553729> 2022年11月4日閲覧.
- OpenAI. n.d. Timeline. <https://openai.com/timeline/> 2022年11月4日閲覧.
- Stability AI. 2022. Stable Diffusion Public Release. <https://stability.ai>

[/blog/stable-diffusion-public-release](#) 2022 年 11 月 4 日 閲覧.

(北海道大学)

総称文の表出主義的意味論
An Expressivist Semantics for Generics

水谷亮介

Abstract

The purpose of this paper is to propose a semantics for generics based on expressivism. The meaning of generics is not clear, despite the simplicity of its surface grammatical structure. After reviewing some problems with Asher et al.'s possible world semantics and Leslie's disquotational semantics, I argue that expressivism is useful for understanding the meaning of generics. I integrate Asher et al.'s semantics and Leslie's suggestion and apply a framework of the plan expressivist semantics proposed by Gibbard to a semantics for generics.

1 研究テーマ

本稿の目的は、総称文の意味論として、表出主義に基づく意味論を新たに提案することにある。総称文とは、集団を漠然と一般化して事柄を述べる文であり、「犬は哺乳類だ」「アヒルは卵を産む」「ダニはライム病を媒介する」のような文がその典型例として挙げられる¹。

総称文の論理形式については比較的、研究者間で比較的合意が取れている[5]。すなわち「 K は F だ (K s are F)」という形の総称文は、「 $Gen\ x\ [Kx]\ [Fx]$ 」という論理形式を持つとされる。「 Gen 」は量化子であり、変項 x を束縛する。「 K 」は制限域であり、変項 x の動く領域を示している。「 F 」は作用域であり、変項 x の性質を述べる部分である。よって、「 $Gen\ x\ [Kx]\ [Fx]$ 」を読み下せば、「 K であるような総称的な x について、 x は F だ」というふうになるだろう。

問題は、「総称的な (Gen)」とはどういう意味かである。総称文の意味論を構築するのは、非常に難しいことが知られている。総称文は、量に関する含意が一定ではなく、その意味するところが曖昧であるように見えるからである。例えば、文「犬は哺乳類だ」は、例外なくすべての犬が哺乳類であることを表している。他方、「アヒルは卵を産む」は、例外なくすべてのアヒルが卵を産むことを意味しているわけではない。現実にはたかだか半数程度のアヒルが卵を産むだけであるのに（卵を産むアヒルはメスのアヒルだけである）、「アヒルは卵を産む」は真と見做される。このように、総称文は、量に関する含意が個々の文によって異なっており、したがってその真理条件も個々の文によって異なっているように見える。それゆえに総称文は、見た目の単純さとは裏腹に、意味論的に分析するのが困難なのである。

2 研究の背景・先行研究

総称文の意味論に対するアプローチのなかでも広く知られているものの一つは、アッシャー、モロー、ペルティエらが一連の論文 [1, 7] において提示した意味論である。アッシャーらの意味論は、総称文を一種の様相文と見做して可能世界意味論を適用したものであり、次のように記述される²。

$$(1) \quad [[Gen\ x\ [Kx]\ [Fx]]]^w = 1 \text{ iff } \forall x \in D: \forall w' \in N(w, \|Kx\|): [Fx]^{w'} = 1$$

直観的には、この意味論は、次のように述べている。すなわち、「KはFだ」が真なのは、各々の対象について、それがKであるとき普通成立していることがすべて成立しているどの世界でもそれがFであるとき、かつそのときに限る、と。単純化して言えば、「KはFだ」が真なのは「どのxについても、xがKならば、普通、xはFである」ということが成り立つちょうどその場合である、とこの意味論は分析していることになる。

確かにこの意味論だと、「カラスは黒い」のような総称文の意味は適切に捉えられている。アルビノのカラスが存在することから、すべてのカラスが黒いわけではないことは明らかである。それにもかかわらずわれわれは、「カラスは黒い」を真と見做す。このように総称文は例外許容的だが、この重要な特性を、アッシャーらの意味論はうまく捉えている。この意味論によれば、「カラスは黒い」の真理条件は、どのxについても、xがカラスならば普通xは黒い、ということである。「普通」という条件が付されていることにより、アルビノのカラスのような例外は許容される。だから、アルビノのカラスがいるとしても、やはり「カラスは黒い」は真なのである。

しかしこの意味論には欠陥があるとの指摘が為されてきた。ここではレズリーの批判 [4] を取り上げよう。レズリーによれば、アッシャーらの意味論は (2) のような文の意味を適切に予測・説明することができない [4, p. 8]。

$$(2) \quad \text{アヒルは卵を産む。}$$

この文は真であるように思われる。しかし、このことをアッシャーらの意味論ではうまく予測できない。現実世界 w において (2) が真であるためには、任意の対象 x について、 $N(w, \|duck(x)\|)$ に属するすべての世界で、「 x は卵を産む」ということが成り立たねばならない。このとき、 x がオスであるような世界は $N(w, \|duck(x)\|)$ に属していないと考えられる。 x が卵を産みかつオスであるということは、 x がアヒルであるとき普通成立していることではないだろうからである（オスのアヒルが卵を産むような世界は、普通とい

うよりむしろ異常だろう)。しかしだとすると、 $N(w, \|duck(x)\|)$ に属するすべての世界において、「 x はメスである」が真であることになり、そうすると、総称文の真理条件に従えば、「アヒルはメスである」が w で真であるという予測が成り立つことになる。これは直観に反する。

こうした問題を受け、レズリーが代替案として提出するのは、引用符解除的な意味論である [4, p. 44]。つまりレズリーによれば、総称文の真理条件は (3) のように記述される。

(3) 「 $Gen\ x\ [Kx]\ [Fx]$ 」が真 iff $Gen\ x\ [Kx]\ [Fx]$

要するにレズリーは、意味論においては「 Gen 」を原始概念と見做してメタ言語内で使ってよいし、またそうする以外には「 Gen 」を意味論的に扱う方法はないと考えるのである。だが、なぜレズリーはこのような結論に至ったのか。

レズリーの議論は以下のように展開されている。レズリーはまず「総称文獲得のパラドックス」と自身が命名する次の問いを提起する。

(4) **総称文獲得のパラドックス**

子供は総称文を、他の量化文よりも早い時期に習得してしまう。だが音声として現れることもなく、また（理論家がただちには意味論を構築できないほどには）意味も複雑であるはずの「 Gen 」を、子供はいかにして習得しようというのか。

レズリーは (4) を解消する手立てとして、認知システムと言語との関連に着目する。レズリーの仮説によれば、「 Gen 」は、「認知システムが持つ一般化のデフォルトモード (the cognitive system's default mode of generalizing)」を起動させるものである [4, p. 23]。「認知システムが持つ一般化のデフォルトモード」(以下、レズリーに倣いこれを単に「デフォルトメカニズム」と呼称することにする)とは、生まれながらにしてヒトの脳に備わっている基本的・原始的な情報収集メカニズムのことである。このメカニズムがはたらくために、われわれヒトは少数の例の観察から一般則を導く——帰納的推論をする——傾向を持つことになる。総称文の真理条件が一見複雑に見えるにもかかわらず、子供が総称文を獲得するのが容易なのは、子供の脳には生まれながらにしてこのデフォルトメカニズムが備わっており、「all」「some」などの量化子がない場合にデフォルトで補われる量化子「 Gen 」の意味の計算にこのデフォルトメカニズムが用いられるからである。このように考えれば総称文獲得のパラドックスは解消されると、レズリーは論ずる。

レズリーによれば、総称文の真理条件の複雑さは、デフォルトメカニズム

の複雑さに起因している。デフォルトメカニズムの際立った特徴は、第一に、種の特徴的側面を埋めるようはたらくことである。特徴的側面とは、動物種の場合、その鳴き声、生殖方法などであり、社会種・人工物種の場合、その目的・機能である。デフォルトメカニズムは第二に、顕著な性質（典型的には、人間にとって危険な性質）については一般化を早める傾向にある。したがって、われわれが総称文「KはFだ」を真と見做す理由は、性質Fが顕著な性質だからかもしれないし、性質Fが種Kを特徴づける性質だからかもしれないし、大多数のKが性質Fを持っているために単に量的観点から見て一般化しても問題ないからかもしれない。

レズリー曰く、こうしたデフォルトメカニズムの特徴を踏まて総称文の真理条件を書き下そうとすれば、次のように選言的に書くほかない。

- (5) 「 $Gen\ x [Kx] [Fx]$ 」が真 iff 次の条件のいずれかを満たす³。
- a. FであるようなKが存在する（ただし、Fが自然種Kの特徴的側面に関係している場合）。
 - b. Fは種Kの機能ないし目的である（ただし、Fが人工物種Kや社会種Kの特徴的側面に関係している場合）。
 - c. FであるようなKが存在し、かつ他のKたちはFである傾向を持つ（ただし、Fが顕著な性質を表す場合）。
 - d. ほとんどすべてのKがFである（上記以外の場合）。

しかし、この真理条件を総称文の意味と見做すことはできないとレズリーは言う [4, p. 44]。「何かが赤いとは、標準的観察者が標準的状况でそれを観察すると赤い経験をするということだ」とする色の傾向説を、述語「赤い」の意味論に組み込もうとするのが奇妙であるのと同じく、(5)もまた総称文の意味論としては相応しくない。結局、「*Gen*」は原始概念として未分析のまま意味論の記述に用いるのが最善だ。これがレズリーの考えである。

3 筆者の主張

(5) はあまりに複雑で不自然であり、意味論としての説得力に欠けることは確かだろう。しかし、レズリー自身の立場である引用符解除的な意味論もまた、説得力に欠けるように思われる。メタ言語に「*Gen*」を残してしまうと、総称文についての論理法則が定式化できなくなるからである。「鳥は飛ぶ。キョロちゃんは鳥だ。ゆえに、キョロちゃんは飛ぶ」のような推論は、演繹的に妥当ではないものの、阻却可能的 (defeasibly) には妥当であるが、このことがレズリーの意味論では説明できない⁴。

そこで本稿では、上に掲げた難点を克服しうる意味論として、表出主義の立場に基づくものを提案したい。表出主義は、もとはメタ倫理学の理論の一つとして提唱されたものだが、現在ではさまざまな領域、例えば認識様相や曖昧述語の意味論への応用が試みられており [8, 6]、総称文に関して表出主義を適用することは決して突飛な発想ではない。

表出主義とは、文の意味を、文の発話によって表出される心的状態によって捉える立場である。ギバード [2] によれば、文「 φ 」の発話は、 φ という判断を表出する（例えば文「地球は丸い」の発話は、地球は丸いという信念を表出する）。それゆえ、文「 φ 」の意味を説明するためには、「 φ という判断」がどんな心的状態なのかを説明すれば十分だということになる。

レズリーは総称文と認知システムとの関連を探る際、基本的に「『K は F だ』と判断するとはどういうことか」を考えているが、その考え方を、ギバード的な表出主義の枠組みと組み合わせたとき、自然に示唆されるのは (6) のような表出主義的分析だろう。

- (6) 「 $Gen x [Kx] [Fx]$ 」と判断する iff 判断者のデフォルトメカニズムが性質 F を種 K に一般化する。

表出主義であればメタ言語に「*Gen*」を残さないのが、引用符解除の意味論よりも説明力がある。

しかし (6) のような意味論には問題が残されている。第一の問題点は「デフォルトメカニズムが性質 F を種 K に一般化する」というのはどういう心的状態なのか、そもそもこれは心的状態を記述したものなのかそれとも脳状態を記述したものなのか、といった部分がおおむね曖昧である点であり、第二の問題点は (6) のような記述だけでは合成的な意味論にはなっていないという点である。とりわけ第二の問題点には困難がある。そもそも一般に、表出主義に基づいては合成的意味論を構築することは難しいとされているからである（この問題はフレーゲ = ギーチ問題と呼ばれる）。

そこで、(7) のように、計画概念を使って定式化することを提案したい。

- (7) 「 $Gen x [Kx] [Fx]$ 」と判断する iff すべての x について、 x が K であるとき、 x は F であると推測する計画を持つ。

このように考えても、レズリーの洞察の重要な部分は取りこぼされない。レズリーの論考のなかで重要だったのは、われわれが総称文を真と見做すときの条件 (5) は、われわれの関心や態度に大きく依存しており、総称文の発話はその関心・態度の表出と見做せるという点にあったからである。また (7) の

ように考えれば、心的状態についてははっきりとするし、また、合成的意味論を構築することも比較的容易になる。計画概念を用いた表出主義の意味論は、ギバードがすでに定式化を与えているからである。

ギバードは、「超計画 (hyperplan)」および「超状態 (hyperstate)」という概念装置を用いて表出主義の合成的意味論の構築を試みている。超計画とは可能世界と類比的な概念装置であり、任意の可能な状況における任意の行為について、それを許可するか拒否するかのいずれかを定めたものとして定義される。可能世界が「物事がどうあるか」を完全に規定するものであるとすれば、超計画は「どの状況 C についても、 C において何をするか」を完全に規定するものであると言ってもよいだろう。そしてギバードは、可能世界 w と超計画 π との組 $\langle w, \pi \rangle$ を「超状態」と呼ぶ。超状態とは、信念・計画のいずれについても完全に意見を固めた状態のことである。超状態は、いかなる平叙文 S についても、 S が真であるかどうか聞くと、必ず「はい」か「いいえ」かで答え、「分からない」と答えることはないという状態であると理解することができる。超状態 $\langle w, \pi \rangle$ は、真理条件意味論での指標と類似のはたらきをする。真理条件意味論では、指標に相対的に文に真理値を与えることで、文の意味すなわち文の真理条件を明らかにする。これに対し、ギバード的な表出主義の意味論では、どんな超状態が文 S の発話に同意しうるのかを明らかにすることで、文 S の意味すなわち S の発話によって表出される心的状態を明らかにしようとするのである。

以上のような枠組みを用いたとき、(7) のような意味の記述は (8) のような形式意味論として定式化できることになるだろう⁵。

$$(8) \quad \begin{aligned} & \llbracket \text{Gen } x [Kx] [Fx] \rrbracket^{\langle w, \pi \rangle} = 1 \\ & \text{iff } \forall x \in D: \forall w' \in N(\pi, \|Kx\|): \llbracket Fx \rrbracket^{\langle w', \pi \rangle} = 1 \end{aligned}$$

(8) は、直観的には次のように述べている。超状態 $\langle w, \pi \rangle$ を抱く者が「 K は F だ」に同意しうるのは、各々の対象について、超計画 π に従えばそれが K であることからそれが F であることを推測する、と。きわめて単純化して言えば、「 K は F だ」と述べることは「 K を見たら F と思え」と命令するようなものだ、とこの意味論は分析するのである。

上の定義 (8) を見れば明らかなように、形式的には、表出主義による総称文の意味論は、(1) とほとんど変わらない。表出主義の意味論は、アッシャーらの定義 (1) を、真理条件を与えるものとしてではなく心的状態の同意条件を与えるものとして解釈することを提案するものだと言ってもよいだろう。しかもそれでいて (8) は、(1) に向けられていた批判を回避することができる。

(2) アヒルは卵を産む。

(2) が問題だったのは、アッシャーらの意味論においては $N(w, \parallel duck(x) \parallel)$ の解釈が「 x がアヒルであるとき (w の基準で) 普通成立していることがすべて成立している世界の集合」というものだったからである。(2) が現実世界で真であることにより、「 x がアヒルであるとき (現実世界の基準で) 普通成立していること」のうちには「 x は卵を産む」が含まれていなければならない。「 x が卵を産みかつオスである」ということは普通ではないため、「 x がアヒルであるとき普通成立していること」のうちには「 x はメスだ」という命題まで含まれていなければならない。その結果、「アヒルはメスだ」という文までも真と見做さざるを得なくなってしまうていた。これに対し、(8) の意味論では、 $N(\pi, \parallel duck(x) \parallel)$ の解釈は「〈超計画 π に従えば x がアヒルであることから推測することになること〉と整合的な世界の集合」というものでしかないので、その集合に、「 x が卵を産みかつオスであるような世界」が属していても問題ない。なぜならば、 x がアヒルであるときにわれわれが x について推測する事柄のうち、 x の雌雄に関する事柄は含まれていない、ということはあるからである (実際、通常われわれが受け入れている計画は、何かアヒルであるということからその雌雄に関することを推測するような計画ではないだろう)。したがって表出主義では、(2) が真であることから「アヒルはメスだ」という文まで真であると予測されることはない。

4 今後の展望

本稿では、総称文の表出主義的意味論の概略を描いた。今後は、アッシャーらの意味論やレズリーの意味論、またその他の立場に基づく意味論とのより詳細な比較を通じて、総称文の意味論的分析を進めるとともに、表出主義の利点・欠点をさらに浮き彫りにし、表出主義の可能性を探求するつもりである。

注

¹ 本稿が議論の対象とするのは、特徴づけ総称文 (characterizing generics) ——種に属する個々のメンバーについての一般化を表現する総称文——のみであり、「恐竜は絶滅した (dinosaurs are extinct)」のような直接種述定 (direct kind predication) については取り上げない。両者の区別については [5] を参照されたい。

² ただし各記号の意味は次の通りだとする。[.] は、可能世界に相対的に、文を引数として真理値を返す関数である。 $\parallel \cdot \parallel$ は文を引数として命題 (=文を真

とする可能世界の集合)を返す関数である(つまり、 $\|\varphi\| = \{w \mid \llbracket \varphi \rrbracket^w = 1\}$)。 D は議論領域である。 N は、世界と命題とを与えると、その命題が成り立つときにその世界から見て普通成立していることがすべて生じている世界の集合を返す関数である。また、 x は、対象 x の名前を表す。

³ なお、議論が煩雑になるのを避けるため、ここでの真理条件の記述はやや簡略化してある。

⁴ この問題についてはレズリーも自覚しており、[3]において反論している。本稿ではこの反論の成否について論ずる紙幅の余裕がない。

⁵ ただし、 N は、超計画と命題とを入力すると、その命題が成り立つときにその超計画に従えば推測することになることがすべて生じている世界の集合を返す関数であるとする。

文献

- [1] Asher, Nicholas & Morreau, Michael (1995). What Some Generic Sentences Mean. In Greg N. Carlson & Francis Jeffrey Pelletier (eds.), *The Generic Book*. University of Chicago Press. pp. 300–339.
- [2] Gibbard, Allan (2003). *Thinking How to Live*. Harvard University Press.
- [3] Leslie, Sarah-Jane (2007). Generics and the Structure of the Mind. *Philosophical Perspectives* 21 (1): 375–403.
- [4] Leslie, Sarah-Jane (2008). Generics: Cognition and Acquisition. *Philosophical Review* 117 (1): 1–47.
- [5] Leslie, Sarah-Jane & Adam Lerner (2022). Generic Generalizations. In Edward N. Zalta & Uri Nodelman (eds.), *The Stanford Encyclopedia of Philosophy* (Fall 2022 Edition). URL = <<https://plato.stanford.edu/archives/fall2022/entries/generics/>>.
- [6] MacFarlane, John (2016). Vagueness as Indecision. *Aristotelian Society Supplementary Volume* 90 (1): 255–283.
- [7] Pelletier, Francis Jeffrey & Asher, Nicholas (1996). Generics and Defaults. In *Handbook of Logic and Language*. Amsterdam [etc.]; Cambridge, MA.
- [8] Yalcin, Seth (2007). Epistemic Modals. *Mind* 116 (464): 983–1026.

(九州大学)

ヒルベルト前期形式主義の方法論

Hilbert's Methodology in his first period of formalism

井上朋彦

Abstract

David Hilbert was the most important figure of formalism. He used the "axiomatic method" in his research on the foundations of mathematics: to axiomatize the theory and then to meta-study the conditions the axiomatic system must satisfy. Then, depending on the situation at the time, he studied the proofs of consistency in mathematics, seeking relative consistency proofs, or "direct" proofs using "early proof theory". In the standard view, Hilbert's research on mathematical foundations can be divided into two periods. This paper examines Hilbert's methodology before the establishment of "Beweistheorie" and the transition of his method for proving consistency, including the developments during 1917-20, which fall between the first and second periods of his formalism.

(1) 研究テーマ

ダフィット・ヒルベルト (David Hilbert, 1862-1943) は 19 世紀末から 20 世紀初頭の時代にかけて活躍したドイツの数学者であり、数論、代数学、幾何学、解析学、数理論理学等数学のさまざまな分野で傑出した成果を残した。数学の基礎についての研究にも精力的に取り組み、1920 年代頃には証明自体を数学的対象として扱う「証明論 *Beweistheorie*」を考案し、ゲッティンゲン大学の共同研究者等とともに数学の無矛盾性を数学内部で証明することを目指した。数学の基礎付けに対する彼らの思想的立場は一般に「形式主義」と言われる。

彼らの研究活動、いわゆる「ヒルベルト・プログラム」の最盛期は証明論提示以後の 1920 年代であるが、それ以前にもヒルベルトの数学基礎論研究は独自の研究成果を残してきた。ヒルベルトの数学の基礎に対する当初の関心は 19 世紀における数学の諸分野の発展や当時の数学の哲学の議論に喚起されたものであり、彼の研究はそうした 19 世紀数学の知的文脈の延長上にあることをまず理解しなければならない。ヒルベルトの数学基礎論研究は幾何学の研究で用いた「公理的方法 *axiomatisches Methode*」を応用する形で始まり、この方法に見られる数学の公理的提示と公理系に対するメタ研究は、

彼が一貫して持ち続けた研究態度であった。

本研究は証明論が確立される前のヒルベルトの数学基礎論研究の方法論について、先行研究を手掛かりとしてその変遷を詳細にたどったものである。

(2) 研究の背景・先行研究

以下、ヒルベルトの数学基礎論研究の時代区分と、証明論提示前の方法論について先行研究を確認する。

(2.1) ヒルベルトの数学基礎論研究の時代区分

ヒルベルトが数学基礎論研究に取り組んだ期間は 1899 年出版の『幾何学基礎論 (*Grundlagen der Geometrie*)』 [Hilbert 1899] から 1930 年代に出版されたベルナイスとの共著『数学の基礎 (*Grundlagen der Mathematik*)』 [Hilbert and Bernays 1934, 1939] までの約 40 年間である。 [Sieg 1999] や [野本 2019] によると、この期間を主要著作物の出版時期に従って「前期形式主義」と「後期形式主義」という二つの時期に区分する見方が標準的であるという。前期形式主義は 1900 年の論文「数概念について (*Über den Zahlbegriff*)」 [Hilbert 1900a] から 1905 年の論文「論理学と算術の基礎について (*Über die Grundlagen der Logik und der Arithmetik*)」 [Hilbert 1905] までの時期である。この時期はまだ記号論理学を用いた数学の形式化は実行されておらず、自然言語で記述された公理系や「モデル」に依存したアプローチをとっていたのが特徴である。一方、後期形式主義は 1922 年の論文「数学の新しい基礎付け (*Neuebegründung der Mathematik*)」 [Hilbert 1922] から 1931 年のゲーデルの「不完全性定理」までの時期である。この時期には [Hilbert 1922] で新たに示した方法「証明論 *Beweistheorie*」を用いて数学の無矛盾性証明研究が進められたが、クルト・ゲーデル (Kurt Gödel, 1906-1978) が証明した不完全性定理により、ヒルベルト等の試みは当初の形で達成することが不可能であることが明らかとなった (cf. [野本 2019, pp. 232-233])。

(2.2) 前期と後期の中間期

上の時代区分では前期と後期の間に長い空白期間を見ることが出来るが、この期間にヒルベルトが全く数学基礎論の研究を行っていなかったわけではない。

[Mancosu 1999, 2003] は、1910 年代にヒルベルト等のゲッティンゲン大学のグループがラッセルとホワイトヘッドの共著『プリンキピア・マテマティカ (*Principia Mathematica*)』 (以下, *PM*) の研究に取り組んでいたことを伝えている。曰く、1914 年のゲッティンゲンの数学コロキウムでハインリヒ・

ベーマン (Heinrich Behmann, 1891-1970) が *PM* についての講義を行ったことをきっかけに、同コロキウムでは数理論理学や集合論といった数学基礎論に関連する講義が急増している。ベーマンは当時、ヒルベルト指導下の学生であり、*PM* を題材にした博士論文を準備していた ([Mancosu 2003, pp. 179-181])。

ヒルベルトはその業績に比して生前の出版物はそれ程多くはないため、彼に関する歴史研究で重要となるのが講義録である。[Moore 1997] は 1917/18 年冬学期にゲッティンゲン大学で行われた講義「数学の原理 (Prinzipien der Mathematik)」に注目し、この講義の意義を次の二点に集約する：(1) 命題論理の「ポスト完全性 1」を初めて証明したこと、(2) 1 階述語論理の形式系を初めて提示したこと。1917 年頃のヒルベルトはラッセルの論理主義に接近しており、この講義の成果もそうした傾向を背景としたものである。[Sieg 1999] は、1917/18 年冬学期講義に加えて 1920 年冬学期と夏学期の講義も考察対象に含めることで、この時期のヒルベルトが論理主義的傾向から次第に離れて構成的数論の研究を始め、やがて有限主義に基づいた証明論に至るまでの過程を検討している。

(2.3) 証明論前夜のヒルベルトの方法論

ヒルベルトが数学の基礎付けの方法として最初期から説き続けた方法は公理的方法、ないし公理論であった。[林 2000] は、ヒルベルトの公理論と 20 世紀数学の公理論との相違点に注目する。すなわち、ヒルベルトは「最初の 20 世紀数学者」でありながらも、彼の説く公理論はブルバキの「構造主義」と異なり、数学の抽象構造ではなくて統語論的な論理を強調するものである(因みに 20 世紀数学で主流となったのはブルバキの公理論であった)。ヒルベルトの公理論は数学の基礎に対する関心から要請されたものであったが、そこには「ゴルダン問題」における論争に端を発し、「解析学の算術化」や「幾何学の統一」等、19 世紀数学において主に論じられたさまざまなトピックがその背景としてあった。ヒルベルトが自身の公理的方法を用いて数学基礎論研究に本格的に取り組むようになったのは 20 世紀に入ってからであるが、彼の方法論に大きな影響を及ぼしたのは、こうした 19 世紀数学の問題群だったのである。

[Ogawa 2004] は、前期形式主義時代のヒルベルトが数学基礎論研究の方法として公理的方法を採用した理由について考察している。曰く、ヒルベルトは数学の「厳密性 rigor」を保証する手段として公理的方法が最適だと考えており、その理由は次の二点に集約される：(1) 公理系により理論を組織化することで、理論に対する理解を深めることができること、(2) 数学的判

断と数学的推論を客観化することができること。[林 2000] の説くようにヒルベルトの公理論は論理を強調するものであったが、そこにはヒルベルトは「知識の完全な論理的基礎付け」に必要な厳密性を保証する手段として公理的方法が最適だと見なしていたことがあったのである。

ヒルベルトは数学の基礎について独自のアプローチをとったが、デデキントやクロネッカーといった彼以前のアプローチの影響を受けていなかったわけではない。[Sieg 2013b] は、ヒルベルトの実数論の方法に見られる彼らの影響と相違点を考察している。数学の基礎に対するヒルベルトの最初の成果は1900年の「数概念について」であるが、そこでは『幾何学基礎論』で用いた公理的方法を応用する形で実数論を含む算術の公理系を提示している。実数概念に対するアプローチで、ヒルベルトもデデキントも幾何学的表象に訴える点は共通している ([Sieg 2013b, pp. 77-80])。しかし、無矛盾性概念の理解については両者に懸隔を見ることができる。デデキントは自然数概念を特徴づける際に、可算無限集合に相当する「一重無限システム」を構成し、それが一意に定まることを示すことで、可算無限集合が無矛盾であることを示そうとした。一方、ヒルベルトは無矛盾性概念を公理からの有限回の論理推論で矛盾が導かれないことだと捉えていたが、具体的な推論規則についてはまだ提示していなかった。これらの無矛盾性概念に対して、Sieg はデデキントの捉え方を「意味論的」だとし、ヒルベルトの捉え方を「疑似統語論的」だとしている ([Sieg 2013b, p. 81])。

その後、ヒルベルトは1904年に代数的操作を用いた純粹統語論的な無矛盾性証明の方法を発表するも²、数学的帰納法の使用についてポアンカレの批判を受け、この方法はしばらく放置される (cf. [林・八杉 2006, pp. 174-186]: [野本 2019, pp. 250-253])。1920年になると、1904年に検討したような構成的数論に再び立ち返り、夏学期講義では非常に制限された形の数論に対する無矛盾性証明も示している (cf. [Sieg 1999, pp. 111-115]: [Ewald and Sieg 2013, pp. 280-283])。1920年の構成的数論の研究は後の「有限数学 finite Mathematik」に直接的につながるものであり、ここでの研究の成果が後期最初の成果として示されることになる。

(3) 筆者の主張

ヒルベルトの数学基礎論研究は前期形式主義、後期形式主義という二つの時期に区分する見方が標準的であるものの、その区分に入らない後期直前の時期にもヒルベルトは数学基礎論研究に取り組み、独自の成果を残していた。また、証明論確立前の前期時代から、ヒルベルトはそれまでとは異なる独自

の方法、公理的方法を説き、その下で算術の無矛盾性証明を与えようと模索していた。

本論文の研究は、こうした先行研究の成果を踏まえた上で、前期・後期形式主義の狭間に当たる 1917-20 年間の取り組みも含めて、証明論確立前のヒルベルトの公理的方法を彼が公理系に求めた条件から概括し、さらに当該時期の無矛盾性証明の方法を相対的無矛盾性証明と「前期証明論」との変遷から捉え直そうとする試みである。

以下、(1) ヒルベルト前期形式主義の公理的方法の要点、(2) 証明論確立前の無矛盾性証明の方法の変遷、について紙幅の許す限り論じる。

(3.1) ヒルベルト前期形式主義の公理的方法

ヒルベルトは「数概念について」 [Hilbert 1900a] で、数概念に対するアプローチとして「生成的方法」と「公理的方法」の二つを挙げた上で、数学基礎論に対する研究では後者が優位であることを述べる ([Hilbert 1900a], pp. 180-181). ヒルベルトが生成的方法に言及する際は、デデキントやクロネッカーらの方法が念頭に置かれており、彼らの方法に内在する限界を克服するのが公理的方法であった。

ヒルベルトの公理的方法とは、幾何学や算術といった特定の理論の公理系を提示した上で、その公理系に対するメタ研究を行うことである。公理系の構築についてヒルベルトは次のように述べる：

ある科学の基礎を研究しようというときには、その科学の基礎になる基本的概念の関係を正確にかつ完全に記述する公理系をうちたてる必要がある。設定された公理は、基本的概念の定義である。
([Hilbert 1900b, pp. 299-300]: 一松訳, p. 12)

すなわち、公理とは基本的概念が定義された命題であり、公理系とはある理論を展開するための必要最小限な公理の集まりのことである。[Hilbert 1900a] では演算規則や順序関係等、数の間の関係を自然言語を用いて書き下すことで、算術の公理系が立てられた³ ([Hilbert 1900a, pp. 181-183])。

ヒルベルトは様々な言葉をもって公理系に対する条件を述べるが、それらを次の四点に集約することが出来る： i. 存在を主張する形式、ii. 諸公理の従属性・独立性証明、iii. 公理系の完全性証明、iv. 公理系の無矛盾性証明。

i. 存在を主張する形式

「存在を主張する形式」とは公理系が扱う対象についての条件であり⁴、『幾何学基礎論』 [Hilbert 1899] におけるヒルベルトの公理論の最も革新

的な点とも言える。[Hilbert 1899] では公理系の対象に関して、諸公理が規定する対象どうしの関係を満たすような事物の領域をあらかじめ仮定し、これを一つの確定した体系と見なしている。ユークリッド『原論』等、ヒルベルトより前の公理論では、点、直線、平面といった幾何学的対象は具体的な経験を基礎として個別に定義されていた。しかし、ヒルベルトはそのようなやり方とは異なり、諸公理の規定する関係を満たす集合であれば、どのような領域をとっても問題はないとした。こうして、公理系の対象について「内容」を除外して「形式」のみに注目することで、より抽象性の高い議論を行うことが可能となった。

ii. 諸公理の従属性・独立性証明

公理間の従属性・独立性とは、公理系の公理が他の公理から証明されるか、あるいはそのようなことがないのかといった概念であり、『幾何学基礎論』では幾何学の公理系をいくつかの小公理系に分割し、公理間の相互依存関係を見ることで諸公理が相互独立であることを証明した。1917年のチューリヒ講演「公理的思考」では、公理系に求められる条件として、無矛盾性証明とともに、従属性・独立性の証明が挙げられている ([Hilbert 1918, p. 148])。

iii. 公理系の完全性証明

「完全性 Vollständigkeit」はヒルベルトの数学基礎論研究において様々な使われ方をしているため、その時々で注意深く検討する必要のある概念である⁵。公理系の完全性については以下のように述べている：

この公理系の「無矛盾性」と「完全性」、つまりこれらの公理から決して矛盾が導かれないこと、またこれらの公理からすべての幾何学の定理が導かれること、この二つのことが証明されなければならない。([Hilbert 1900a, p. 181]: 寺阪・大西訳, p. 203, 強調は筆者によるもの)。

ここでヒルベルトは公理系の公理が、ある理論の任意の命題を導く上で、量的に十分であることを求めている。

iv. 公理系の無矛盾性証明

公理系の無矛盾性はこれら四つの条件の中で最重要概念である。上の引用では完全性ととともに無矛盾性についても言及されており、「これらの公理から決して矛盾が導かれないこと」とある。当時のヒルベルトの無矛盾性概念理解が、デデキントのような 19 世紀的な意味論的理解に対して、疑似統語論的であったのは上で確認した通りである。無矛盾性証明の方法については

次節で確認する。

(3.2) 証明論確立前の方法の変遷

証明論確立前のヒルベルトの数学基礎論の方法、とりわけ無矛盾性証明に対する方法は、相対的無矛盾性証明と「前期証明論」とを行き来するものだったと考えることが出来る。以下、その変遷を簡単に確認する。

『幾何学基礎論』の成果の一つとして幾何学の無矛盾性を実数論の無矛盾性に還元することで相対的に証明したことが挙げられる。しかし、前期時点のヒルベルトは、実数論の無矛盾性は他に還元する先がなく、その無矛盾性を「直接的」に証明する必要があると考えていた (cf. [Hilbert 1900b])。

1904年頃の算術断片に対する「無矛盾性証明」は、証明と同一視された記号列に代数的操作を施すことで、純粹統語論的にその無矛盾性証明を与えようとしたものだった。ここでとられた統語論的方法は、メタ数学と内容的数学との区別も曖昧である等、数学基礎論の手法として未成熟ではあるが、(1) 意味を排除した記号のみを用いて可算無限集合の定義を行い、(2) 数学の証明を記号の有限列と見なし、(3) 公理と推論規則のみを用いて「証明」を導出し、そこから相矛盾する結果が導かれないことを検証する等、後の証明論との共通点も見ることが出来る。このような証明論確立前にとられた純粹統語論的な証明方法を、簡単のため「前期証明論」とする。しかし、1904年の方法はポアンカレにより致命的な欠点を指摘され、挫折することになった。

1910年代になり、*PM*の研究に本格的に取り組むようになると、ヒルベルトはラッセルの論理主義に急速に接近するようになった。1917年頃には、実数論の無矛盾性はヴァイアーシュトラスやデデキントにより整数論の無矛盾性に還元されたと述べた上で、さらに整数論の無矛盾性は「論理学 *Logik*」を除いて還元する学問分野はないとしている⁶ ([Hilbert 1918, p. 153])。すなわち、ここでは『幾何学基礎論』でとった相対的無矛盾性証明が実数論に対して検討され、整数論を経て、最終的に論理学へと至っている。

しかし、後期形式主義直前の1920年頃になると、主に還元可能性公理に対する不満から論理主義的アプローチから離れ、再び前期証明論的方法を模索するようになった。この時期にヒルベルト等が中心的に研究したのが構成的数論である。さらに、1920年冬学期講義では非常に制限された自然数の体系に対して帰納法の図式を用いた無矛盾性証明が与えられている。ここでの議論は深められて翌年には講演で発表され、さらにその翌年には論文として出版された。これが「数学の新しい基礎付け」であり、ここでラッセル等の論理主義やブラウワー等の直観主義とも異なる思想的立場の方法として「証明論」を提示した。

(4) 今後の展望

以上、証明論確立前の数学基礎論におけるヒルベルトの方法を手短に論じてきたが、ここでの議論はまだまだ予備的考察の域を出ない。筆者の最大の狙いとしては、20世紀初頭の数学基礎論史を哲学史の枠内で解釈することにある。この目的の実現のためにも、未だ数理論理学を用いた方法の整備が行き届いていなかったヒルベルト前期形式主義の方法論及び数学の哲学に対する考察をより深めていくことが急務であると考えられる。

ヒルベルトの数学の哲学について今回取り上げることはできなかったが、数学的对象の存在と無矛盾性概念とを結びつけた「存在＝無矛盾性のテーゼ」や、彼が青年期から抱き続けていた「全数学問題の可解性公理」等が主要なトピックとして考えられる。また、彼特有の「完全性」概念の多義的な使い方についても、彼の数学思想との関連から議論できるかもしれない。

注

¹ ポスト完全性とは、任意の命題に対してその命題か、その命題の否定が証明できることをいう。アメリカの論理学者エミール・ポスト (Emil Leon Post, 1897-1954) に因む (cf. [野本 2019, pp. 310-312])。

² 1904年にハイデルベルクで講演を行い、翌年にその講演を基にして論文「論理学と算術の基礎について」 ([Hilbert 1905]) を発表した。

³ この場合の算術は実数論も含まれており、実数の連続性を「アルキメデスの公理」と「完全性の公理」で定義している。なお、この算術の公理系は抽象代数学で言うところのアルキメデス実数体に相当するものである。

⁴ ヒルベルト以前の公理論と差別化する形でこのような言葉が使われ始めたのは1922年以降である。なお、「存在を主張する形式」は *existentiale Form* に対する [吉田・瀧野 1993] の訳に則っている。

⁵ ヒルベルトが数学基礎論研究において「完全性 *Vollständigkeit*」に言及する際は、(1) 公理系の完全性、(2) 実数の完備性、(3) ポスト完全性、(4) 意味論的完全性、のいずれかを指しているものと考えられる。

⁶ この時期、とりわけ前期時代のヒルベルトは、*Logik* を非常に広い意味で用いているため、もちろん現代的な「論理学」の理解とは齟齬が生じる。当時の使い方に即して理解する必要があるだろう。

(5) 参考文献

林晋 [2000] 「ヒルベルトと二十世紀数学 公理主義とは何だったか?」, 『現代思想』, 2000年10月臨時増刊, 青土社, 30-41.

林晋・八杉満利子訳・解説 [2006] 『ゲーデル 不完全性定理』, 岩波文庫, 73-310.

- Hilbert, D. [1899] *Grundlagen der Geometrie*, in Festschrift zur Feier der Enthüllung des Gauss-Weber-Denkmalles in Göttingen, Teubner, pp. 1-92. 寺阪英孝・大西正男訳『幾何学の基礎』, 共立出版, 1970).
- [1900a] “Über den Zahlbegriff”, *JDMV*. Bd. 8, 1900 (『幾何学の基礎』付録VI, 共立出版).
- [1900b] “Mathematische Probleme”, Paris. Rep. in [Hilbert 1935] (一松信訳『ヒルベルト 数学の問題』, 共立出版, 1969).
- [1905] “Über die Grundlagen der Logik und Arithmetik” in *Verhandlungen des Dritten Internationalen Mathematiker-Kongress*, Teubner, pp. 174-185. (『幾何学の基礎』付録VII, 共立出版).
- [1918] “Axiomatisches Denken”, *Math. Ann*, 78: 405-15, 1918. (中村幸四郎訳『幾何学基礎論』, ちくま学芸文庫).
- [1935] *Gesammelte Abhandlungen III 2* Aufl. Springer.
- Hilbert, D. and Bernays, P. [1934] *Grundlagen der Mathematik*, Vol. I, Springer (吉田夏彦・淵野昌訳『数学の基礎』, シュプリンガー・フェアラーク東京, 1993).
- [1939] *Grundlagen der Mathematik*, Vol. II, Springer.
- Mancosu, P. [1999] “Between Russell and Hilbert: Behmann on the Foundations of mathematics”, *BSL*, 5(3): 303-330, rep. in [Mancosu 2010, pp. 159-175].
- [2003] “The Russellian Influence on Hilbert and his school”, *Synthese*, 137: 59-101, rep. in [Mancosu 2010, pp. 176-198].
- [2010] *The Adventure of Reason: Interplay Between Philosophy of Mathematics and Mathematical Logic, 1900–1940*, OUP.
- Moore, G. H. [1997] “Hilbert and the emergence of modern mathematical logic”, *Theoria*, 12(1): 65-90.
- 野本和幸 [2019] 『数論・論理・意味論 その原型と展開 知の巨人たちの軌跡をたどる』, 東京大学出版会.
- Ogawa, Y. [2004] “The Pursuit of Rigor: Hilbert’s axiomatic method and the objectivity of mathematics”, *Annals of the Japan Association for Philosophy of Science*, 12(2): 89-108.
- Sieg, W. [1999] “Hilbert’s Programs 1917-1922”, *BSL*, 5: 1-44, rep. in [Sieg 2013a, pp. 91-127].

— [2013a] *Hilbert's Programs and Beyond*, OUP.

— [2013b] 'Method for real arithmetic' in [Sieg 2013a, pp. 73-90].

Ewald, W. and Sieg, W. [2013] "Introduction to the 1920 Lectures" in *David Hilbert's Lectures on the Foundations of Arithmetic and Logic 1917-1933*, OUP.

(名古屋大学)

合理的なアクラシアはいかにして可能か
How Is Rational Akrasia Possible?

高山馨

Abstract

This paper presents arguments in defense of the possibility of rational akrasia. It is commonly thought that acting against one's best judgment is always irrational. However, some philosophers, following Robert Audi, have argued that it can be rational. I suggest that the difference between proponents and opponents of the possibility of rational akrasia is not only a difference in what to consider most important in the discussion (e.g. the substantive merits of an action or the procedural consistency of practical reasoning), but also a difference in how they think about the rationality of an action.

(1) 研究テーマ

本稿が論じるのは、「アクラシアは常に非合理である」という一般的見解を拒否し、「合理的なアクラシアが可能である」と主張する議論である。

(2) 研究の背景・先行研究

「自身の最善の判断（ないし、より良い判断）に反して行為すること」とされるアクラシア^{1,2}（ないしは「意志の弱さ」）は、実践的非合理性の典型例の一つだと考えられてきた（Stroud and Tappolet 2003; Tappolet 2013）。デイヴィドソンは、アクラシアをめぐる現代の議論の古典である Davidson（1970）において、アクラシア的行為は可能であるという主張を擁護したものの、そのような行為は、デイヴィドソンの言う「自制の原理（principle of continence）」に反しており、非合理的だとみなしていた。「自制の原理」とは、端的に述べれば、「自身の最善の判断と一致するよう行為せよ」という要求である。自制の原理を合理性の原理だとみなすならば、自身の最善の判断に反して行為することとして特徴づけられるアクラシアが常に非合理だとされるのは、自明であるように思われる。

しかしながら、ロバート・アウディを始めとする論者の中には、自身の最善の判断に反することが合理的になる場合がある、すなわち、合理的なアクラシアが可能であると主張する者がいる（Audi 1990; McIntyre 1990; Arpaly

2000, 2003; Brunero 2013; Hrishikesh 2017)。

以下、最初期に議論を提示した Audi (1990) を中心に、合理的なアクラシアを擁護する議論をごく簡単に概観する。あらかじめ先行研究における大きな応答方針の一つを確認しておくならば、それは次のようなものになる。すなわち、「行為者のより良い判断それ自体も間違いうるため、誤った最善の判断に反して行為する方が、結果的に合理的な行為を行うことになる場合がある」という方針である。

また、語彙の確認として、「最善の／より良い判断 (best/better judgement)」とは、「道徳的、非道徳的考察を含む、あらゆる関連する考察を総合すれば、行為者 A は行為 ϕ すべき」という判断を指す。ここでの「考察 (consideration)」を「理由」と同一視することで、この判断は次と同値になると考えられる。すなわち、「A は ϕ する決定的な／最大の理由を持っている」。なお、本稿では、「決定的な／最大の理由」は、「A が持っているさまざまな ϕ する理由、 ϕ しない理由を総合すると、 ϕ する理由の方が相対的に強い」という理由の相対的な重みづけを意味する³。

2.1 アウディの議論

Audi (1990) は、以下の事例を用いる (Audi 1990, 276-277)。

相応の罰を与えようとするジョン：

ジョンは、彼の娘が土曜の夜に外出することを許可すべきでないと信じている。なぜなら、外出を禁止することは、娘が勉強すべきときに電話をしていたことに対する相応の罰だと考えているからである。相応の罰を与えるべきだという考えにコミットしているジョンは、いざ娘に外出禁止を言い渡そうとした際、娘がかなり動揺するであろうことを悟り、厳しく叱ることで許すことにして、外出を許した。

ジョンは、自分は娘に外出を許可すべきでないと信じているが、それに反して外出を許可する。一見すると、ジョン心的状態は整合しておらず、その振る舞いは非合理的である。しかし、アウディによれば、上記の事例において、ジョンを単純に非合理とみなすことはできない。なぜなら、その判断は、あくまでも、ジョンの事例をどのように詳細にするかに依存するからである。アウディの議論を簡潔に述べると以下のようなになる (Audi 1990, 277-279)。

1. 娘に罰を与えないことにうまく整合するような心的状態が存在すると

する（例えば、罰を与えることは人を傷つけるだろうという信念や人を傷つけないという欲求、罰を与えることは娘から反発を引き起こすという信念と娘に反発されたくないという欲求など）。

2. そうした欲求や信念は「ジョンの性格に深く根ざした正当な理由 (good reasons rooted deep in his character)」を構成する。
3. ジョンの持つ「娘に罰を与えるべきだ」という信念それ自体は、ジョンの性格や利害関心、理想とは整合しておらず、非合理的な信念である
4. それゆえ、ジョンが罰を与えるべきだという自身の最善の判断に反して行為することは、それに一致して行為するよりも、ジョンの心的状態全体にうまく整合しうる。
5. したがって、最善の判断に反することは合理的でありうる。

アウディは、ジョンの心的状態の全体的な整合性のうちに、ジョンの合理性が保たれていると考えている。つまり、一見すると局所的に不整合が生じているように思える行為者の心的状態も、全体的／総合的に見ると問題なく整合している。それゆえ、最善の判断に反することも合理的になりうるのだ、というわけである。ここでのジョンの事例は、アウディのメインテーゼである「合理性は全体論的な概念でなければならず、そうであるならば、ある種のアクラシア的行為は合理的でありうる」(Audi 1990, 280) という主張を支持するために提示されたものである。

他方、この議論においてアウディは、合理的な行為とはどのようなものかという点に関して、いくつかのモデルを提示し、そのうちの一つである「理由のバランスモデル (the balance of reasons model)」を支持する。このモデルによれば、合理的な行為とは、行為者の持つ十分に正当な理由に基づいて、適切な方法でなされる行為である。したがって、上述の議論のステップ 2-4 を踏まえると、ジョンが、ある限定された範囲では、最善である判断に反して行為することは、その結果として自身の持つ正当な理由に基づいて行為することとなり、当該のアクラシアによる行為が合理的な行為とみなされるようになる。

とはいえ、アウディは、たとえ合理的なアクラシアが可能であるとしても、アクラシアが、ある程度、その合理性に反しているということ（つまり、ある程度、非合理であることを認めた上で、「最高度に合理的でなくとも、ある行為はバランスの上で合理的でありうる」と主張するのである (Audi 1990, 280)。ここで認められる、アクラシアの合理性はあくまでも相対的なものだ

という点は、3節でその問題点を確認する。

2.2 アウディ以後の議論

Audi (1990) 以降のアクラシアの合理性に関する議論は、基本的に、アウディの議論の枠組みを引き継いでいる。そして、行為者の最善の判断の特権性を弱めることで、そのアクラシア的行為の実質的な合理性を示すという形式を採ることが多い。

McIntyre (1990) において、マッキンタイアは、行為者の「より良い判断」は実際には適切に包括的な判断ではなく、関連するすべての考察事項を考慮に入れていない可能性がありうると主張する。具体的には、『ハックルベリー・フィンの冒険』の主人公ハックのような、次の事例を念頭に置いている (McIntyre 1990, 380; 同様の事例を用いる議論として、Bennett 1974; Arpaly 2003; Brunero 2013; Hrishikesh 2017)。

ハック・フィンの事例：

ハックは、逃亡奴隷のジムを、その「所有者」であるワトソン夫人の元へ返すかどうか迷っている。ハックは、当時の支配的な価値観に則り、ジムはワトソン夫人の「所有物」であり、正当な持ち主に返還すべきだと信じている。そのため、逃亡を助けたことに良心の呵責を覚え、ジムをワトソン夫人の元へ返すことを決心する。しかし、いざその機会を得た際、決心を実行に移せない自分に気づき、ハックは自分の弱さを責め、友人を守るためにジムを探しにきた賞金稼ぎたちに嘘をつく。

マッキンタイアによれば、ジムをワトソン夫人の元へ返すべきだというハックの判断は、主にジムをワトソン夫人の所有物だとみなす考えに基づいており、ハックの持つジムへの友情や愛着といった強い感情、その他の関連する要因を無視している。そのため、ハックの「より良い判断」は、実際には十分に包括的な判断ではなく、関連するすべての考察事項を考慮に入れていない。そのため、この時ハックが最大の理由を持つ行為とは、実はジムの逃亡を助けることなのだ、というわけである。それゆえ、マッキンタイアは、「アクラティックな行為者は、最大の理由を持っている行為を自分が行なっていると信じていなくても、それにもかかわらず、行為者が追求している行為は、行為者が追求すべき最大の理由を持ったものであるかもしれない」と主張するのである (McIntyre 1990, 385)。マッキンタイアによれば、行為者が実際には自身の持つべき最大の理由を適切に認識していなかったとして

も、実質的にその理由に基づき行為することが可能である。そして、 ϕ する最大の理由に基づき行為することは合理的であるため、合理的なアクラシアが可能となるのである。

同様に、Arpaly (2000, 2003) でも、行為者の最善の判断の特権性は低く見積もられる。とりわけアーパリーは、「合理性の理論は、行為者の最善の判断に何か特別なものがあると仮定すべきではない。行為者の最善の判断は、単なる信念の一つにすぎない」とまで述べている (Arpaly 2000, 512)。アーパリーによれば、われわれは自分が何をすべきかについて誤った信念を持つ可能性があり、その信念に反して行為することで、正当な理由のために行為することができる。そして、正当な理由のために行為することは、合理的に行為することなのだと主張するのである。

(3) 筆者の主張

合理的なアクラシアの可能性を擁護する議論を概観したが、先に述べた通り、その基本的な方針は次のようなものであった。すなわち、「行為者のより良い判断それ自体も間違いうるため、誤った最善の判断に反して行為する方が、結果的に合理的な行為を行うことになる場合がある」というものである。

しかしながら、上述のオーディに加え、マッキンタイアやアーパリーたちも、アクラシアを単に合理的だと捉えているわけではない。彼／彼女らの主張は、自身の「最善の判断」に反して行為する方がそれに従うよりも、より合理的でありうるという、あくまでも相対的な評価における合理性の主張である。すなわち、自身の判断に従う行為と比較して、それに反して行為する方がより合理的な場合がありうる、と主張しているのである。

この点に関して、一つの単純な問題は、Brunero (2013) が指摘するように、自身の最善の判断に従う行為とそれに反する行為の両方もが非合理であるかもしれない、というものである。つまり、ある選択肢は他の選択肢よりもより良いものでありながら、結局どちらも悪い選択肢である可能性がある (Brunero 2013, 551)。例えば、のっぴきならない状況において、指を1本犠牲にして失うか、それとも2本失うか、といった場面があてはまるかもしれない (指を2本失うよりも、1本失うほうがより良い選択肢であろう)。しかし、この点に関しては、それぞれの個別事例の詳細を特定することで、たとえ「合理性」が相対評価における語として用いられていたとしても、非合理な中でのどんぐりの背比べという事態に該当しないケースを見つけることは可能であろう。そうであるならば、「自身の最善の判断に反して行為するほうが、相対的にも実質的に合理的でありうる」という主張は、さしあたり

維持することができる。

より根本的な問題点、ないし相違点は、Stroud and Svirsky (2019) が指摘するとおり、合理的なアクラシアの可能性を擁護する論者たちがアクラシアによる行為の実質的なメリットを強調するのに対して、アクラシアは常に非合理だという見解を維持する論者は、そもそもその点を問題として扱わないかもしれない、ということである。なぜなら、アクラティックな行為者の行為が結果的にどれほど成功を納めたとしても、その行為者の実践的推論には、依然として手続き上の欠陥が存在するからである。たとえその行為が実際には正当な理由によって支持されているとしても、別の行為を支持する判断を下した行為者がそのことに気がついていない以上、その行為者の推論ステップには何らかの誤りがあることが示されている。推論に失敗しているため、推論の基礎ないしは帰結となりうる信念や意図などの心的状態が整合しているとは言えず、そのアクラシア的行為は常に非合理的なのである。

伝統的見解の擁護者がこのように主張するのに対し、合理的なアクラシアの擁護者は、あくまで、たとえ行為者の心的状態が局所的に不整合であったとしても、実質的に行為が正統な理由に支持されているならば合理的なのだ、と主張する。両者のこのすれ違いは、議論における強調点の違い (e.g. 行為の実質的なメリットか、実践的推論の手続き的な整合性)、ひいては行為の合理性をどのようなものとして捉えるかの違いに帰着すると思われる。次節では、この点を確認する。

(4) 今後の展望

3 節の最後で触れたように、アクラシアを非合理とみなす伝統主義者と、アクラシアが合理的でありうるとする論者の相違について、その問題点を端的に述べるならば、両者の「行為の合理性」理解が異なっているのではないかと、ということである。「アクラシアは常に非合理である」という主張と、「合理的なアクラシアが可能である」という主張における合理性理解が異なっているならば、その点を整理しない限り、議論の成立は難しい。したがって、合理的なアクラシアの議論においては、「アクラシア」がどのような現象を指示するのかといった問題設定の確認に加えて、そもそも行為の合理性をどのようなものとして捉えるかという点が重要な論点の一つとなりうるのである。

最後に、行為の合理性についてよく区別される二つを確認し、その区別によってアクラシアをめぐる論争を整理し直す可能性について述べることで、本稿を閉じる。

行為の合理性については、主に次の二つの見解がある (Hinchman 2013)。

第一に、「合理的であるとは、行為者の心的状態が適切に関係し合っていることだ」とする見解、第二に、「合理的であるとは、行為者が自身の持つ理由へ適切に応答することだ」とする見解である。この二つは、いわゆる「構造的合理性 (structural rationality)」と「実質的合理性 (substantive rationality)」に対応すると考えられる (Broome 2007, 2010; Kiesewetter 2017; Worsnip 2021)。そして、私の見るところ、アクラシアの合理性をめぐる二つの立場のうち、合理的なアクラシアの可能性を否定する論者たちは構造的合理性を、そして合理的なアクラシアを擁護する論者たちは実質的合理性を採用する傾向にある。ただし、中には Brunero (2013) のように構造的合理性を採用しつつ、合理的なアクラシアの存在を擁護する論者もいるため、これはいまだ見通しに過ぎない。この二つの合理性を用いてこの論争を整理し直し、そのうえで、どちらの合理性が行為を考えるうえで適切なのかを考えることが、今後の課題である。

注

1. より正確には、「行為者 A は自分は ϕ すべきであると信じており、かつ ϕ することを意図しない / ϕ しない」と、差しあたり、定義することができる。ここで「 ϕ することを意図しないこと」と「 ϕ しないこと」を選言的に並列する理由の一つは、実践的推論の帰結が意図であるか、あるいは行為であるか (あるいは規範的信念であるか) を、本稿では、オープンにしているからである。
2. 本稿では、「アクラシア (akrasia)」と「意志の弱さ (weakness of will)」は相互に交換可能な語として扱う。両者を概念的に区別する議論については、Holton (1999) を参照のこと。
3. ここでの整理は、笠木雅史氏の 2020 年度東北大学での集中講義「倫理思想各論・倫理学特論Ⅲ」授業資料に依っている。

(5) 参考文献

- Arpaly, N. (2000). On Acting Rationally Against One's Better Judgement. *Etics*, 110: 488-513.
- , (2003). *Unprincipled Virtue: An Inquiry into Moral Agency*, New York: Oxford University Press.
- Audi, R. (1990). Weakness of Will and Rational Action. *Australasian Journal of Philosophy*, 68: 270-281.
- Bennett, J. (1974). The Conscience of Huckleberry Finn. *Philosophy*, 49:

123-134.

- Broome, J. (2007). Does Rationality Consist in Responding Correctly to Reasons? *Journal of Moral Philosophy*, 4(3), 349-374.
- . (2010). Rationality. in O'Connor, T. and Sandis, C. (eds.) *A Companion to the Philosophy of Action*, Blackwell, 285-292.
- Brunero, J. (2013). Rational Akrasia. *Organon*, F 20 (4): 546-566.
- Davidson, D. (1970). How Is Weakness of the Will Possible? in Davidson 1980, 21-42. [柴田正良訳「意志の弱さはいかにして可能か」『行為と出来事』(服部裕幸・柴田正良訳)、勁草書房、所収]
- . (1980). *Essays on Actions and Events*, Oxford: Clarendon Press. [服部裕幸・柴田正良訳『行為と出来事』、勁草書房]
- Hinchman, E. (2013). Rational Requirements and 'Rational' Akrasia. *Philosophical Studies*, 166: 529-552.
- Holton, R. (1999). Intention and Weakness of Will. *Journal of Philosophy*, 96: 241-262.
- Hrishikesh, J. (2017). What's the matter of the Huck Finn? *Philosophical Explorations: An International Journal for the Philosophy of Mind and Action*, 20: 70-87.
- Kiesewetter, B. (2017). *The Normativity of Rationality*, Oxford University Press.
- McIntyre, A. (1990). Is Akratic Action Always Irrational? in O. Flanagan and A. Rorty (eds.). *Identity, Character, and Morality*, Cambridge, Ma: MIT Press, 379-400.
- . (2006). What Is Wrong with Weakness of Will? *The Journal of Philosophy*, 103(6): 284-311.
- Stroud, S. and Svirsky, L. "Weakness of Will", *The Stanford Encyclopedia of Philosophy (Fall 2019 Edition)*, Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/fall2019/entries/weakness-will/>>.
- Stroud, S., and Tappolet, C. (eds.). (2003). *Weakness of Will and Practical Irrationality*, Oxford: Clarendon Press.
- Tappolet, C. (2013). Weakness of Will. In Hugh LaFollette (ed.), 2013, *International Encyclopedia of Ethics*. Wiley-Blackwell. 4412-21.
- Worsnip, A. (2021). *Fitting Things Together: Coherence and the Demands of Structural Rationality*, Oxford University Press.

(東北大学)