新進研究者 Research Notes

日本科学哲学会
科学基礎論学会

第 3 号　（2020 年）

松本将平　知覚的知識に関するセラーズの見解をいかに修正すべきか ・・・・・・・・・・・・・・・・・・・・・ 1
藤原諒祐　コネクショニスト認識論の批判的検討 ・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・ 9
山銅康弘　科学哲学からみた心理学における再現性のタイプについて ・・・・・・・・・・・・・・・・・・・・・ 18
水上拓哉　ソーシャルロボットの倫理における仮想的行為者性概念の可能性と限界 ・・・・・・・・ 27
姜雪菲　われわれは自律的ロボットの行為に責任を負いうるか ・・・・・・・・・・・・・・・・・・・・・・・・ 36
榎本啄杜　情報概念の真理性は正当化できるか ・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・ 46
溝越大秦　哲学的治療において必要なもの ・・ 55
伊沢亘洋　科学者はどのように非認識的価値を考慮するべきか ・・・・・・・・・・・・・・・・・・・・・・・・・ 63
角田健太郎　なぜ古典論理に計算論的意味を与えるのか
—SLCにもとづくDummett的観点の拡張の試み— ・・・・・・・・・・・・・・・・・・・・・・・ 69
稲田進一　存在論的多元主義の可能性 ・・・ 77

JAPANESE STUDENT RESEARCH NOTES OF PHILOSOPHY OF SCIENCE
Philosophy of Science Society, Japan
Japan Association for Philosophy of Science
No. 3

Contents

Shohei Matsumoto: How Should We Modify the Sellars' View on Perceptual Knowledge? ····· 1

Ryosuke Fujiwara: A Critical Study on Connectionist Epistemology ····· 9

Yasuhiro Sando: On The Types of Replicability in Psychology from The Point of View of Philosophy of Science ····· 18

Mizukami Takuya: The Possibilities and Limitations of Virtual Moral Agency in Ethics of Social Robots ····· 27

Jiang Xuefei Are We Responsible for Autonomous Robots' acts? ····· 36

Enomoto Takuto Can we justify that the concept of information requires truth? ····· 46

Taishin Mizokoshi What is needed in philosophical therapy ····· 55

Koyo Izawa How should a scientist consider non-epistemic values? ····· 63

Kentaro Tsunoda Why giving computational interpretation to classical logic? -an attempt at extending Dummett's conception of meaning based on SLC- ····· 69

Shinichi Inada The possibility of ontological pluralism ····· 77

知覚的知識に関するセラーズの見解をいかに修正すべきか
How should we modify the Sellars' view on perceptual knowledge?

松本将平

Abstract

Among the conditions of attribution of perceptual knowledge presented by Wilfrid Sellars, a condition called "the second hurdle" is criticized by Robert Brandom as being too strong. Brandom presents a counter-example to it and therefore argues that it is not always necessary. Michael Williams, however, points out a problem of the example and claims that the second hurdle weakened properly is necessary for perceptual knowledge. My points in this paper are as follows. Although Brandom's example is problematic, Williams' criticism of Brandom is inconclusive, and Brandom's view still seems to be correct. Nevertheless, I claim, Williams' explanation is basically correct, because only few cases are appropriate examples of knowledge that do not satisfy the second hurdle.

(1) 研究テーマ

知覚的知識（知覚において獲得される非推論的知識）はいかにして可能なのか。この問いに対しウィルフリド・セラーズが提示した回答は、一方で知識という概念に対する重要な洞察を含みながらも、もう一方で「悪名高い強い内在主義」(Rosenberg 2007, p. 248)であり、問題を抱えているように見える。彼の抱える問題をうまく回避できるように、彼の立場を修正することはできないか。これが本研究のテーマである。

(2) 研究の背景・先行研究

まずは、「経験論と心の哲学」（EPM）§35 において示されるセラーズの立場を確認することから始めたい。彼は、まず「出発点」の立場を提示した上で、二つのハードルを課すことで、彼自身がコミットする知覚的知識の帰属条件を与える。

本題に入る前に一つ注意をしておく。セラーズは EPM においては基本的に「報告」（知覚によって主体に引き起こされる、文トークンの（外的ないし内的）発話）をベースに認識論的立場を提示する。それは、我々の心的な出来事や状態を問題にする際、それらの特徴を明らかにするために、それらを我々の言語的な振る舞い及びその傾向性に近似して議論するという戦略(「言

語行動主義」）をセラーズがとるためである。言語行動主義においては、「p と考えること」は、一次的な意味としては「『p』と述べること」と、二次的な意味としては「『p』と述べる短期間の直前傾向性」とみなされる（“The structure of knowledge “（SK） II, §10）。EPM における「報告」は、知覚に際して生じる「述べる」ことだと解釈できる。そして、セラーズにおいては、上の意味における「p と述べる」ことの短期間の直前的傾向性が、p という信念を持っているということである（Sellars 1969, 71；笠木 2014, スライド 8）。以上を踏まえた上で、彼の見解を見ていくことにしよう。

「出発点」は次のような立場である。例えば、ある主体 S が、知覚の標準的状況のもとで目の前に実際に φ である（例：赤い）ような対象があるときそのときに限って「これは φ である」と報告する傾向性（以下これを傾向性 D とよぶ）を持っているとしよう。そして彼は実際に標準的状況のもとで対象を知覚し、当該の傾向性の表出として当該の報告をなした（あるいはその報告をなす短期的傾向性としての信念を獲得した）としよう。彼がまさにその傾向性の表出としてその報告をなした（信念を獲得した）がゆえに、その報告（ないし信念）は真である見込みが高く、それゆえに S は目の前の対象が φ だという知識を持っている。これが出発点のアイデアである（これは、いわゆる信頼性主義が課す知識の条件に他ならない）。

しかし、セラーズは「出発点」の立場に満足しない。というのも、知識とは規範的概念であり、その規範性は「出発点」の条件では確保されていないとセラーズは考えるからである。それゆえ彼は、S に当該の知識が帰属されるためには、出発点の条件に加えて当該の報告が「権威」を持っていなければならない、と述べる。これが「第一のハードル」である。では、報告はいかにして権威を持つのか。それは、その報告の権威の存在が認められることによってである(Rosenberg 2007, 248)。どういうことか。例えば、傾向性 D を実際に身につけている人 S がいるとしよう。この時、我々は S がそのような報告をなしたこと（あるいは S にそのような報告をなす短期的直前傾向性としての信念が帰属されていること）から実際に S の目の前に φ であるようなものがあるのだと推論することができる。このような推論を、ロバート・ブランダム（2000）に倣って「信頼性推論」と呼ぶことにしよう。第一のハードルで要求されていることは、S の当該の報告ないし信念に対して信頼性推論をなして良いとみなしている人が存在することなのである。「出発点」の条件は報告者が事実として標準的状況において世界と適切に対応した報告をなす傾向性を持っているかどうかが眼目であったのに対し、「第一のハードル」はそこに〈知識帰属者による評価〉という観点を加えるのである。

しかし、第一のハードルまででは、例えば文トークンの形で測定した温度を報告する温度計などにも知識が帰属されることになる。ゆえにセラーズは、以上に加えて「第二のハードル」を課す。これが要求することは、S 自身がその報告ないし信念の権威を認識しているということである。どういうことか。EPM §36 および “The structure of knowledge” （以下 SK と略記）第 II 部§40 を見る限り、〈目の前の対象について「これは φ である」という自分の報告が実際に目の前の対象が φ であることの信頼可能な徴候である〉という一般的知識を主体自身が持っており、それゆえ信頼性推論をなすことができるということをセラーズ自身は要求していると解釈するのが自然である。ゆえに、セラーズにおける知覚的知識の帰属条件は次のように纏められる。

（セラーズにおける知覚的知識の帰属条件）
S が、目の前の対象が φ であるという知覚的知識を持っているためには、傾向性 D の表出として「これは φ だ」という報告を実際になしているかあるいはそれをなす短期間の直前的傾向性を持っており（出発点）、その報告ないし信念についての信頼性推論を他ならぬ S 自身がなすことができることが必要である（第二のハードル）。[i]

しかし、主体自身が当該の報告ないし信念を信頼性推論によって正当化できることが、本当にその人への知覚的知識の帰属のために必要なのかについては疑問の余地がある。また、本稿では詳述できないが第二のハードルは額面通り受け取ると正当化の循環を引き起こす懸念がある（Sosa 2009）。第二のハードルは問題含みなのである。

そこでブランダムは、セラーズの基本的な枠組みは受け入れつつも、第二のハードルについては「多分この点でセラーズは言い過ぎている」（Brandom, 1997, p. 157）と述べ、第二のハードルを捨てることを提案する。そして彼は、やはり本稿では詳述できないが、代わりに彼特有の信念帰属条件を課すことで温度計などを知識帰属者から除外するのである。

そこでブランダムは、第二のハードルの反例として次のケースを提示する。

陶片を見るだけでそれがトルテカ陶器かアステカ陶器かを弁別する能力を身につけた専門家がいるとする（以下、この陶片鑑別師を、ジェイ・ローゼンバーグ（2007）にならって「メグ」と呼ぶことにする）。また、メグは自分の弁別を正当化するために陶片の特徴を引き合いに出すことができないとする。陶片を見る時、彼女は、あるものはトルテカ陶器、またあるものはアステカ陶器だと単にそう思い、例えば「これはアステカ陶器だ」と報告するのである。さらに、彼女は、自分がトルテカおよびアステカの陶片に関する信

> 頼可能な非推論的報告者であるとは信じていないとする。それゆえ、ある破片がトルテカ陶器とアステカ陶器のどちらであるかを同僚に報告したり出版したりする時に、彼女は自分の観察報告をそのまま信用するのではなく、必ず顕微鏡による確認ないし化学的な分析を行う。しかし彼女の同僚たちは、何年にも渡り彼女と仕事をする中で、彼女が実際に陶器の弁別に関する信頼可能な報告者であるということに気づいている。（Brandom 2000, 第 3 章）

この場合、メグは（第一のハードルは越えているものの）第二のハードルを超えていないため、セラーズの基準に従えばメグには陶器に関する知覚的知識は帰属されない。しかしブランダムによれば、「[...] 彼女は実際には当該の破片がトルテカ様式であることを顕微鏡や試薬を用いる以前からすでに知っていたのだ、と同僚たちが述べることは、彼女が正しいことが明らかになった場合には合理的であるように思われる」（Brandom 2000, 98-9）。

しかし、マイケル・ウィリアムズは、この事例に対し次の(1)(2)の批判を投げかける（Williams 2015, 174）。

(1) メグはそもそもその陶片が例えばアステカ陶器のかけらだということを信じていると言えるだろうか。その陶器鑑別師は当該の信念を持っているわけではなく、その陶片はアステカ陶器ではなかろうかとにらんでいるだけであろう。それゆえに、我々は確かにメグの報告を我々の判断の証拠に使うことはできるものの、それは単に彼女を「情報源」として扱っているに過ぎないのであって、彼女は知識を持っているわけではないのではないか。

この批判に際して、ウィリアムズは知識帰属の目的に関するエドワード・クレイグ（1990）の洞察を引き合いに出す。クレイグ曰く、我々は知識という概念を、信頼できるインフォーマントに「旗を立てる」ために用いる（Craig 1990, 11）。その際クレイグは、「インフォーマント」と単なる「情報源」を区別し、単なる情報源ではなくインフォーマントにこそ知識が帰属されるのだと論じる(Craig 1990, Sect. V)。クレイグにおいては、インフォーマントは、次の①と②を満たしている情報源のことである。①その情報源から情報を得るために聞き手に要求される能力が、せいぜい相手に質問をしたり相手の言葉やジェスチャーを理解したりできるだけの言語能力で十分であり、また②その情報源は聞き手に対しコミュニケーションや協働の中で情報を提供する（それゆえに信念や意図を持ってコミュニケーションに参加している）。ウィリアムズはこのクレイグの洞察を受け入れ、メグはインフォーマントではなく（というのもメグは当該の信念を持っていないので②が満たされないため）、単なる情報源でしかないため当該の知識は帰属されない、と論じる。

(2) また、仮にメグが当該の信念を持っているとすると、彼女はどんなメ

カニズムで獲得したのかもわからない信念を何の根拠もなく信じていることになる。しかしこれはローレンス・バンジョーが信頼性主義への反例として提示した千里眼の事例（BonJour 1988）と変わらないため、千里眼の事例における主体の信念を知識とみなすのが不合理であるように、陶器鑑別師の信念を知識とみなすのも不合理である。

以上よりウィリアムズは次のように結論づける：メグに当該の知識を帰属させることはできないのであり、知識を帰属される主体はいくつかの種類の認識的責任を負うことができる必要がある（Williams 2015, 175）。

認識的責任について論じる余裕はないが、次のことを指摘しておく。ウィリアムズはそのような責任を課すことにおいて、自身の信念（例えそれが非推論的であっても）を正当化する能力を持っていることを認識主体に要求している [ii]。ただし、彼はセラーズの第二のハードルをそのまま引き受けるわけではない。主体の信念は、それを疑うべき理由がない限りにおいてデフォルトで権威が帰属されるのであり、文脈的に適切な疑いが生じた場合にのみ主体にはその信念を正当化する必要が生じるのである（Williams 2015, 177）。つまりウィリアムズは主体が当該の信念を正当化できなければならない状況を制限することで、セラーズの第二のハードルを低くするのである。

しかし、例えここまで弱めたとしたとしても、本当に第二のハードルは必要なのだろうか。この点について検討をしたい。

(3) 筆者の主張

確かに、上で見たウィリアムズの批判はもっともらしい。とはいえそれは決定的なものではない。問題は、ブランダムの立場にあるのではなく、彼が出した例の側にあるのである。そこで、次のように事例を書き換えてみよう。

メグは陶片を見るだけでそれがトルテカ陶器かアステカ陶器かを弁別する能力を持っているが陶片の特徴に訴えて自分の報告を正当化することはできない。彼女は遺物の整理の際、陶片が出るたびに、例えば「これはアステカ陶器だと思う。後で調べておいて頂戴」と言って同僚に手渡す。しかし彼女自身は、自分の報告の内容を確信しているが、いちいちそれが真であるかどうかを確かめたりはしないし、そうする必要に迫られたこともこれまでなかった。ゆえに彼女は自分の当該の報告が信頼可能かどうかを知らず、自分が陶片に関する信頼可能な非推論的報告者であるとは信じていない。また彼女は実は無意識的に嗅覚で陶片を鑑別しており、加えて、彼女は自分の報告の正しさに関する最小限の感受性を持っている。すなわち、もし自分が嗅覚的判断の標準的状況にいない時には無意識的に判断を保留するし、判断に自信が持てない時には無意識的に陶片を匂いの嗅ぎやすい位置に移動させること

ができてしまう。一方、いつも破片を手渡される彼女の同僚たちは、長い付き合いの中で、彼女が実際に陶器の弁別に関する信頼可能な報告者であり、しかも実は嗅覚で判断しているということを知っている。

こちらのケースであれば、メグは単にその陶片がアステカ陶器だとにらんでいるだけではなく、確かに信じている。ゆえに(1)の批判は当たらない。また、(2)の批判は、そもそもブランダムに対する論点先取であるように思われる。ブランダムであれば、その信念ないし報告の信頼可能性を知識帰属者が保証できるならば、例え千里眼のような事例であっても知識帰属は適切だと述べるであろう。それゆえ、確かに修正前のメグの事例は千里眼事例と変わらないかもしれないが、だからといってそれが知識の例として不適切だ、とは簡単には言えないはずである。

ところで、修正前のメグおよび千里眼の事例は、信念形成メカニズムが全くもって不明である。彼女らに知識を帰属させることへの抵抗感はこの点にあるのではないだろうか。修正前のメグや千里眼の事例においては、彼女らの信念形成メカニズムが不明であるがゆえに、その知覚の標準的状況も不明である。そのため、彼女らがそもそもセラーズの出発点の条件すらクリアできているかどうかさえ知識帰属者には判定し難く、それゆえに彼らが本当に信頼できるインフォーマントであるかどうかも判断できないであろう。

そこで、上の例では次のような条件をメグの例に追加した。(a)実は（無意識に）嗅覚で判断を下しており、(b)自分がいま、嗅覚的判断がうまく下せない状況（非標準的状況）にいるかどうかを（無意識に）判断して、非標準的状況にいる場合には判断を差し控える(標準的状況においてのみ判断を下す)ことができ、(c) (a)及び(b)を同僚は知っている という三点である。

この状況においては、彼女が信頼可能な報告者だということを知っている人（e.g. 同僚たち、あるいはその同僚たちから彼女の報告が信頼可能だということを教えてもらった人たち）は、たとえメグの陶器鑑別の標準的条件に関する知識などを持っていなくても彼女からその陶片についての情報を得ることができる。加えて、メグは当該の信念を持っており、聞き手と何らかの目的を果たすために協働できるし、中央アメリカの陶器に関するプロなので聞き手が欲しい情報を慮って付加的な情報を付け加えることもできる。以上より彼女は、先の①と②を共に満たしているため、単なる情報源ではなくインフォーマントであろう。そのため、たとえ自分の報告を正当化できないとしても、修正版のメグに知識を帰属させることには問題はないのではないだろうか。ゆえに本研究では、たとえウィリアムズの前提を認めても、ウィリアムズの主張に反して、弱められた第二のハードルが必要だとまでは言えな

いであろうと結論づける。

ただし、このようなケースは、元々のブランダムの事例にさらにいくつもの条件が加わったものであり、事例としてはより一層稀なものとなってしまっている。このような知識帰属のケースはほぼありそうにない。信頼できるインフォーマントに旗を立てるという目的のもとでなされる知識帰属においては、ほとんどの場合において、やはりウィリアムズが提示する弱められた第二のハードルを越えることは必要とされるであろう。更に言えば、おそらく大抵の状況においては、自分で自分に「旗を立てる」ことができなければならないだろう(Craig 1990, Sect. VIII)。というのも、他人が代わりに私の信念の信頼性を保証してくれることなどほとんどないだろうからである。そのため、基本的には、主体は（適切な疑いが生じた時に）自分の信念の信頼可能性を引き合いに出すことができる必要がある。ゆえに、本稿で提示した改善版陶器鑑別師事例のようなほとんど起こりそうにないケースを除けば、ウィリアムズの説明は基本的には正しいと考えて良いのではないだろうか。

(4) 今後の展望

本研究では、ウィリアムズの批判に対してブランダムを擁護しつつ、両者の対立を知識帰属の状況という観点から見直すことにより調停することを試みた。その際、本研究では、クレイグの洞察をガイドとして知識帰属が適切であるような状況の特定を行なったが、これはあくまで暫定的なものであり、知識帰属が適切であるような状況の具体的な特徴づけはさらに行われる必要がある。加えて、ウィリアムズの肩を持つなら、彼がセラーズの第二のハードルを弱める時に提示した「文脈的に適切な疑い」という考えや、最終的に彼が主体に課す「認識的責任」という考えは、さらなる明確化が必要である。どういう状況・文脈のもとであれば知識帰属が適切であったり疑いが適切であったりするのか。また、主体が当該の知識を帰属されるために負わねばならない責任とはどのようなものか。これを具体的に明確化することが今後の課題である。

加えて、以上の洞察においては、成熟した大人の人間が認識主体として想定されていた。しかし、動物や子供に対しても、我々はしばしば知識を帰属させるように思われる。動物や子供に知識は帰属されうるのか。されるならばその条件は何か。この問題に答えることも今後の課題である。

---

[i] 第二のハードルが第一のハードルを含意していることに注意されたい。

[ii] ウィリアムズ(2015)は、認識的責任として次の三つをあげる：(1)説明責任(accountability)、(2)適切な真摯さ(due diligence)、(3)賠償責任(liability)。

それぞれ、(1)自分の持っている信念を説明する責任 (2) 適切かつ真摯に信念を取り扱うことによって認識的義務を満たす責任 (3) 個々の信念について、それが実際には信頼可能ではないということが判明した場合に、その信念の権威が失われるという形で認識的制裁を受けるという責任。ウィリアムズは、(2) の責任に関連して、知識主体であるためには自分自身の信念を正当化する能力が必要だと論じる。(Williams 2015, 175－6)

(5) 参考文献

BonJour, L. (1985). *The structure of empirical knowledge*. Harvard University Press.

Brandom, R.（1997）. "Study guide", *Wilfrid Sellars, Empiricism and the philosophy of mind, with an introduction by Richard Rorty and a study guide by Robert Brandom*, Harvard University Press.

——.（2000）. *Articulating reasons: An introduction to inferentialism*, Harvard University Press.

Craig, E. (1990). *Knowledge and the state of nature: An essay in conceptual synthesis.*

笠木雅史 (2014)「セラーズの所与の神話批判はどの程度成功しているのか？」, 第 73 回日本哲学会、北海道大学, 6 月 29 日.

Rosenberg, J. F.（2004）. "Sellarsian seeing: In search of perceptual authority", *Perception and reality, Ralph Schumacher*（ed.）, Mentis Verlag, Reprinted in Rosenberg（2007）.（References are to the reprinted edition.）

Sellars, W.（1956）. [EPM], "Empiricism and the philosophy of mind", Feigl and Scriven eds., *Minnesota studies in the philosophy of science*, vol. 1, University of Minnesota Press.

——.（1969）. "Language as thought and as communication", *Philosophy and Phenomenological Research* 29, 506–27; reprinted in Scharp, K. & Brandom, R. B. (2007). *In the space of reasons*. 57–80.（References are to the reprinted edition.）

——.（1975）. "The structure of knowledge", H-N. Castañeda（ed.）, *Action, knowledge, and reality: Studies in honor of Wilfrid Sellars*, Indianapolis: Bobbs-Merrill: 295–347.

Sosa, E.（2009）. "The mythology of the given", *Reflective knowledge: Apt*

belief and reflective knowledge, Oxford University Press.
Williams, M. (2015). “Knowledge in Practice”, in *Epistemic Evaluation: Purposeful Epistemology*, 161–185.

（東京大学）

新進研究者 Research Note

コネクショニスト認識論の批判的検討

A Critical Study on Connectionist Epistemology

藤原諒祐

Abstract

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

(1) 研究テーマ

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

(2) 研究の背景・先行研究

2.1. 自然化された認識論，科学的心理学と認識論・科学哲学

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

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

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

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

2.2. 心の哲学における消去主義

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

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

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

2.3. ポール・M・チャーチランドのコネクショニスト認識論

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

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

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

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

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

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

(3) 筆者の主張

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

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

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

3.1. 第二認識論としてのコネクショニスト認識論

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

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

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

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

3.2. 第一認識論としてのコネクショニスト認識論

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

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

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

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

3.3. 方法論的多元主義の可能性

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

第一認識論としてみても，コネクショニスト認識論が ISA アプローチより

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

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

(4) 今後の展望

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

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

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

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

(5) 参考文献

Churchland, P.M. (1979). *Scientific Realism and the Plasticity of Mind*, Cambridge University Press.

Churchland, P.M. (1981). “Eliminative Materialism and the Propositional Attitudes”, in *The Journal of Philosophy*, 78, pp.67-90.

Churchland, P.M. (2007a). “What Happens to Reliabilism When It Is Liberated from the Propositional Attitudes?”, in *Neurophilosophy at Work*, Cambridge University Press, pp.88-112.

Churchland, P.M. (2007b). “Chimerical Colors: Some Phenomenological Predictions from Cognitive Neuroscience”, in *Neurophilosophy at Work*, Cambridge University Press, pp.161-198.

Churchland, P.M. (2012). *Plato’s Camera: How the Physical Brain Captures a Landscape of Abstract Universals*, MIT Press.

Goldman, A.I. (1986). *Epistemology and Cognition*, Harvard University Press.

Quine, W.V.O. (1969). “Epistemology Naturalized”, in *Ontological Relativity and Other Essays*, Columbia University Press, pp.69-90.

Ramsey, W., Stich, S., and Garon, J. (1991). “Connectionism, Eliminativism, and the Future of Folk Psychology”, in *The Future of Folk Psychology: Intentionality and Cognitive Science*, John D. Greenwood (ed.), Cambridge University Press, pp.93-119.

Thagard, P. (1988). *Computational Philosophy of Science*, MIT Press.

Thagard, P. (1989). “Connectionism and Epistemology: Goldman on

Winner-Take-All Networks", in *Philosophia*, 19(2-3), pp.189-196.
戸田山和久 (1999).「科学哲学のラディカルな自然化」,『科学哲学』, 32(1), pp.29-43.

（東京大学）

新進研究者 Research Note

科学哲学からみた心理学における再現性のタイプについて

On the types of replicability in psychology from the point of view of philosophy of science

山銅 康弘

Abstract

I discuss the types of replicability in psychology from the viewpoint of philosophy of science. Specifically, the discussion between the papers of the philosophers of science Radder and the psychologist Schmidt are outlined. After that, I show that the types of the replicability by Schmidt is difficult from both the viewpoint of methods and purposes, and based on the suggestion from the philosophy of experiment by Radder, I clarify that the view of science established on the basis of the hypothetico-deductive method which Schmidt and others implicitly consider is inappropriate in psychology.

(1) 研究テーマ

科学哲学からみた心理学における再現性のタイプについて議論を行う．具体的には科学哲学者 Radder と心理学者 Schmidt の議論を概観する．その後 Schmidt による再現性の枠組みが方法の観点からも目的の観点からも困難が生じることを示す．また Radder の実験の哲学からの示唆を踏まえることで，Schmidt などが暗黙裡に考えている仮説演繹法的な科学観が心理学においては不適切であることを明らかにする．

(2) 研究の背景・先行研究

実験科学の分野において再現性が問題になっている．心理学の有力ジャーナルで掲載された 100 本の論文を追試したところ，元の結果を再現できたと判断できたのが半数にも満たなかったと報告した Open Science Collaboration(2015)は科学における再現性の危機（replication crisis）を分野の垣根を超えて提起することとなり，さらにはその内容が一般紙にまで取り上げられ学際的，社会的な広がりを見せることとなった．

問題は多層的である．例えば上記において「判断できた」と遠回しな表現を用いたことには理由がある．Open Science Collaboration(2015)は再現で

きたことの指標として統計的有意性や効果量，追試者の主観的評価などを用いていており，ある実験結果を再現するとは実際いかなることなのかを示す絶対的な証拠が存在せず，この問題の難しさがうかがえる．

また科学者が用いる「再現性」や「再現可能性」といった用語の不一致もこの問題をより一層難しくしている．Open Science Collaboration(2015)で用いられている“Reproducibility”という言葉はここ最近で定着しつつある用法とは異なっている．この研究では新しいデータを用いる直接的追試（詳細については後述）を行っているが，現在では“Reproducibility”は同じデータを用いて，記述された同じ分析方法で結果が再現できることを指し，訳語としては「再現可能性」をあてることが多い．対してデータや方法など何らかの異なる操作を行い，オリジナルの研究と同様の結果が得られることは“Replicability”と呼び，「再現性」と訳語を分けているようである（e.g. The Turing Way Community et al., 2019; 高橋 , 2018）．本稿ではこれを踏襲し，以後取り上げる文献で使用されている言葉は適宜筆者が判断し現在での用法で言い換えていることに注意されたい[i].

「再現性の危機」は上記の再現可能性と再現性の両方が問題となっている．しかし再現可能性の問題は哲学的というよりも，いかに科学者が実験手順の十分な記述やデータの透明性を担保できるような研究環境を構築できるかという制度設計に関する話が主で,社会科学的側面が強いため今回は扱わない．本稿において中心的な話題となるのは再現可能性ではなく再現性である．

再現性の問題は科学哲学において中心的な話題として取り上げられることは少ない．それは実験の哲学が科学哲学の歴史において過小に評価されてきたことと軌を一にしている．しかし数少ない例外としてオランダの哲学者 Hans Radder が精力的に再現性についての議論を展開してきた．本稿では再現性のタイプにはいくつあるのかという問題を扱い，心理学を対象として取り上げることで昨今の再現性の危機と呼ばれるものについて科学哲学の観点から分析を行う．

具体的には再現性概念の議論の変遷について，主に物理学や化学を対象としたうえで三つの概念を提示した Radder(1992),およびそれを参考に二つの概念に縮約した心理学者の Schmidt(2009)，またこれに対して批判的なコメントを加えた Radder(2015)を軸に議論を行う．その理由として Schmidt の提示した枠組みが現在の心理学では共通了解なものとなりつつあり（e.g. Makel et al., 2012; Earp & Trafimow, 2015），その影響関係に Radder がいることで，再現性という概念が科学においても哲学においても興味深いものとなっているからである．本稿では Schmidt の議論を主に Radder の提示し

た観点から批判を行うが，最終的には以上に挙げた文献すべてについて批判するという立場をとる．

Radder(1992)は再現性を「何を再現しているのか（of what)」という観点から 3 つに分類している．一つ目は「物的実現（material realization）の再現性」，二つ目は「実験プロセス全体の再現性」．三つ目は「実験結果の再現性」である．

このうち最も説明が容易なのは実験プロセス全体の再現性で，これはオリジナルの研究と同様の理論解釈のもと，同様の手順を用いて実験を行い，結果が再現できることを意味している．次に実験結果の再現性は，同様の理論解釈のもとで行うが，結果に至るプロセスはオリジナルの研究と同一である必要はない．最後に物的実現の再現性であるが，これはオリジナルの研究と同様の手順を用いるものの，追試を行うメンバー内で特定の解釈が正しいという共通の信念を必要とせず，そのため実験手順は日常的な言語で記述されている必要がある．例として物的実現を主題として扱った Radder(2012)ではニュートンの質量とアインシュタインの質量で物体を量り同一の結果が得られることなどを挙げている．

このような着眼点は Radder が注力してきた実験の哲学についての背景を概観しておくとより理解できる．Radder(2012)によると，ポパーやハンソン，クーンに代表される科学哲学は，科学をあたかも理論的，解釈的活動だけであるように分析され，実験などによる生産的行為，すなわち物的実現は存在しないか，よくて理論の応用としての派生物だとみなしてきた．確かに生産的行為，本稿で扱う範囲に限定すれば実験は理論的な分析をもとに実行され，実行されたプロセスは理論的に記述される．しかし実験によってひとたび作り出されたものには実在論的位置を占めているため，完全に理論に依存的なものではない．

Radder はさらに物的実現と理論解釈の相互依存性についてさらに分析しているが，さしあたりここで留意すべきは，再現性の種類という問題においては理論の側面と生産の側面という二つの次元から見る必要性があるということである．物的実現の再現性とは，理論を共有しなくとも実現されるような，相互依存性をなるべく生じさせないために要請される再現性ということになる．さらなる詳細については続く Schmidt との議論で明確になると思われるため，次に進むとする．

Schmidt(2009)は Radder の議論を踏まえ，まず三つを(1) 物的実現の再現性，(2) 理論的解釈の再現性，(3) 異なる物的手続きによる実験結果の再現性と表現する．そこから，(2)と(3)の違いは小さいと述べており，(1)をオリジ

ナルの研究と同様の手順で行う「直接的追試（direct replication）」，(2)と(3)を異なる手順で行う「概念的追試（conceptual replication）」と新たに名付けている．ここで違和感を覚えるのは，おそらく Schmidt の用語法であればむしろ違いが小さいのは同様の手順を行っているという(1)と(2)の方であろう．これは Schmidt が Radder の三つの再現性の例として選んだものが誤っていること[ii]に恐らく起因していると思われ，この点については留意されたい．

また Schmidt によるこのような分け方は伝統的に，心理学者である Lykken(1968)などによる追試研究の分類に近く，Lykken は「文字通りの追試（literal replication）」，「操作的追試（operational replication）」，「構成的追試（constructive replication）」の三つに分けており，そのうち「文字通りの追試」は再現可能性のことを指すためそれを除いた二者と中身はかなり類似しており，なじみ深いものだと考えているようである．

最後に Radder(2015)であるが，Radder は Schmidt が二番目の再現性である「実験プロセス全体の再現性」ないし「理論的解釈の再現性」を排除していると述べている．ただし Schmidt(2009)は上記のように「違いが小さい」と述べるだけで排除したとは書いていない．上記のように Schmidt が再現性の例示を誤っているため，排除したと解釈するのが妥当だと判断したのだと思われる．その結果，Radder は一番目と二番目の区別が消失してしまっている点を批判している．「物的実現の再現性」と「実験プロセス全体の再現性」は，オリジナルの研究の手順を正確に再現する必要があるという点では同じであるが，理論を共有している必要性があるかどうかに二つの再現性は違いがある．一番目と二番目を区別しなくて良いのは「物的実現の共通言語記述の全ての側面が実験プロセスの理論的記述で考慮されている場合にのみ[iii]」可能だと述べている．特に心理学では適切で正確な理論が欠如しやすいということから，理論になるべく依存しない形での再現性が重要になってくるため，物的実現の再現性と実験プロセス全体の再現性は分ける必要があるのだという．次節ではより詳細に彼らがどのような観点から再現性を見ているかを明らかにし，またその妥当性を考える．

(3) 筆者の主張

Schmidt はどのような観点から Radder の「物的実現の再現性」を「直接的追試」と呼びなおし，「実験プロセス全体の再現性」と「実験結果の再現性」を「概念的追試」とまとめなおした，ないしは「実験プロセス全体の再現性」を排除し「実験結果の再現性」を「概念的追試」と呼びなおしたのだろうか．

Schmidt(2009)では続く記述で，再現性の機能について述べている．直接的追試にはサンプリングエラーの制御，内的妥当性，研究不正対策，大きな集団や異なる集団への一般化可能性を満たす機能があると述べており，概念的追試には初期の実験の仮説を検証する機能があるとしている．しかし特段機能から再現性のタイプを考えたという記述はなく，むしろ再現性を二つのタイプに分けた結果割り振られる機能について述べているという意味合いが強い．おそらく Schmidt は実際にどう実験を行うのかという方法の観点から再現性を二つにまとめ上げたのだと思われる．そしてこのような方法の観点からの定義は直接的追試について述べている心理学の文献でよくみられるものである（e.g. 加藤, 2018）.「物的実現の再現性」はオリジナルの研究と同様の手順を踏んで行うのに対し,「実験結果の再現性」は結果が再現できれば良く，プロセスは故意に手順を変えているという点で異なる追試なのである[iv].

しかしこの方法の観点からの再現性の分類は上手くいかない．なぜなら直接的追試はオリジナルの研究と同様の手順を踏まなければならないわけだが，それはどこまで正確に記述されたオリジナルの実験手順に合わせるか，ということが問題になる．完全に同一の実験を目指そうにも，部分的に同一の実験を目指そうにも失敗することを以下で示す.

まず直接的追試とは完全にオリジナルと同様の実験を行うことだと考えてみる.ここでは心理学で人を対象とした実験について,「誰に対して行うのか」と「誰が行うのか」という二点に注目してみよう．まず「誰に対して行うのか」であるが，完全に同じ性質を持った人々を別の集団から同じ数だけ選択するということが現実的に不可能であることは明らかであろう．またオリジナルの研究で参加した人に対して再度同じ実験を行うことで，完全に同じ性質を持った人々を対象にすることができるように思われるが，既にその実験を行ったという知識はオリジナルの実験の前には持っていなかったはずで，これはもはや同じ性質を持った参加者（被験者）とは言えない.

次に「誰が行うか」であるが，これも上記と同様で，同じ性質を持った別の実験者を選ぶことは難しい，なぜなら実験者は実験を遂行するための何らかの技能を持っているが，その技能は暗黙知であるため明文化できないことが多い．それゆえある実験を再現できなかった場合に，オリジナルの研究の仮説が間違っていたのか，それとも単に実験者の能力が不足していたために再現できなかったのかは一意的に決まらないということを Collins は多くのケーススタディから明らかにしてきた（e.g. Collins, 1985/1992）.

それでは完全に同一な実験はあきらめ，直接的追試は部分的に同一な実験

を目指すものだとすると，今度は概念的追試との区別が判然としなくなる．ある実験が行われ，その実験手順を元に直接的追試を行う際，どの手順は踏襲する必要があり，どの手順は変更，ないし無視しても影響がないのかを判断しなければいけないが，その基準はいかにして設けられるのか．これはもはや再現性を方法の観点からだけでは分析することはできず，その追試実験が何を目的として行われるのかという目的の観点が必要となってくるのである．

直接的追試と概念的追試について，Schmidt(2009)を参照した Earp & Trafimow(2015)は，直接的追試が特定の要因結果の検証を求めるもので，概念的追試は基礎となる理論や現象を検証しようとするものであると目的の観点から記述している．それでは引き続き方法の観点から直接的追試とは部分的に同一な実験を目指すものだとして，特定の要因結果の検証が求めるという目的の観点を追加で導入し検討しよう．

これによってオリジナルの研究のどの手順は踏襲する必要があり，どの手順は無視しても大きな影響がないのかを判断できるようになると思われる．しかしこれは理論負荷的な行為であることに注意する必要がある．オリジナルな実験が一体どういった理論や現象を検証しようとしているものなのかを，追試者は解釈した上で実験手順を取捨選択し，直接的追試を行わなければならない．ところがこれは先ほどの概念的追試の目的そのものである．従って以上のような観点からでは直接的追試が不可能であると言えよう．Schmidt の直接的追試と概念的追試の枠組みには困難があることが分かる．

(4) 今後の展望

直接的追試の存在に対する過度な信頼の弊害をここでは展望として述べておく．先までの議論は Schmidt による直接的追試，概念的追試の枠組みに対し，Schmidt 的な用法である方法や目的の観点から検討した．この理論的な観点が含まれていないという Schmidt の用法は，言い換えれば「実験プロセス全体の再現性」を排除したという Radder の矛盾のない Schmidt 解釈を採用したうえで検討を行ったということに近いといえる．したがってまだ，直接的追試を物的実現の再現性ではなく実験プロセス全体の再現性として考えるという可能性が残されている．

ところが Radder による実験の哲学の立場からの示唆として，理論解釈と物的実現という二つの次元の問題を元に分析すれば，それもまた困難が生じる．共有された理論解釈のもと，同一の実験手順によって結果を再現するということが科学において重要だという考えは厳密な仮説演繹法を念頭に構築

されているという点に注目されたい[v]．Radder がもともと対象として扱っていた物理学や化学においては，かなりの精度で同様の実験手順を行うことができるため，その結果再現できないことが仮説の棄却を意味し，実際的には論文が撤回される[vi]．

しかし心理学においては完全に統制された条件を実現させることはできない．閉じた実験系（Radder 2012）においてのみ，実際の実験手順が理論的記述と一致する．別の表現をすれば実際の実験手順のほぼ全てを記述することができる．心理学は物理学や化学とは異なり参加者の性質などを完全に統制することは不可能なため，閉じた実験系は実現できず，実際の実験手順が理論的記述と一致しない（例えば個々の参加者の性質などは十分には把握できないため全てを記述することはできない）．したがって厳密な仮説演繹法はここでは成立せず，再現ができないことを理由としてオリジナルの研究を撤回する，ということは心理学という科学において要請されることはない．むしろ再現できないことが，オリジナルの研究で扱われている何らかの対象や概念に対する理解を促進させ，新たな研究へとつながっていくという有益な機能さえ持ちうる．

まとめると，厳密な仮説演繹法を持たない科学においては再現できないことがすぐさま科学としての要件を満たさないということを意味しない．現在「再現性の危機」と呼ばれ議論されている問題点は，その全てが真の危機だというわけではなく，心理学のような科学では原動力と呼べるものさえそこには含まれてしまっているのである．

本稿では理論的分析に終始したが，現在では間違っていると判断される実験結果とその理論的解釈が，結果的には後続の研究を後押ししたというケーススタディを用いて，実証的にも明らかにしていきたいと考えている

---

[i] 前もって述べておくと，Radder(1992) や Radder(2015)では「実験結果の再現性」を“replicability”と呼び，それ以外を“reproducibility”と呼んでいる．また Schmidt(2005)や Earp & Trafimow(2015)においては直接的追試，概念的追試について“replication”を用いているが，これらは全て本論では「再現性（replicability)」に統一している．

[ii] 具体的に理論的解釈の再現性の例として，Schmidt は光の速度には限界があるというアインシュタインの仮説を少なくとも二つの完全に異なる実験系から検証したことが挙げられているが，この例は Radder(1992)では取り上げられていない．

[iii] Radder(2015), p.9, 拙訳．

[iv] Radder(1992)では「実験結果の再現性」を異なる実験手順だけでなくオリジナルの実験手順による結果の再現も含んでいると包含的な概念として

述べている．他方 Schmidt は直接的追試と概念的追試を排他的な概念として捉えており，ここからも Schmidt の枠組みは Radder の枠組みの単なる縮約ではないようである．

[v] 心理学者が心理学に対して仮説演繹法を明示的に使用したものとして Chambers(2017)がある．

[vi] Collins(1985/1992)は物理学であっても再現できたか否かの判断に社会的な力学が作用することを主張するが、本稿ではこれ以上は深入りしない．

(5) 参考文献

加藤司，2018,「『パーソナリティ研究』の新たな挑戦―追試研究と事前登録研究の掲載について」，『パーソナリティ研究』, advpub.

高橋康介，2018，『再現可能性のすゝめ―RStudio によるデータ解析とレポート作成―』，共立出版．

Chambers, Christopher D., 2017, *The Seven Deadly Sins of Psychology A Manifesto for Reforming the Culture of Scientific Practice*, Princeton: Princeton University Press.

Collins, Harry M., 1985/1992, *Changing Order: Replication and Induction in Scientific Practice*, Chicago and London: University of Chicago Press.

Earp, Brian D. and David Trafimow, 2015, “Replication, Falsification, and the Crisis of Confidence in Social Psychology”, *Frontiers in Psychology* 6(621), pp.1-11.

Lykken, David T., 1968, “Statistical Significance in Psychological Research”, *Psychological Bulletin* 70(3, Pt.1), pp.151–159.

Makel, Matthew C., Jonathan A. Plucker, and Boyd Hegarty, 2012, “Replications in Psychology Research: How Often Do They Really Occur?”, *Perspectives on Psychological Science* 7(6), pp.537–542.

Open Science Collaboration (OSC), 2015, “Estimating the Reproducibility of Psychological Science”, *Science* 349(6251), pp.943–951.

Radder, Hans, 1992, “Experimental Reproducibility and the Experimenters' Regress”, *PSA* 1992(1), pp. 63-73.

---, 2012, *The Material Realization of Science: From Habermas to Experimentation and Referential Realism*, Boston: Springer.

---, 2015, “Causality, Theory Ladenness and Reproducibility in Experimental Science”, Conference: 2015 OZSW Conference, At VU University Amsterdam.

Schmidt, Stefan, 2009, “Shall We Really Do It Again? The Powerful

Concept of Replication Is Neglected in the Social Sciences", *Review of General Psychology* 13(2), pp.90–100.

The Turing Way Community, Becky Arnold, Louise Bowler, Sarah Gibson, Patricia Herterich, Rosie Higman, Anna Krystalli, Alexander Morley, Martin O'Reilly and Kirstie Whitaker, 2019, "The Turing Way: A Handbook for Reproducible Data Science (Version v0.0.4)", *Zenodo*, http://doi.org/10.5281/zenodo.3233986

（東京大学）

ソーシャルロボットの倫理における仮想的行為者性概念の可能性と限界
The possibilities and limitations of virtual moral agency in ethics of social robots

水上拓哉

Abstract

This paper addresses the problems of the moral status of social robots by examining Coeckelbergh's concepts of "virtual moral agency" and "virtual moral responsibility." We first briefly summarize the discussion of morality in robot ethics and then present Coeckelbergh's approach. Coeckelbergh's "relational turn" is useful in analyzing social robots because it is impossible to understand the moral competence of social robots by only analyzing the robots themselves. However, the concept of "virtual moral agency" creates conceptual confusion in the ascription of moral responsibility.

(1) 研究テーマ

ロボット倫理学においてはしばしばロボットの道徳的身分、すなわちロボットは道徳的行為者であるか否かという問題に焦点が当てられてきた。本論文では、その中でも社会的役割を担うことが期待されるソーシャルロボットに対象を限定し、その道徳的身分を考える上で M. Coeckelbergh が提案する「仮想的道徳的行為者性（virtual moral agency）」という概念がもたらす貢献とその限界について批判的に検討する。

(2) 研究の背景・先行研究

2.1. ロボット倫理学における行為者性をめぐる混乱

ロボットの引き起こす倫理的問題について取り組む研究領域にロボット倫理学(Lin et al. 2014, 2017)がある。これまでのロボット倫理学において特に重点的に論じられてきたのは「ロボットは道徳的行為者（moral agency）たりうるか」という問題、すなわちロボットの道徳的身分（moral status）についての議論であった。一般的に私たちが道徳的責任を帰属させるのは、何らかの観点において十分な道徳的行為者性をもち、その行為者のすることやしたことの正しさを問えるような存在だけである。だから、ロボットの道徳

的価値や責任について考察する上でもその道徳的身分が問われなければならない、というのが基本的な構図である(Coeckelbergh 2009a, p.218)。

ロボットの道徳的行為者性について、これまでどのような論争が展開されてきたかについては Gunkel(2012)や Sullins(2006)が詳しい。基本的には、ロボットの（将来的なものも含む）能力が人間のそれと比較検討され、ロボットが何らかの水準で道徳的行為者といえるかどうかが議論されてきた。あるものが行為者であるための必要条件のラインナップについて統一的な見解は存在しないが、Himma(2009)が「標準的見解」として提出しているように、少なくとも自由意志や意図の有無が問題になることが多い。

2.2. ロボットの道徳的能力――Coeckelbergh の関係論的転回

初期のロボット倫理学においては、ロボットが何らかの意味で自由であり、また意図をもっているかが議論の中心であった。このとき問題となってきたのは、ロボットが実際にこれらの性質を有しているかどうかである。自由意志や意図といった心的状態を何らかの意味で拡張し、広い意味でこれらの性質をもっていると主張する立場もあるが、そのような立場でもこの「実在指向」は同様である。この考えに基づき、初期のロボット倫理学においてはロボット側に分析の焦点が当てられ、ロボットが本当に何を考えているのか、何らかの意味で心的状態を実際にもっているのかが議論されたきたのである。

一方、この実在への指向を拒絶し、アピアランスに基づいたロボット倫理学を展開しているのが技術哲学者の M. Coeckelbergh である。Coeckelbergh は、ロボットが（その実在とは独立的に）倫理的場面において人間・非人間の境界を容易に乗り越える存在だと考える。つまり、彼にとってロボットは、想像力をかきたてる、感傷的でアピアランスドリブンな存在なのだ(Coeckelbergh 2009a, p.218)。現在のロボットが、人間と同じ意味で心をもった存在だということは難しいが、私たちはしばしば、そういったロボットがあたかも心を有しているかのように振る舞うことがある。そして、それが人間のロボットに対する生き方なのであれば、私たちの倫理的配慮はロボットではなくむしろ人間、すなわち「私たち」が何を考え、感じ、想像するのかに向けられるべきだ、というわけだ(Coeckelbergh 2009a, p.219)。

2.3. 仮想的な道徳的行為者性

Coeckelbergh の提案するロボット倫理学の方法論的転回は、いわば人間中心的かつ関係論的な転回である。では、このアプローチにおいてロボットの道徳的行為者性や道徳的責任の問題はどのように再構成されるのだろうか。

Coeckelbergh は別の論文で、ロボットの道徳的身分を理解するために「仮想的な（virtual）」道徳的行為者性という概念を提案している(Coeckelbergh 2009b)。

ここで Coeckelbergh が問題視するのは、ロボットの道徳的身分の帰属が、自由や意識といった証明困難な条件に依存して行われているということだ。もし仮に技術的人工物が心的状態や意識をもつことが原理的に可能だとしても、それは、個別具体的な人工物がそのような属性をもっているのだと証明する方法があることを含意しない。そしてそれゆえに、道徳的行為者性の標準的見解は苦境に陥るのである(Coeckelbergh 2009b, p.182)。

そこで Coeckelbergh は、この道徳的行為者性と「他者の心」の問題の関係性を、ロボットの文脈で積極的に利用することを提案する。すなわち、私たちは人間同士の道徳的行為者性および道徳的責任の帰属において、実際に特定の人間が心的状態や自由意志を持ち合わせているか否かを確かめていないのだから、ロボットにおいても同様のアプローチが取れるのではないか、と主張するのだ。そして、実在とは独立した形で、ロボットがユーザに対してどのように現れるのか、その経験に基づいて帰属できるような行為者性について言及するために Coeckelbergh が作り出したのが、「仮想的道徳的行為者性（virtual moral agency）」の概念である。

Coeckelbergh の仮想的道徳的行為者性の概念は、いわゆる行動主義（behaviorism）にコミットするものではない。Coeckelbergh は、私たちが心的状態と呼んでいるものが公に観察できる行動パターンに過ぎない、という主張に同意も否定もしないからだ(Coeckelbergh 2009a, pp.219-220)。心についてあくまでも不可知論的な立場を取りつつ、ロボットが与える経験に基づき道徳的行為者性概念を再構築する。それによって Coeckelbergh は、ロボットの道徳的身分に関する厄介な問題を避けつつ、従来の倫理学の用語を用いてロボットの道徳的価値を説明しようとしているのだ。

(3) 筆者の主張

3.1. ロボットが「ソーシャル」になるとき――仮想的行為者性の可能性

Coeckelbergh の関係論的転回、およびそれに基づく道徳的行為者性の再構築は、ソーシャルロボットの倫理の概念的基礎を与える可能性を秘めている。というのも、ソーシャルロボットの「ソーシャル」性は、ユーザが相互行為の中で見出していくものであり、その道徳的影響力についても人間側の想像力が中心的要素となるからである。

私たちにはコンピュータを、わずかな社会的合図（social que）だけで社会

的存在として扱ってしまう傾向性がある。これは今日、「メディアの等式（media equation)」と呼ばれるものである(Reeves and Nass 1996)。つまり私たちは、コンピュータをその内実ではなく、アピアランスに基づいて自分自身と同一視してしまうのだ。

道徳的行為者性の標準的見解においては、ロボットが特定の能力を有しているかどうかが問題になってきた。だが、ソーシャルロボットにおいてはまさにこの「能力」というものが当該のシステムのみで分析しきれない場合がしばしばある。あえて頼りないロボットを作ることで人間の優しさを引き出し、共生的関係を作り出そうとする岡田美智男の「弱いロボット」の研究（岡田 2012）はその好例だといえよう。ソーシャルロボットが道徳的行為者性の必要条件を満たしているのか否かが関係論的に見いだされるものならば、道徳的行為者性そのものもシステムのみに帰属させられるものではない、ということになる。

3.2. 仮想的な責任帰属の実践とは――仮想的行為者性の限界

仮想的行為者性の概念は、ソーシャルロボットの道徳的身分を考える上で有効である。では、道徳的行為者性の標準的見解が想定していたような責任帰属の実践に立ち返るとき、仮想的行為者性の概念はどのような示唆を与えるのだろうか。

Coeckelbergh は仮想的道徳的行為者性をもつロボットに問う責任も同様に仮想的なものでよいと主張し、これを「仮想的道徳的責任（virtual moral responsibility)」と呼んでいる。つまり、道徳的行為者性がアピアランスのみに基づいて帰属できたように、道徳的責任についてもアピアランスに基づき仮想的なレベルで問うことができると主張するのである。このとき、Coeckelbergh は、ロボットにおいては Himma（2009）がいうところの「不快な心的状態」、つまり苦痛を感じたり後悔しているように見えるのであれば、仮想的道徳的責任の帰属としては十分だと述べている(Coeckelbergh 2009b, p.185)。

だが、この仮想的責任の内実は仮想的行為者性に比べていささか表層的なものに見える。以下では Coeckelbergh の仮想的道徳的責任についての議論を確認し、その概念がなぜ表層的なものに留まっているのかを分析する。

3.2.1. 仮想的行為者性を認める基準の曖昧さ

Coeckelbergh が指摘するように、たしかに私たちの道徳的行為者性や道徳的責任の帰属の実践は、実在へのアクセスをせずに行っているといえるかも

しれない。たとえば、ある殺人事件の犯人の責任について考えるときには、その犯人の心についての探求に基づいてその答えを出そうとはしない。より日常的な場面でも、私たちはいわゆる「他者の心」の問題についていちいち悩むことはせず、実際に心をもつ存在として扱っているに過ぎない。

だが、この私たちの道徳的実践をそのままロボットに適用できるとは限らない。上記のような実践は、あくまで人間同士において（ひとまず）うまくいっているものではあるが、そこにロボットという技術的人工物を入れても問題がないとは限らない。

すでに言及したように、私たちは特定の状況の下では、ソーシャルロボットに対して容易に行為者性を帰属させてしまう。メディアの等式に関するNassの研究は、コンピュータがユーザに対して、実態よりも高度なものとして現れる可能性を示唆してきた。また、NeffとNagyは、マイクロソフトのチャットボット“Tay”がユーザから攻撃的な発話を学習して炎上した問題について、言説分析を行った(Neff and Nagy 2016)。調査は事件当時Tayについて言及したツイートを収集し分類するという単純なものではあるが、NeffとNagyは、ユーザの反応としては、Tayを人間の行動の犠牲者として扱うものと、Tayを（いわゆる）人工知能の潜在的な脅威として扱うものがあると分析している。Tayのプログラムに実装されている機能は、人間の心的能力と比べれば極めて単純なものに過ぎないが、それでも私たちは、特定の状況においては、そのような存在に行為者性を帰属させたくなるのだ。

Coeckelberghは仮想的行為者性の定義において、ロボットが行為者として現れる時間的範囲について明確に規定していない。メディアの等式がいうところの「等式」は、いわば錯覚のようなものであり、長期的な関係の構築においてもその等式が保たれ続けるとは限らない。したがって、メディアの等式が示唆するような事実から、私たちが行っているような、「他者の心」の問題を回避する形での行為者性の帰属の正当化に繋げるというのは飛躍がある。

Gerdesは、Coeckelberghの仮想的な道徳的行為者性の概念について、このような概念拡張を導入することは、長期的な視野においては人間同士の関係性を損なう危険をはらんでいると批判している(Gerdes 2016)。先述したように、ロボットが行為者性の必要条件を有しているように見えるためのアピアランスの達成がそこまで高いハードルではないことを考えれば、この危険性はより現実的なものになるだろう。

3.2.2. 「仮想的」を付けることによる責任帰属の構造の崩れ

そしてこのように人間の経験に基づいて行為者性を拡張することは、もと

もとの「標準的見解」で想定されていたような責任帰属の実践において、厄介な概念的混乱を招くことになる。

標準的見解においては、道徳的行為者であることは道徳的責任が問えるための必要条件であった。もともと責任帰属の問題を考える上でのひとつの中心的な基準が道徳的行為者性という概念なのだ。

では、「仮想的」な水準においてはどうだろうか。仮想的道徳的行為者だといえるロボットが仮想的道徳的責任をももっている、といえるとは限らないと思われる。ここで確認したいのは、道徳的行為者としてみなされるアピアランスと道徳的責任をもつものとしてみなされるアピアランスの基準に乖離がある、ということだ。すでに述べたように、仮想的道徳的行為者性の帰属については、ユーザ個々人がロボットのアピアランスをどのように経験するかに依存しており、あるロボットが短期的に道徳的行為者のように見えてしまう可能性はある。だが、このような水準のロボットは仮想的道徳的責任をもつともいえるのだろうか。おそらく Coeckelbergh の立場においてはもつ、ということになるのだろう。Coeckelbergh にとって仮想的道徳的責任が要請するのは、責任帰属対象が苦しんでいるようなアピアランスであった。この水準であれば、道徳的行為者のように見えるロボットが道徳的責任をもっているようにも見える、という構図を作ることはできそうだ。

しかしながら、Coeckelbergh がここで想定している「責任」は、極めて限られた意味での責任でしかない。反省しているように見える、苦しくなっているように見えるという要素は、たしかに私たち人間の責任帰属の実践においても重要ではあるが、そのすべてではない。したがって、仮想的道徳的責任の内実は（仮想的道徳的行為者の概念と比べ）本来の道徳的責任の内実からかけ離れたものになってしまっているのだと思われる。

これに対して、ここでいう仮想的道徳的責任は、あくまでロボットの文脈での責任概念の定義を試みるものであって、その文脈においては処罰に対して苦痛を感じるアピアランスだけで十分なのだ、という反論はありうる。だが、その場合、そのような責任の定義がロボットの文脈で正当化できる根拠を提出しなければならないし、少なくとも Coeckelbergh(2009b)はその議論を十分にしていない。

Coeckelbergh がここで想定している責任概念は、あくまで事後的な処罰に指向したものである。その一方で、そのロボットが将来的にその振る舞いを改善させる、という側面には言及されていない。だが、「責任」の定義においてはむしろ、そのような側面が重視されることも少なくない。

ここでは一例として少年犯罪と少年法における責任の関係を取り上げる。

秋本(2015)によれば、少年法上の責任概念は、再非行可能性と矯正可能性（成長発達可能性）という、非行少年の要保護性が主軸となっている。つまり、ここでの「責任」は、非行や犯罪をなした少年たちの行為に対応した事後的な責任非難を意味せず、処罰や制裁、非難としての法的責任を問うているわけではないのだ。また、髙内(2003)が指摘するように、少年法上の法的責任とは「保護処分を受けるべき地位」として解釈されている。すなわち、（過去の行為の非難としてではなく）将来的な非行反復可能性、また保護処分による矯正教育を通して成長発達を遂げる可能性に基づいて与えられるのがここでの「責任」なのだ。

もしロボットに対して（少年犯罪のように）苦痛を感じるよりもむしろ矯正･成長発達に関するアピアランスを求めるということが､「ロボットの責任」の意味として適切だということになれば、反省する素振りを見せるだけのロボットは、仮想的責任すら帰属できない存在だということになるだろう。

Coeckelbergh は道徳性についての標準的見解である「道徳的行為者→道徳的責任」という構図を議論の出発点としている。そして、ロボットの道徳的身分の考察という文脈でアピアランスの重要性を指摘し、仮想的な水準で行為者性の再定義を行っている。すなわち、ここで「標準的見解」の構図は、「仮想的道徳的行為者→仮想的道徳的責任」という形で拡張されている。しかし、これまで検討してきたように、この拡張版の構図が元の構図の関係性を保つことができるのは、極めて限定的な意味での「責任」においてのみであった。

したがって、もし仮想的道徳的行為者性をロボットの道徳的身分の考察の基礎に据えるのであれば、そこから導かれる仮想的道徳的責任の内実はどのように定義されるべきなのか、という規範的な問いにも取り組まなければならないだろう。そして、もし仮にロボットの文脈においても責任概念は苦痛のアピアランスを超えるものが要請されるのであれば、仮想的行為者性を帰属できる水準と仮想的道徳的責任を帰属できる水準は大きく乖離したものになる。その場合、もともと想定していた「標準的見解」の構造は保たれていないのだから、そのような段階のロボットの道徳的身分を、道徳的行為者性の概念を用いて説明しようとすることの妥当性は疑わしいものになるだろう。

(4) 今後の展望

私たちがソーシャルロボットを人間と近い社会的存在であるかのように扱うという心的傾向は、たしかにソーシャルロボットの道徳的身分を考える上で無視できるものではない。しかし、ロボットを自分と同一視するユーザが

いたとしても、そのユーザが「本当に」そのロボットを同一視しているとは限らない。現在のソーシャルロボットの文脈では、むしろユーザは騙されているのではなく、むしろアピアランスを手がかりに、社会的存在である「体で」扱うという場合が多い。この場合、ソーシャルロボットのユーザがもつ感情は、手品ショーやテーマパークのキャラクター・グリーティングで生じるような「虚構的な」感情の一種として考えられるだろう。このように考えたとき、Coeckelbergh の仮想的行為者性の概念は、K. Walton や M. Ryan などの虚構論を取り入れつつ、ある種の虚構的な行為者性として再構築可能かもしれない。この点については今後詳細に検討していきたい。

(5) 参考文献

Coeckelbergh, M. (2009a). "Personal robots, appearance, and human good: A methodological reflection on roboethics." *International Journal of Social Robotics*, *1*(3), 217–221.

Coeckelbergh, M. (2009b). "Virtual moral agency, virtual moral responsibility: on the moral significance of the appearance, perception, and performance of artificial agents." *AI & society*, *24*(2), 181–189.

Gerdes, A. (2016). "The issue of moral consideration in robot ethics." *ACM SIGCAS Computers and Society*, 45(3), 274-279.

Gunkel, D. J. (2012). *The machine question: critical perspectives on AI, robots, and ethics*. MIT Press.

Himma, K. E. (2009). "Artificial agency, consciousness, and the criteria for moral agency: what properties must an artificial agent have to be a moral agent?" *Ethics and Information Technology*, *11*(1), 19–29.

Lin, P., Abney, K., & Bekey, G. A. (2014). *Robot ethics: the ethical and social implications of robotics*. The MIT Press.

Lin, P., Abney, K., & Jenkins, R. (2017). *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence*. Oxford University Press.

Neff, G., & Nagy, P. (2016). "Talking to bots: Symbiotic agency and the case of Tay." *International Journal of Communication*, *10*, 17.

Reeves, B. & Nass, C. (1996). *The media equation: how people treat computers, television, and new media like real people and places*, Cambridge University Press.

Sullins, J. P. (2006). "When is a robot a moral agent?" *Machine ethics*, 151–160.

岡田美智男. (2012). 『弱いロボット（シリーズ　ケアをひらく）』. 医学書院.

秋本光陽. (2015). 「犯罪少年の『責任』はいかにして組織されうるか」 『ソシオロゴス』, *39*, 191–210.

髙内寿夫. (2003). 「現行少年法における『責任』概念について」 『法政理論』, *35*(4), 75–113.

（東京大学）

われわれは自律的ロボットの行為に責任を負いうるか
Are we responsible for autonomous robots' acts?

姜雪菲

Abstract

Robots are now becoming more and more autonomous. If completely autonomous robots are eventually invented, are we responsible for those robots' acts? If not, it will be the case that there exists nobody responsible for those robots, which is called "the responsibility gap." However, some philosophers say the responsibility gap cannot happen in the first place, and we still are responsible for autonomous robots because of several reasons. In this paper, I argue against these claims; that is, I try to defend the occurrence of the responsibility gap. Then, I would like to suggest my own solution to that problem.

(1) 研究テーマ

ロボット倫理は応用倫理の一部門である。概して、ロボット倫理とは、ロボットの利用によって生じる倫理的な問題を扱う分野である。様々な問題の中で、近年非常に関心を集めているのは、自律的ロボットの行為に対して誰が道徳的責任を負うべきなのかという問題である。もし将来、機械学習などの技術によって完全に自律的なロボットが実現されれば、我々はそのようなロボットの行為に対して責任を負いうる適切な人間を見つけられない状況に陥る可能性がある。この懸念は多くの場合、「責任ギャップ」と呼ばれている。一方、「責任ギャップ」の発生を否定し、そのような問題は生じない、あるいは少なくとも必ず発生するとは限らないと主張する立場がある。本稿では、こうした責任ギャップの発生に否定的な立場の議論を整理して応答を行う。

本稿の構成は以下のとおりである。まず研究の背景において、「責任ギャップ」とはどのような倫理的問題なのか、さらにそれがいかに生じるのかを説明する。そして、この問題に対して様々な否定的な立場の主張を先行研究として提示する。次に筆者の主張で、提示したそれぞれの立場に対して応答を行い、「責任ギャップ」は十分に発生しうる問題であり、それに対して解決策を講じることは重要な課題であると結論付ける。最後に、今後の展望において、一つの可能な解決策を提示する。

(2) 研究の背景・先行研究

ロボットは次第に自律性を獲得する方向へと発展している。完全に自律的なロボットがまだ実現されていない現段階では、プログラマーや操作者など、監督や制御の役割を果たしている人たちこそロボットの行為に道徳的責任を負うべきだと考えている。しかし、もし将来的に完全に自律的なロボットが実現されれば、ロボットの行為に対する道徳的責任を、上記の人々に帰すのは困難になる。なぜなら、自律的ロボットが独力で意思決定ができるようになれば、そのようなロボットの行為を制御し予測することは困難になるからである。それゆえ、我々は責任ギャップに直面する。以下では、この問題をより明確にするために、まず道徳的責任とはどういうことなのかについて説明し、道徳的責任はどのような条件に基づいて、ある特定の人間に帰属するのかを確認することにする。

概して、道徳的責任を負うとは、「ある人の行為が正と不正、徳と悪徳、善と悪などの用語で評価されることができ、そしてその人が自分の行為とその行為によって導く結果に対して非難や賞賛を受けることもできることを意味している」（Mckenna and Pereboom 2016）。従来、道徳的責任の対象は成熟した人間に限定されている。動物や幼い子どもは道徳的配慮の対象にはなるが、自ら責任を負うための十分な能力を持っていないため、我々は彼らを道徳的行為者とはみなさない。

では、我々はどのような条件の下で、ある人の特定の行為に対して、それを非難（賞賛）に値すると考えるのだろうか。基本的には、その人が自分の行為をコントロールすることができて、かつ自分が何を行っているのかを知らなければならないことが要求される（Talbert 2019）。これらはそれぞれコントロール条件と認識的条件として規定される。例えば、ある人が悪い催眠術師によって強引に催眠術にかけられ、術師の指示に従って詐欺を実行したとしよう。この場合、催眠された人は悪いことを行ったが、催眠されている状態のため、自分の行為をコントロールすることができないし(コントロール条件の欠如)、しかも自分が何をしているのかに関しても十分な知識をもたないので(認識的条件の欠如)、自分の行為とその結果に関して責任を要求されることはないだろう。

上記の二つの条件のもとで、自律的ロボットが利用される場面を考えよう。例えば、自律武器システム（AWS）が司令官から命令を受けて、自ら敵軍に攻撃を行うとする。それらは人間の兵士と同じように、具体的な攻撃の対象を自分で判断し、どのように攻撃を行うのかも自分で決める。すなわち、AWSは戦場においてはほとんど人間の兵士と同じような役割を果たしている。AWSがこのような能力を持つのは、主に実装される学習技術によるものであ

る。このような技術はAWSの自律性を支えるが、AWSの本来の開発者たちも含めて、AWSが学習技術を通して習得する新たな情報、そしてその情報を用いてAWSが下す判断に関して、それを予測したり、それに干渉したりすることはできない。このとき、開発者や司令官の本来の意図や命令から外れて、AWSが有害な結果を引き起こしたならば、誰が道徳的責任を負うべきなのだろうか。

責任ギャップが発生しうると考えている論者たちは、このような場面では責任を負いうる適切な対象はいないと主張する（Matthias 2004, Sparrow 2007, Danaher 2016）。例えば、ロバート・スパローは、AWSが引き起こした損害に対して、プログラマー、司令官、あるいはAWS自体が責任を負う可能性を検討したが、責任を負うべき対象は存在しないと結論づけ、そのような帰結を招くAWSの開発自体を控えるべきだと主張するに至った(Sparrow 2007)。

一方、責任ギャップの発生を否定する論者たちは、それぞれの理由に基づき、自律的ロボットによって生じる損害に対して責任を負う対象を特定することは可能だと主張し、責任ギャップをめぐる議論に異を唱える。筆者の見解では、こうした試みはおおよそ二つの方向性に分類できる。第一の方向性は、たとえロボットが自律的になっても、依然としてそれに対する何らかのコントロールをもつ人間が存在するため、責任ギャップは生じないというものである。第二の方向性は、自律的ロボットの行為はある程度、または全面的に人間に依存するので、責任は関連する人たちが負うべきだという考えである。前者としては、Himmelreich（2019）やSchulzke（2013）が挙げられ、後者としてはNyholm（2018）やRobillard（2018）が挙げられる。以下では、それぞれの主張の要点を示そう。

(2)-1 自律的ロボットに対しても何らかのコントロールをもつ対象がいる

ヨハネス・ヒンメルライヒは、たとえAWSが完全に自律的になったとしても、上位にいる司令官は、命令を下すことによって最終的な結果に対しある程度のコントロールをもつことができるので、結果に対して責任を負うべきであると主張する。ここでヒンメルライヒは、たとえ不適切な目標への攻撃といった、望ましくない結果を実際に引き起こすのがAWSであるとしても、その発生は司令官たちが下した命令まで遡ることができると述べる。たとえ司令官らがAWSの特定の攻撃や、その攻撃に伴う具体的な結果に対して直接コントロールすることができなくとも、誰かが攻撃されるという確率的な結果、そして不適切な目標に危害を加えるリスクに対してはコントロー

ルすることができるからである。加えて、戦争で、本来の意図と外れた結果が発生すること自体は一般に予見可能であるため、司令官は AWS が機能不全などによって、望ましくない結果をもたらす可能性があることは認識できる。したがって、司令官らはコントロール条件だけでなく、認識的条件も満たすことになり、AWS によって引き起こした損害に対して道徳的責任を負いうることになる。ゆえに、ヒンメルライヒは、たとえロボットが完全に自律的になっても、責任ギャップは生じないと結論づけるのである。

同様にして、マーカス・シュルツケは、戦争中のＡＷＳ使用に焦点を当て、AWS の行為に対しては責任を負う適切な対象が存在し、責任ギャップは生じないと主張した。しかし、ヒンメルライヒとは異なり、シュルツケは責任帰属の対象には司令官だけではなく、AWS の開発者たちも含まれると主張する。なぜなら、以下で述べるように、司令官も開発者もＡＷＳに制限を加える対象である点では、共通する振る舞いをもつからである。

シュルツケによると、司令官の道徳的責任は、軍事的ヒエラルキーというシステムによって定められるものである。戦争に関するすべての意思決定は、ヒエラルキーの上位にいる人たちが行い、下位にいる兵士たちは実行できる行動の範囲が限定され、特定の行動のみが許可または禁止される。このようにして、兵士の自由は厳しく制限されているため、兵士は、指示された通りに遂行した自分の行為に対して責任を負わされることはほとんどない。代わりに、責任は意思決定者たちに分配されている。ゆえに、将来的に人間の兵士が AWS に置き換えられたとしても、同様のシステムによって、責任はヒエラルキーの上位で意思決定をつかさどる人間たちに帰属することになる。

さらに、シュルツケは AWS の開発者も責任を負うべきであると主張する。シュルツケによれば、ロボットの操作は初期の設定に強く依存しているため、ロボットが損害を引き起こしうるのは、開発者が適切な制約を加えることに失敗しているからである。AWS が新たな情報を学習することができるにもかかわらず、それらの総合的な行動の範囲は初期の設定によって制限されるため、それらの能力と傾向は開発者まで遡ることができる。ゆえに AWS が誤った学習を開始しないよう制限を加えないことは、開発者たちの過失であるとシュルツケが述べた。

つまりシュルツケから見ると、司令官や開発者は自律的ロボットに制限を加えることができるため、もし損害が生じれば、彼らが適切な制限を設定することに失敗したことに対して責任を負うべきである。それゆえ、自律的ロボットの行為に対しては責任を負う対象が存在し、責任ギャップは生じない。

(2)-2 ロボットの行為者性を否定する

スヴェン・ニーホルムとマイケル・ロビラードは、ロボットが完全な行為者性を持たないという点で共通の見解をもつ。ただし、ニーホルムはロボットがある程度の行為者性を持つことを認める一方、ロビラードはロボットが全く行為者性を持たないと考えている。

自動運転車や軍事ロボットなどの自動化されたシステムの行為者性について、ニーホルムは人間―ロボットの協同という形で理解すべきであると提案する。なぜなら第一に、これらの自動化されたシステムの動きは特定のルールや原則に規制され、しかもそれらの動作に干渉することができる権威者が存在するからである。例えば自動運転車の場合、自動運転車が独力で自分の行き先を決めることはできない。すなわち、目的地は人間によって設定されているため、自動化されたシステムの自律性とは単に人間が設定した目標に向けて実行する動作のことにすぎない。この意味で、それらは独力で行為を行うとは言えず、必ず誰かほかの対象と一緒に行動しなければならないといえる。ゆえに、たとえ自動化されたシステムがほとんどの作業を自分で行ったとしても、それらの行為は他の対象による監督に依存しているので、それらの行為に対して責任を負うべきなのは、監督の役割を果たしている人間であるとニーホルムは主張する。

ロビラードはさらに強くロボットの自律的な行為者性を否定する。ロビラートによれば、ロボットの行為者性は企業や政府などと類比的に考えられるべきである。すなわち、ロボットは単に「物理的な形で例示化された社会的に構築された組織にすぎない」（P.713）ため、それらのあらゆる行為には人間の本来の意図と決定を超えることはない。そして責任はこの組織を構成する人間たちが負うべきである。もちろんここで集団が意思決定を行うという主張自体に対して、つまり集団主義（collectivism）に対して反論を加えることはできるが、ロビラードはその種の反論は既に議論されている問題であって、AWS に特有の問題ではないとして退ける（P.713）。

以上のように、ニーホルムとロビラードはともに、ロボットの行為に対する責任は関連する人間に帰属すべきであり、責任ギャップは発生しないと結論づけている。

(3) 筆者の主張

私の考えでは、責任ギャップに対する上記の二つのタイプの反論は、いずれも責任ギャップの発生を完全に排除することに失敗している。すなわち、私の考えでは責任ギャップは依然として生じうる問題である。

まず、自律的ロボットに対するコントロールすることができる対象がいるという主張する論者たちに対して、以下の点を指摘して応答したい。

第一に、AWS の司令官だけが、最終の結果に対してコントロールをもち、一般的な予測もできる、というわけではない。例えば AWS の開発者も AWS を開発するかどうかを決めることができるという点では、最終の結果に対してコントロール持っているといえる。このような漠然とした基準に基づいて責任を負うべき対象を決めることは、関連のない対象へと責任の範囲を過剰に拡大してしまう恐れがある。

第二に、確かに現在は軍事ヒエラルキーの上位の対象が下位の行為を統率し、責任を負うというシステムが機能しているかもしれない。しかし、このようなシステムがずっと維持され続けるかどうかは誰も保証することができない。ルチアーノ・フロリディ（2017）は、技術だけからなるループを三次技術と呼ばれ、それが実現されれば人間のユーザーは、「もはやループの中には存在せず、せいぜいそれに接しているだけである」（P.38）と考えている。このことは、AWS に対する軍事ヒエラルキーに対しても当てはまりうる。すなわち、軍事システムの全体がロボットや AI によってなされる状況は想定可能である。もしそのような状況になれば、統率役のロボットに責任を帰属させない限り、同様のプロセスで自律的ロボットの行為に対する責任を誰かに帰属することはできない。

第三に、たとえ司令官や開発者がロボットに制限を加えることができるとしても、すべて可能な結果を予測することはできない以上、あらゆる望ましくない結果の発生を防ぐことは不可能であると思われる。自律的ロボットに加える制限には限界があるのであり、限界を超えた事象に対して、制限を加えたものに全ての責任を負わせることは不合理である。

以上のように、ヒンメルライヒとシュルツケは自律的な行為者の行為が行為者自身以外に帰属するケースやシステムを持ち出して、自律的ロボットにも適用しようとしているが、両者の想定はいずれも不十分である。すなわち、両者は、自律的なロボットの出現によって、標準的な責任帰属条件が完全に適用不可能になる可能性に、十分な注意を払っていない。

次に、自律的ロボットの行為者性を否定する論者たちの主張について、以下の点から応答しよう。

第一に、確かに、自動運転車や軍事ロボットなどの自動化したシステムの最終的な目標は人間によって設定される場合が多い。しかし、たとえロボットがそれ自身では目標を持たないとしても、設定された目標に向けて自らで動作する過程では、どのようにタスクを達成するのかについて、自身で判断

を下し、具体的な行為を実行することができると想定されている。つまり、この過程において、自律的ロボットは従来人間の役割を完全に代替することができるようになる（Coeckelbergh 2016）。そうであれば、ロボットの動作を監督し、権威のある人が実際の状況でロボットの一連の動作全てに介入することはないと思われる。そして、自律的ロボットが独自の意思決定を経て、最終的な目標から逸脱し、損害等、意外な状況が発生させることは十分にあり得る。少なくともこのような状況に対しては、責任ギャップは依然として発生しうると考えられる。

第二に、個人でも、集団でも、意図されていない結果に対してすべての責任を持たせることは不適切である。例えば、人間の兵士が意図的に不適切な対象を攻撃する場合、司令官にそれに対する責任があるが、兵士自身に全く責任がないとはいえない。同じように、自律的ロボットの開発または利用において、ロボットが全く意図せざる行為を行った場合でも、司令官や開発者が道徳的に非難されることは妥当ではないと私には思われる。

以上のように、責任ギャップの発生を否定する論者たちの主張はいずれも不十分であって、決定的ではない。すなわち、責任ギャップは依然として発生しうる問題であり、それに対する対処法を考えることは重要である。

(4) 今後の展望

責任ギャップが発生しうることが認められるならば、次の課題は、それが深刻な影響を及ぼす前に、予め解決策を講じることである。ここではその解決に向けて、一つの方法を提案したい。それは、自律的ロボット自身に道徳的責任を帰属することによって、責任ギャップを埋めるというものである。ここで、ロボットに責任を持たせるという提案自体は既に提出されている（Bechtel 1985, Stahl 2006）。しかし、彼らはいずれもロボット自身が道徳的責任を負うことができるように明確な基準を記述できていないという問題を抱えている。

では、自律的ロボットはどのような条件を満たせば、道徳的行為者とみなされるのだろうか。標準的な枠組みのもとでは、道徳的行為者は成熟した人間に限定されるため、道徳的行為者性の条件はそれに応じて、自由意志や人間性などであると考えられてきた。しかし、これらの条件は、特定の行為を個人に帰属するという道徳的責任の一つの側面だけを重視するものである。こうした条件は、道徳的責任が、行為者と道徳的コミュニティ(moral community)における他のメンバーとの関係において果たす役割について、十分に表していない(Watson 2004)。ゆえに、我々は自由意志や理性など人間

特有の性質に頼らずに、道徳的行為者であることの条件を再考する必要がある。そこで、精神をもたない行為者が道徳的行為者となるための条件を整備することが、本論で論じてきた問題のもっとも見込みのある解決策であると私は考える。

(5) 参考文献

1. Bechtel, W. (1985) “Attributing responsibility to computer systems.” *Metaphilosophy* 16 (4):296-306.
2. Coeckelbergh, M. (2016) “Responsibility and the Moral Phenomenology of Using Self-Driving Cars.” *Applied Artificial Intelligence* 30:8, 748-757
3. Danaher, J. (2016) “Robots, Law and the Retribution Gap.” *Ethics and Information Technology* 18 (4):299–309.
4. Himmelreich, J. (2019) “Responsibility for Killer Robots.” *Ethical Theory and Moral Practice* 22 (3):731-747.
5. Matthias, A. (2004) “The responsibility gap: Ascribing responsibility for the actions of learning automata.” *Ethics and Information Technology* , 6: 175–183.
6. McKenna, M. & Pereboom, D. (2015) *Free Will: A Contemporary Introduction* Routledge..
7. Nyholm, S. (2018) “Attributing Agency to Automated Systems: Reflections on Human–Robot Collaborations and Responsibility–Loci.” *Science and Engineering Ethics* 24 (4):1201-1219.
8. Robillard, M. (2018) “No Such Thing as Killer Robots” *Journal of Applied Philosophy* 35 (4):705-717.
9. Sparrow, R. (2007) “Killer Robots.” *Journal of Applied Philosophy*, Vol. 24, No. 1.
10. Schulzke, M. (2013) “Autonomous Weapons and Distributed Responsibility.” *Philosophy and Technology* 26 (2):203-219.
11. Stahl, B. (2006) “Responsible computers? A case for ascribing quasi-responsibility to computers independent of personhood or agency.” *Ethics and Information Technology* 8 (4):205-213.
12. Watson, G. (2004) “Two faces of responsibility.” *Agency and Answerability: Selected Essays*, 260-288.
13. フロリディ, L, 2017, 春木良且・犬束敦史訳者代表, 『第四の革命―情報

圏(インフォスフィア)が現実をつくりかえる』 新曜社.

（大阪大学）

情報概念の真理性は正当化できるか
Can we justify that the concept of information requires truth?

榎本　啄杜

Abstract

The concept of information is ambiguous and has different meanings depending on the context in which it is used. On the other hand, there is a strong belief that “information” that generates knowledge must encapsulate truth. Luciano Floridi is the defender of this idea, but in his argument, there seems to be some leaps in logic and inaccuracies in formula. In this paper, I try to reconstruct his argument and interpret it as reasonable one. However, some questions still remain, so I shall mention them at the end.

(1) 研究テーマ

知識論の分野において、情報概念がどのように使用されているかを検討することを通じて、情報概念の分析・精緻化を試みる。

(2) 研究の背景・先行研究

情報の哲学における主要な問題として、情報概念で知識をどう説明するかというものが挙げられる。たとえば、F. ドレツキ（1981）は知識と情報の関係について、以下のように述べている。

> この意味[1]では、誤情報は情報の類ではない。（中略）情報とは知識を生み出すことができるものであり、知識が真理を要求する以上、情報も真理を含意するはずである。（Dretske 1981, p.45）

ここでは、知識のもつ真理性は、それを引き起こす情報の真理性に由来していると考えられている。ここで問題となるのは、知識にとって必要とされる情報である「意味論的情報（semantic information）」が真理性をもつことは必ずしも自明ではなく、当該分野においても論争が続いているということである。

これを受けて L. フロリディ（2011）は、意味論的情報が真理性をもつことを示すために、形式的な議論を行っている（Chapter 4）。しかし、この議

論は本来あるべき前提や推論過程が省略されており、一見すると真理性を獲得するまでの過程に飛躍があるように思える。また、論理式に不正確な部分があることによって、解釈に困難をきたす箇所も見受けられる。本稿では、これらの不足部分を補うことを通じて当該議論を再構成し、筋の通ったものとして解釈し直すことを試みる。

フロリディの議論を概観する前に、前提となる背景を確認しておく[2]。情報科学やそれに関連する分野[3]で広く受け入れられている意味論的情報の定義として、GDI（General Definition of Information）[4]がある。これはデータ概念に基づくもので、意味論的情報は「（ⅰ）1つ以上のデータで構成されており（ⅱ）well-formed で[5]（ⅲ）意味をもつ」という3つの条件で定義される。定義からも明らかなように、GDI は真理値に関する条件をもたない。フロリディによると、GDI は真理中立的であるために誤情報やトートロジーも情報の一種として認められることになってしまうため、意味論的情報の必要条件ではあるが、十分条件にはなっていない。これを解決するためには、GDI の定義に「（ⅳ）真である」を付け加えなければならない。

上記を達成するために、背理法を用いて以下のような意味論的議論（semantic argument）がなされる。使用する語彙は以下の通り。

D = $\{p_1, \cdots, p_n\}$；D は空でもよい命題の集合（GDI の集合）[6]
φ, ψ；D 上の命題変数
S = $\{i_1, \cdots, i_n\}$；S は空でもよい意味論的情報の具体例の集合
x, γ；S 上の命題変数
t (φ)；φ は真である（事実式）
f (φ)；φ は偽である（事実式）
t / f (φ)；φ は真である、または偽である（事実式）
T(φ)；φ はトートロジーである
C(φ)；φ は自己矛盾式である
H(φ)；φ の情報量[7]
P(φ)；φ の確率[8]

なお、S は D の部分集合であり、意味論的情報はすべて命題の形をとる。また、彼によると、意味論的情報のモデルは①P.1～P.4 の 4 つの原理を必ず実装し、②R.1～R.3 の 3 つの要請をできるだけ満たすことが求められる。4 つの原理（P）と 3 つの要請（R）は、以下の通りである。

P.1：情報量の非負性

$(\forall x H(x) \geq 0)$

P.2：情報量の加法性

$(\forall x \forall \gamma((x \neq \gamma) \rightarrow (H(x \cup \gamma) = H(x) + H(\gamma))))$

P.3：逆関係原理（Inverse Relationship Principle（IRP））

$(\forall \varphi(P(\varphi) = 1) \rightarrow (H(\varphi) = 0))$ [9]

P.4：情報量が 0 の命題は意味論的情報ではない

$(\forall \varphi((H(\varphi) = 0) \rightarrow \neg(\varphi \in S)))$

R.1：情報量は物理的なダメージがなくてもシンタクス的に減少しうる

R.2：情報リポジトリに任意の命題（事実式）を追加し続けても情報量は増加しづらい

R.3：情報量は物理的にもシンタクス的にも意味論的にも減少しうる

なお、P.1 と P.2 は変数 x, γ が用いられているため S の議論領域、P.3 と P.4 は変数 φ が用いられているため D の議論領域であると想定されている。前述したように S は D の部分集合であるため、D の議論領域である原理は S にも適用可能であるが、その逆は成り立たない。

まず、S と D を同じ集合だと仮定する。つまり、すべての命題は意味論的情報であると仮定する。すると、P.3 と P.4 から次のように言える。

$T(\varphi) \rightarrow (P(\varphi) = 1)$

$(P(\varphi) = 1) \rightarrow (H(\varphi) = 0)$

$\underline{(H(\varphi) = 0) \rightarrow \neg(\varphi \in S)}$

$T(\varphi) \rightarrow \neg(\varphi \in S)$

S と D は同じ集合であるのに、トートロジーは意味論的情報ではないという事実が導かれた。また、次のようにも言える。

$(P(\varphi) = 0) \rightarrow (P(\varphi) < 1)$

$(P(\varphi) < 1) \rightarrow (\varphi \in S)$

$\underline{C(\varphi) \rightarrow (P(\varphi) = 0)}$

$C(\varphi) \rightarrow (\varphi \in S)$

この結論は Bar-Hillel－Carnap Paradox[10] を引き起こす。P.3（IRP）によ

れば、出来事の生起確率が低くなればなるほど情報としての価値は増すため、確率が 0 になる自己矛盾式は非常に大きな情報量をもつ。しかし、$p \wedge \neg p$ が非常に大きな情報量をもった情報として認められるというのは、避けられるべき事態である。GDI は確率についての前提を含まないために問題とならなかったが、この 2 つの難点は確率的アプローチをとると不可避なものである。

S と D を同一視するこの考え方、つまり GDI こそが意味論的情報であるという考え方は、前述のように様々な分野で受け入れられている。しかし、いくつかの原理を適用するだけで上記のような矛盾が生じてしまうため、この仮定を棄却し、これらの難点を排除するような修正を加える必要がある。これは、GDI に以下の規則を付加することで解決できる。

［1］ $\forall \varphi((T(\varphi) \vee C(\varphi)) \rightarrow \neg(\varphi \in S))$

これにより、すべての命題のうち、トートロジーと自己矛盾式以外の命題が意味論的情報になる。つまり、真偽はどうあれ、事実式でなければ意味論的情報としては認められない。こうして導かれた GDI＋［1］のモデルは、意味論的情報が真理性をもたないと考える論者でも受け入れられるものである。

ところで、現時点のモデルにおいて、P.1 と P.2 から次のように言える[11]。

［2］ $\forall \varphi \forall \psi((\varphi \neq \psi \wedge t/f\ (\varphi) \wedge t/f\ (\psi)) \rightarrow (0 < H(\varphi) < H(\varphi \cup \psi) > H(\psi) > 0))$

［3］ $H(\cup_1^n \varphi) < H(\cup_1^{n+1} \varphi)$

［3］は［2］を一般化したものである。これによると、任意の命題をとり続ければいくらでも総情報量が大きくなり続け、物理的にダメージを与えでもしない限りは、減少することはありえない。

しかし、ある命題のすぐ後にそれを否定する命題をとる操作を想定すれば、現在のモデルは、シンタクス的に情報量が減少してしまう現象を説明できないことに気が付く。たとえば、命題 p を受け取った後に命題¬p を受け取ったとすると、どちらの方が信頼できる命題であるかを判断できない以上、新しい命題 $p \vee \neg p$ は情報量をもたない。4 つの原理だけではこの現象に太刀打ちできないため、現在のモデル GDI＋［1］は、さらに R.1 と R.2 の 2 つの要請を満たすべきであると考えられる。

R.1：情報量は物理的なダメージがなくてもシンタクス的に減少しうる

［4］ $\Diamond(H(\cup_1^n \varphi) > H(\cup_1^{n+1} \varphi))$（physically and syntactically）

R.2：情報リポジトリに任意の命題（事実式）を追加し続けても情報量は増加しづらい

［5］ $P(H(\cup_1^n \varphi) \geq H(\cup_1^{n+1} \varphi)) \geq P(H(\cup_1^n \varphi) < H(\cup_1^{n+1} \varphi))$

R.2 は R.1 をエントロピー的に修正したものである。これによって、今や GDI はトートロジーや自己矛盾式、さらには複数合わさることで矛盾してしまうような事実式の集合を排除できるようになった。

ところが、この改訂版 GDI ではシンタクス的な情報量の減少は説明できるが、意味論的な減少（semantic loss）はまだ説明できていない。元の真なる命題に誤った言明を加えてしまえば、元の命題が情報量のストックに貢献していた分の情報量は減少してしまうにもかかわらず、手持ちの規則だけでは情報量が増加することになってしまう。偽である命題はいくらでも作ることができるので、このような仕方で情報量が単調に増加してしまう場合もまた排除しなければならない。これを考慮すると、GDI はさらに R.3 の要請を満たすべきである。

R.3：情報量は物理的にもシンタクス的にも意味論的にも減少しうる

［6］$\Diamond(H(\cup_1^n \varphi) > H(\cup_1^{n+1} \varphi))$（physically, syntactically and semantically）

R.3 は R.1 と R.2 を含んでいるため、R.3 を満たす規則を付加すれば、すべての要請を満たすことができる。こうして得られる規則が、次である。

［7］ $\forall \varphi((\varphi \in S) \rightarrow t(\varphi))$

［7］は定義から自明に［1］を排除しており、R.1 と R.2 は R.3 に含まれているため、［7］だけでこれまでの問題をすべて解決することができる。よって、意味論的情報のあるべきモデルは GDI＋［7］である。

(3) 筆者の主張

以上がフロリディの議論の概要である。［1］を導くまでは比較的明快に議論が進むが、3 つの要請を持ち出してから［7］を導き出すまでの過程に、一見すると飛躍があるように思える。そのような感覚は、ここに論理式の誤りと推論過程の省略があるために生じるものである。これを補ってやれば、フロリディは当初の宣言通り背理法を用いて、段階を踏んだ議論を行っていることがわかる。

まず、［2］と［3］は D 上の変数を用いて記述されていたが、正しくは S 上の変数を用いて、以下のように記述されるはずである。

$$［2'］\ \forall x \forall \gamma((x \neq \gamma \wedge t/f\ (x) \wedge t/f\ (\gamma)) \rightarrow (0<H(x)<H(x \cup \gamma)>H(\gamma)>0))$$

$$［3'］\ H(\cup_1^n x) < H(\cup_1^{n+1} x)$$

この段階（GDI＋［1］）で意味論的情報として認められているのは、トートロジーと自己矛盾式を除くすべての命題であり、真偽を問わず事実式の集合である。［2］と［3］は事実式について述べているのに加えて、議論領域が S である P.1 と P.2 を適用していることは明らかであることから、S 上の変数である x と γ を用いて［2'］［3'］のように記述することが求められる。

さらに、フロリディの記述には以下の推論が欠落している。現在、GDI＋［1］と S は同じ集合であると仮定している。S（つまり GDI＋［1］）では P.1 と P.2 から［2'］と［3'］が成り立っているはずである。しかし、この集合の一部の要素において、R.1 と R.2（［4］と［5］）で説明されるような現象（S 上の変数を用いて$\Diamond(H(\cup_1^n x) > H(\cup_1^{n+1} x))$）が実際に成り立っており、［2'］［3'］と矛盾してしまう。具体的には、非負性をもつ情報量の加法性が GDI＋［1］の内部では明らかに成り立っていない。これは、GDI＋［1］と S を同一視する仮定によって生じた矛盾であるため、この仮定は誤りである。つまり、S は GDI＋［1］の集合よりもさらに小さい集合であり、その S は、上記の矛盾を引き起こした原因である、合わさると矛盾している事実式を GDI＋［1］から除いたものと一致する。

R.3 についての議論も同様に言える。上で導かれた暫定版 S を仮に S'とすると、S'と S は同じ集合であると仮定できる。S'では［2'］と［3'］が成立するはずだが、これと矛盾するはずの R.3（$\Diamond(H(\cup_1^n x) > H(\cup_1^{n+1} x))$）が S'内で成り立ってしまい、矛盾する。これは S'と S を同一視する仮定によって生じた矛盾であるため、この仮定は誤りである。つまり、S は S'よりもさらに小さい集合であり、その S は、上記の矛盾を引き起こした原因である、偽の事実式を S'から除いたものと一致する。

このように背理法を適用し続けて最後まで残ったのは、以下のように真の事実式だけである[12]。

<table>
<tr><td>①トートロジー</td><td colspan="2">②合わさると矛盾する事実式</td></tr>
<tr><td>①自己矛盾式</td><td>③偽の事実式</td><td>④真の事実式</td></tr>
</table>

この図全体で D の集合、つまり GDI である。①まずトートロジーと自己矛

盾式が含まれることで矛盾が生じたため、これらを D から除いた。この時点で残っているのは事実式の集合である。導かれた集合を D'とすると、②合わさると矛盾してしまう事実式が含まれることで矛盾が生じたため、これも D'から除いた。導かれた集合を D''とすると、③偽の事実式が含まれることで矛盾が生じたため、これも D''から除いた。④すると残るのは真の事実式だけである。これを S と同一視してもこれ以上の矛盾は特に生じないため、S の要素であれば、それは真の事実式であると結論してもよい。

フロリディの議論を再構成する作業は以上である。後半から飛躍していたように思えた当該議論も、4 つの原理と 3 つの要請の正当性を受け入れる限り、一貫して背理法を用いた合理的な論証になっていると言ってよいだろう。また、当該議論は集合 S と D の議論領域の違いに大きく依拠していることもわかった。もしも［2］［3］が原文の通り D の議論領域であるとすれば、同じ D 上で［2］［3］と R.1～R.3 を同時に受け入れることになってしまい、議論の出発点で矛盾が生じてしまう。当該議論における背理法は、S と D の議論領域の違いがなければ成り立たないものだと言える。

(4) 今後の展望

「筆者の主張」において、「4 つの原理と 3 つの要請の正当性を受け入れる限り（中略）合理的な論証になっている」、「当該議論は集合 S と D の議論領域の違いに大きく依拠している」と述べた。これはつまり、前提の正当性が失われることがあれば、合理的な論証とは言えなくなることを示唆している。この点において、以下の疑問が残る。

たとえば P.1 情報の非負性について、当該議論においては H(φ)の計算方法が明記されていないため、これだけを見ていては、どのような値をとるのかが不明である。確率的アプローチをとるのであれば対数関数を用いた自己情報量の計算方法[13]が一般的だが、この計算方法であれば、どのような命題についても定義上自明に負の値をとらない。また、当該議論がなされている章の次の章（Chapter 5）で真理値に基づいた情報量の計算方法が提案されている[14]が、この計算方法でも情報量は負の値をとることはない。すると、P.1 は S 上の議論領域であるだけではなく、D 上の議論領域でもあるということになる。その他の原理及び要請についても、その正当性を考察すべき点が今後の課題として残されている。

(5) 参考文献

Bar-Hillel, Y., and Carnap, R. 1953, 'An Outline of a Theory of Semantic

Information' repr. in Bar-Hillel (1964), 221-274.

Dretske, F. I. 1981, *Knowledge and the Flow of Information* (Oxford: Blackwell).

Floridi, L. 2005, 'Is Semantic Information Meaningful Data?' Philosophy and Phenomenological Research, 70(2), 351-370.

Floridi, L. 2010, *Information: A Very Short Introduction* (Oxford: Oxford University Press).

Floridi, L. 2011, *The Philosophy of Information* (Oxford: Oxford University Press).

---

1 「この意味」とは、直前の文章である以下を指している。「出来事の状態は、それにあたることで適切な観測者が X について学ぶことができるちょうどその程度において X についての情報を含む。(中略) これを『情報』という用語の核となる意味として言及しようと思う。」

2 (Floridi 2011, p.83-84)

3 情報管理、データマイニング、データベース設計や意思決定論などの分野が挙げられている (Floridi 2011, P.83)

4 フロリディ (2005) では同じものが SDI (Standard Definition of Information) と呼ばれている。

5 ここで言う well-formed とは、必ずしも文法的なものでなくてもよい。たとえば、電気の回線が物理的にうまく繋がっていても well-formed である。これにより、命題以外のものでも意味論的情報として認められる。

6 ここでは簡単のため、GDI の代わりに命題を用いて議論が進められるが、議論の進め方と結論的には本質的な影響はない。

7 「情報量」という言葉における「情報」は、ある事象の確率によって決まる数学的な量のことであり、フロリディが真理性をもたせようとしている意味論的情報とは異なる。D 上の変数についても情報量を計算できるのはそのためである。

8 原文 (Floridi 2011, p.98) では S 上の変数 x を用いて P(x)と記載されているが、確率は D 上の変数についても計算できる (実際にフロリディも D 上の変数について確率を用いている) ので、P(φ)の誤植だと思われる。

9 この論理式は、IRP から導かれはするが IRP そのものではない。本議論ではこの論理式しか用いないためにこう記載していると考えられる。なお、IRP は 1－P(φ)で表現されるので、P(φ)=1 ならば情報量は 0 になる。

10 Bar-Hillel－Carnap Paradox：理想的な受信者が受け取ることのないような自己矛盾文は、最も大きな情報量を運ぶことになる。(Bar-Hillel and Carnap 1953, p.229)

11 「筆者の主張」で詳しく述べるが、[2] と [3] は本来であれば S 上の命題変数である x と y を用いて記述されるべきである。この誤りが議論をわかりにくくしている要因の一つであると筆者は考える。

12 この図では便宜上互いに排反であるかのように記載しているが、②合わさると矛盾する事実式と③偽の事実式④真の事実式には重複があり、実際には排反なわけではないことに注意。

[13] 自己情報量は $I(\varphi)=-\log P(\varphi)$ で計算される。
[14] （Floridi 2011, p.123-127）参照。対象となる命題が、言及しようとしている現実 w からどの程度離れているかの度合いに基づいて計算される。

（関西大学）

哲学的治療において必要なもの
What is needed in philosophical therapy

（溝越 大秦）

Abstract

The primary question that lies here is "What is needed in philosophical therapy?" In L. Wittgenstein's "Philosophical Investigations", philosophical problems are regarded as an illness to want to look at things in particular way. Wittgenstein compared philosophy to therapies of such an illness. Such an illness means being captured by a certain type of thinking. Because philosophical problems occur at a language level, philosophical therapy has to focus on the connection between words and activities with this to emancipate people from such a type of thinking. A Language- game achieves this by functioning as a model to evince the connection.

(1) 研究テーマ

本論文のテーマは、ルートヴィヒ・ウィトゲンシュタイン（以下 L.W.）『哲学探究』（以下 PI）における哲学的治療である。哲学的治療において何が必要なのか、が本論文の問題意識である。

(2) 研究の背景・先行研究

哲学においては、古来より様々な問題が扱われている。その問題とは、存在とは何か、我々はなぜものを見られているのか、科学はこの世界をどの程度解明できるか、我々はどのように生きるべきか（これは倫理学の問いとも取れる）などである。これらの問題についてあらゆる哲学者たちが論争を繰り広げてきた。しかし、どの問題も明確な回答を与えるにはあまりにも困難であろう。

L.W.はこのような困難に対して次のような態度で接するよう語る。

> 哲学におけるあなたの目的は何か。-ハエにハエとり壺の出口を示してやること。(PI §309)

困難に喘いでいる、つまり哲学的問題について右往左往している人に対してL.W.は哲学的問題を脱し、解決するための出口を教えることに哲学の目的を見出す。彼は哲学者が取る態度を次のようにも例える。

哲学者は、病気をとりあつかうように、問いをとりあつかう。(PI §255)
哲学の方法が一つしかない、というようなことはない、実にさまざまな方法があり、いわば異なった治療法があるのである。(PI §133)

つまり、L.W.にとって哲学はいわばさまざまな病気をさまざまな手段で治療するように、さまざまな哲学的問題をさまざまな手段で解決するものなのである。ではどうすれば哲学的問題が解決されるのであろうか。彼は以下のように述べる。

これらの問題［ =哲学的問題］[1]は、［ …］ われわれの言語の働きを洞察することによって解決される［ …］これらの問題が解決されるのは、［ …］とうに知られていることを整頓することによってである。(PI §109)
もしわれわれが言語という現象を、原初的なその適用法にそくして研究し、その適用例において語の目的とはたらきを明瞭に見渡すことができるのであれば、そうした煙霧［ =哲学的問題］は霧散する。(PI §5)

哲学的問題は言語についての誤解から発生し、言語的なレベルで、既に知られていることを洞察ことによって解決される。L.W.は語の原初的な適用例における語の使用を洞察するために、以下のような言語を想定する。

一つの言語を考えてみよう。その言語は、建築家 A とその助手 B との間の意思疎通に役立つのでなくてはならない。A は石材によって建築を行なう。石材には台石、柱石、石版、梁石がある。B は A に石材を渡さねばならないが、その順番は A がそれらを必要とする順番である。これらの目的のために、二人は「台石」「柱石」「石板」「梁石」という語からなる一つの言語を使用する。A はこれらの語を叫ぶ。-B は、それらの叫びに応じて、もっていくよう教えられた通りの石材を、もっていく。-これを完全に原初的な言語と考えよ。(PI §2)

関口(1996)はウィトゲンシュタインが PI§2 で語の使用に目を向けさせるために、「大工とその助手という二人だけの共同体における、[上の引用にある]たった四つの語からなる(完結した)言語の設定という、単純化と省略と誇張を施した記述」(関口(1996), p.204)を導入したと述べる。

それでは、哲学的治療は哲学的問題に直面している者にどのようなものを振り返させなければならないのであろうか。ならびに、哲学的治療において必要なものは何か。本論文は L.W.が想定する哲学的問題の特徴と、それに対するいわば治療のアプローチについて、本論文は Kuusela (2008)と Morris (2007)の議論に注目する。

Kuusela(2008)は、一定の像の観点で事物の捉えようとする意欲(will)が哲学的問題の特徴であるとし(Kuusela, 2008, p.45)、哲学的治療は当人を「言語（あるいは思考）とは何か比類なきものである(PI §110)」というような問題のあるものの見方から解放させるものに位置付ける。その解放のために哲学的治療は一つの語における異なる使用、あるいは一つの使用における異なる語を導入することを目指す。

Kuusela(2008)は哲学的治療が言語使用の個別的な例の整頓(arranging)によって、一定のものの見方に捉えられている者に哲学的問題を解消するための新しい見方を明瞭に提示する整理(ordering)がなされることによって達成されると述べる [2]。(Ibid., p.45, Ibid., p.89) 整頓、そして整理によって事物の多様な面が現れ、単一の概念下における多様な事例を見ることができる。(Ibid., p.234, 235) これにより普遍的な定義や共通した特徴からその概念下の事物を捉える「特殊な」ものの見方を拒絶し、語の使用を正確に理解することができる。

Morris(2007)によれば、哲学的治療には主に二つの方法がある。一つは文法的記述の訂正及び正しい文法的記述である。文法記述の訂正と正しい文法的記述について彼女は、”This” や”That”が真の固有名だというラッセルの主張に対して、”This is N”と人は言うが、”This is ’This’”と言うことはない、という L.W.の文法的記述を例に挙げる(Ibid., p.75)。これにより、哲学的偏見により得られた平静が打破され、ラッセル（あるいはラッセルを支持するもの）に哲学的偏見を検証させると彼女は指摘する(Ibid., p.78)。

もう一つはものの見方の代替案を提供することである。彼女は例として、PI §67 の引用から、「何か共通するものがなければならない」という考えに縛られている者に対して一本の繊維が一本の糸に通らずとも、多くの繊維が糸の強さをもたらすというアナロジーを挙げ(Ibid., p.75)、L.W.が当人に「代替案の可能性を認めるよう説得する(Ibid., p.80)」のであると主張する。

両者の解釈を踏まえると、哲学的治療は言語を普遍的な理論や共通の特徴でとらえることを拒否し、治療を受ける者の言語に対する誤解の種類、いわば「症状」に応じて多様な個別的事例に目を向けさせ、思考の傾向を矯正し、語の使用を正確に理解させることによって成立するといえよう。

例えば、「ゲーム」という語を理解するために L.W.は PI §66 でカードゲームやチェスを挙げ、勝ち負けのあるものと言ったあと、球戯の例を挙げ、「子どもがボールを壁に投げつけて再び受けとめている場合」(PI §66)には勝ち負けという特性が消滅することに注目している。そこで技術という語でさえ、チェスとテニスでは全く違う様相を見せることも挙げる。

PI §66 は「ゲーム」という語の使用の個別的事例を整理し、「ゲーム」の「アスペクトの変化」(関口(1996))を対話者に起こさせているのであり、「ゲーム」に対する特殊な見方および偏見を取り除いているのである。

ここで本論文は一つ疑問を投じたい。哲学的治療で個別的事例によって「語の使用」に目を向けさせると述べたが、より詳細にこれらを把握すべきではないか。語の使用は語と事物の単なるつながりではなく、使用者の要素に強く関わるものであろう。したがって、言語使用の個別的事例を提示しても、使用者のどの要素に目を向けるかが明確でなければ、治療を受ける者に混乱をもたらすであろう。哲学的治療で、我々はどこに目を向けさせるべきか。

それは語に伴う行為である。L.W.は、「子どもがボールを壁に投げつけて再び受けとめている場合」(PI §66)やチェスとテニスの技術の違いなど、行為を見て、「ゲーム」という語に関する考察を展開している。L.W.は言語および語の使用において語に伴う行為を重要視しているのではないか。Kuusela(2008)、Moorris(2007)、関口 (1996)は取り立てて論じることはないが、本論文は語の使用を理解するためには語に伴う行為が重要なキーワードであることを提唱したい。

(3) 筆者の主張

(3)では、語に伴う行為が語の使用の成立にどのように寄与しているかを、PI で述べられる言語の習得過程に注目して論じる。そして、再び PI §66 の例を考察し、語に伴う行為に注目させることに必要なものを提示する。最後に、哲学的治療の成立に必要なものを挙げる。

PI §2 における語の使用は、たとえば、ただ「台石」「柱石」「石板」「梁石」と突然発声しただけで見出されることはない。これら 4 つの語の使用は A の叫びを受けた B が台石、柱石、石板、梁石を A の元に持ってくことにより見出されるのである。

PI §2 の言語の習得過程について L.W.は以下のように記す。

> 言語のそのような［ =PI §2 に挙げられる］ 原初的諸形態を、子ども[3]は、話すことを学ぶときに用いる。その場合、言語を教えるということ

は、それを説明することではなくて、訓練するということなのである。(PI §5)

子どもたちは、事前の説明によってではなく、「そのような活動を行い、その際そのような語を用い、そのようにして他人の言葉に反応するよう、教育される(PI §6)」。つまり当該の言語ゲームにおける他の参加者による言葉・活動 への反応をとおして語の使用が教えられ、語の用い方が訓練される。例えば、PI §2 で記述されている言語ゲームにおいて「台石」と A が叫んだ時、それを聞いて持っていく時、間違って柱石を持って行った時に、元の位置に戻り、ふたたび A の所に石材を持っていくなどの活動を通して、B は「台石」という言葉を聞き石板でも柱石でも梁石でもない台石を A に持っていくことを学ぶことが挙げられる。

さらに、語の使用を学ぶ際、「直示的教示」が重要であると L.W.は以下のように指摘する。

訓練ということの一つの重要な部分は、教える者が諸対象を指さして、子どもの注意をそれらのものへ向け、それとともに何か語を発すること、例えば「石板」という語を、そのような形をしたものを提示する際に発音することから成り立つであろう。([…]わたくしはこれを「語の直示的教示」と呼びたい。[…]) (PI §6)

直示的教示によって、子どもには語と提示されたものが表象のような形で対応するといえるかもしれない。ただし、語の使用を明瞭に理解する上で、事物との対応関係が語に伴う行為よりも重要であるわけではない。L.W.は以下のように答える。

たとえ直示的教示が表象を喚びおこすのだとしても、[…] しかし、それは一定の教育を伴なって (原文ママ)、はじめて可能になるのである。異なった教育が行なわれれば (原文ママ)、これらの語の直示的教示が同じであっても、まったく異なった理解が生じるであろう。(PI §6)

直示的教示を全く同じ方法でおこなったとしても、A が叫んだ語に B が応えて、直示的教示によって提示されたものに「×」というような印をつけろという教育や訓練が行われれば、PI §2 で挙げられる言語ゲームとは全く違う言語ゲームになり、B による語の使用の理解も全くもって違うものに

なる、という言語観なのである。

したがって、L.W.の言語観において子どもが語の使用を理解するためには語と事物の対応関係よりも、ある言語内で語によって引き起こされる行為が重要であると結論づけられる。ところが、この言語観は PI §2 の原初的な言語に適用されるものである。我々はこの言語観を、哲学的治療で扱われる、高度な概念的思考が要求されるような言語形態に拡張できるのか。

このような問いに対して、高度な教示を要求されるからこそ、語の使用を理解する際に、ある言語内で語によって引き起こされる行為が重要であると本論文は主張する。語の使用の習得について取り上げている Luntley(2013)の議論を援用しながら以下の段落でそれを明らかにする。

ある言語内の他の参加者による言葉・活動への反応をとおして語の使用が教育され、訓練される過程を Luntley(2013) は stimulus-response(SR) training（以下、SR 訓練）と呼ぶ(Luntley, 2013, p.45)。SR 訓練は以下のようにはたらくと指摘される。SR 訓練 は 因果 的に決定された 反応 のパターンや S(stimulus) 下で R(response)を返す傾向を確立する。(Luntley, 2013, p.45) 例えば、PI §2 において B は A が「台石」と叫ぶこと(S)を受けて石を持って行くこと(R)を通して「台石」のはたらきを学ぶのである。

ただし、言語を学ぶ手段は SR 訓練だけではない。Luntley(2013)は、PI §6-10 では、より洗練された訓練の考え方が扱われると指摘する。 SR 訓練では、訓練を受ける者に対して、刺激を受けて規則的に反応することが要求されるのみであるが、PI §6-10 の訓練では注意されている対象へ能動的に反応し、行為を選択する能力が要求される。(Ibid., p.45)

以上から、より高度な教示を必要とする言語であればあるほど、教示を受ける側により多くの行為を選択せねばならないことが分かる。すなわち高度な言語形態を理解するケースだからこそ、ある言語内で語によって引き起こされる行為が重要になる。したがって、哲学的治療を行う際、治療を施すものはある言語内で語によって引き起こされる行為に目を向けさせることが必要なのである。

前段までの議論より、問題を起こしている語それぞれについて、行為を伴う訓練を行うことが哲学的治療の望ましい形と言える。しかし、我々には体力、時間、効率など様々な問題が発生する。そこで、再び、PI §66 で L.W.が「ゲーム」と名のつく様々な具体例を挙げ、「考えるな、見よ!」と語る例について考察する。

PI §66 に挙げられるような、「ゲーム」という語に伴う行為を「見る」ことはもはや語に伴う行為の擬似的な訓練である。擬似的な訓練は、「子どもが

ボールを壁に投げつけて再び受けとめている場合」(PI §66)を想定し、「ゲーム」という言葉に能動的に行為を選択し、使用を理解する過程である。

それを実現する道具としても記述された言語ゲームが用いられる。言語ゲームは「言語と言語の織り込まれた諸活動との総体(PI §7)」であるため、記述された言語ゲームというモデルを通した語に伴う行為に関する擬似的な訓練によって哲学的問題を抱える者は語の使用を明瞭にすることができる。語の使用が明瞭になった者は哲学的問題から解放されるであろう。以上のことより、記述された言語ゲームを道具として用いた語に伴う行為の疑似的な訓練哲学的治療において必要なものなのである。

(4) 今後の展望

哲学的治療を受けた者は、は言語への誤解を捨てて、正しい理解へと向かう。言語への誤解には言語に対する過信のようなものも含まれるであろう。言語に対する過信とは、本来言語で表現できないものも言語で表現することができるという種類の誤解である。

治療を受けた者が、当該の哲学的問題を言語に対する過信であるとして放棄した時、その治療はあくまで当人にとってのみ、あるいは同じ言語（英語や日本語、ドイツ語など）を話す者にとってのみ、もしくは同じ文化・慣習を共有するものにとってのみ適用できるのか。それとも、あらゆる人間に適用できるのか。

この問題は、前期 L.W.哲学において重要な「語りえない」という語について拡張することができる。本論文の哲学的治療の立場に則って言う「語りえない」とはすなわち、本来言語で表現できないにも関わらず、言語に対する過信から生まれた問いに対する形容である。『論理哲学論考』（以下 TLP）のある種数学的な言語観は、言語を普遍的な定義や何か共通した特徴で捉えていたものであった。TLP も PI と同様哲学的問題の放棄に注力したものであるが、言語に普遍性を求めた時点で「語りえない」という語の使用に哲学的問題を招く独断が残されているかもしれない。

個別的事例を重んじる PI の言語観では、言語に対する過信がどのように取り除かれていくかを TLP より詳細に説明することができよう。すなわち誰に、どの文化圏の言語使用者に、そして人類全体にどのような治療が可能なのであろうか。あるいは「文化」や「人類全体」というようなカテゴライズは無用であり、個々人に対する治療のみが個々人に応じて成立するものなのだろうか。

上記の問いに答える際、「語りえない」として哲学的問題を放棄する過程に

TLP とは違った過程を見出す可能性があろう。この問題は TLP と PI それぞれをより正確に理解すること及び両者のつながりをより明確にすることにつながると言える。

（大阪大学）

注

1　引用での［］内はすべて筆者による補足。

2　“ordering”は哲学的問題に対して新しくものの見方を「秩序づける」という意味を強調するために、物の順序を正す「整理」の訳語を、”arranging”は言語の個別的事例を用意し、偏ったものの見方をする相手に対して見せるあるいは考えさせるために事物の配置を整えるところに注意して「整頓」の訳語を適用する。

3　邦訳は「子供」であるが、筆者により「子ども」と修正している。

(5) 参考文献

・Ludwig Wittgenstein, 2010, Philosophical Investigations, translated by G. E. M. Anscombe, P. M. S. Hacker and Joachim Schulte, revised fourth edition by P. M. S. Hacker and Joachim Schulte, Wiley-Blackwell
(『哲学探究』、藤本隆志訳、1976、『ウィトゲンシュタイン全集 8 』、大修館書店)

・Ludwig Wittgenstein, 2009, Logische-Philosophische Abhandlung Tractatus logico-philosophicus, Suhrkamp Verlag Gmbh
(『論理哲学論考』、奥雅弘訳、1975、『ウィトゲンシュタイン全集 1』、大修館書店)

・Oskari Kuusela, 2008, The Struggle against Dogmatism: Wittgenstein and the Concept of Philosophy, Harvard University Press

・Katherine Morris, 2007, “Wittgenstein’s Method: Ridding People of Philosophical Prejudices”, Guy Kahane, Edward Kanterian, Oskari Kuusela (Editor), Wittgenstein and His Interpreters: Essays in Memory of Gordon Baker, Wiley-Blackwell; 1 edition

・Michael Luntley, 2013, ”What’s doing? Activity, naming and Wittgenstein’s response to Augustine”, Arif Ahmed (Editor), Wittgenstein’s Philosophical Investigations: A Critical Guide (Cambridge Critical Guides), Cambridge University Press

・関口浩喜, 1996, 「展望とアスペクト」,日本哲学会,『哲学』第 47 号, 256-265

新進研究者 Research Note
科学者はどのように非認識的価値を考慮するべきか
How should a scientist consider non-epistemic values?

伊沢　亘洋

Abstract

Recently, science has played an important role in policy-making, and in this situation, scientist often must make non-epistemic value judgements. To solve this value-free problem, in 2009, Heather Douglas suggested a new value-free ideal. This new ideal depends on a new distinction between a direct role and an indirect role in value, instead of the classic distinction between epistemic values and non-epistemic values. But, this new distinction has a big problem, and this argument seems no longer successful. Thus, in this paper, I will turn to case study of PM 2.5 in Japan to revise Douglas' ideal.

(1) 研究テーマ

リスクアセスメントのように、科学者が倫理的価値を考慮しなければならないような状況において、価値中立性をどのように考えるべきか。

(2) 研究の背景・先行研究

科学哲学において、伝統的に、科学における「価値中立性の理想」が議論されてきた。例えば、Lacey は、認識的価値[i]のみを考慮して理論選択が行われることが科学の価値中立性であると主張している(Lacey 1999)。

以上のような認識的/非認識的価値の区別に訴える伝統的価値中立性の立場に対して、近年、積極的に非認識的価値（倫理的価値、社会的価値など）を考慮するべきだとする議論が盛んになっている。この流れは、科学が社会的意思決定の中で重要な役割を果たすようになったことが背景にある。このような議論においては、非認識的価値を科学内での判断において、考慮しつつ、価値中立性を維持するということが中心的な問題となっている。本論ではそれらの議論の代表的論者として Douglas を紹介する。

Douglas の新しい価値中立性

Douglas は価値中立性という理想は、政策決定過程における科学の役割を考えれば拒否しなければならないと述べる一方で、「科学の高潔さ(the integrity of science)」を守るために非認識的役割は適切な役割に制限されなければならないと述べる（Douglas 2009, p. 1)。

Douglas は価値を認識的規準、認知的価値、倫理的価値、社会的価値に分類する(p. 92)。認知的価値は研究を豊穣にするような価値のことを指しており、豊穣さ、外的整合性、単純性、予測の精確性などが含まれている。一方、認識的規準は内的一貫性と経験的十分性の二つだけである。

また Douglas は価値の分類とは別に、価値の役割を提案する。Douglas は役割を二つに分ける。一つは直接的役割である。直接的役割において、価値はそれ自体で我々の決定を決め、我々の選択を動機付ける独立の理由として働く(p. 96)。一方、間接的役割において、価値は以下のような役割を果たす。

> 選択にとって利用可能な証拠が十分かどうか、その不確実性の重要性は何かを決定することを助け、誤った選択による潜在的帰結を評価し、そのような帰結が悲惨である

ときには、より多くの証拠を要求することによってこの帰結を和らげるのを助けるという重大な役割を果たす。(p. 97)

Douglas は、価値中立性の条件として、「分離的」という条件をあげている(ch. 6)。これは、特定の研究結果を望んだり、ほかの結果を恐れることが避けられているという条件である。Douglas によれば、価値の働きを間接的役割に制限することで、この「分離的」という条件を満たすことが可能であるとする。以上のような二つの役割を設定した上で、それらの役割が科学におけるそれぞれの領域において制限されるかどうかがさだめられている。

まず、探求する領域を決める際と、方法論を決める際には、直接的役割が許されている。例えばどのような領域を探求するべきかということを決定する際には社会に役立つことや、人命を多く救うことにかなった領域が探求されるべきだろう。また、方法論の選択の際にも人体実験などは倫理的観点から行うべきでない。これらの例においてはいずれも社会的価値、倫理的価値がそれ自体決定の理由として働いている。したがって価値が直接的役割を果たしている。

一方で、間接的役割は、方法論を決めるとき、データを集めるときに働く (pp. 104-105)。例えば有意水準の設定が挙げられている。ある物質の毒性について有意差検定を行ったとする。このとき検定によって帰無仮説（毒性がない）が棄却されなかったにもかかわらず実際には毒性があったような場合には、その結果にもとづいて安全ではない環境基準値が定められ、多くの人の健康を損なうことが帰結する。一方で、この逆の場合、つまり帰無仮説が棄却されたにもかかわらず、実際に毒性がなかった場合には、実際に必要のないほど過剰に安全な環境基準値を定めることとなり、健康を特に損なわないにもかかわらず無駄に企業がその物質を排出しない努力をすることが帰結する。これらの二つの帰結を評価する際に働くのが間接的役割である。またデータの特徴づけにおいても間接的役割は働く(pp. 105-106)。例えば毒性学において毒物によって動物に悪性の腫瘍ができるかを観察するときに、特定の組織を腫瘍とみなすかどうかに関して曖昧なケースがある(Douglas 2000)。このような曖昧なケースに関して、毒物を投与した投与群に含まれる動物の腫瘍を悪性とみなすことは毒性がないとする帰無仮説を棄却しやすくする。このような場合においても何を悪性の腫瘍とみなすかについての境界をどこに引くかという問題は、偽陽性と偽陰性による帰結を評価して決めるべき問題であり、間接的役割が働く[ii]。

間接的役割の解釈

Douglas の価値中立性にとって、間接的役割は非常に重要である。しかし、Elliott は Douglas の間接的役割は次の二つの解釈が可能であるとする(Elliott 2013)。

帰結的な直接的/間接的役割：ある主張を受け入れることで生じさせようとする意図された帰結にもとづいて、価値が科学者の選択に影響を与える時、価値は直接的役割をする。科学者が避けたいと望む誤りに関係している意図しない帰結に基づいて価値が科学者の選択に影響を与えるとき、価値は間接的役割をする。(p. 377)

論理的な直接的/間接的役割：価値が主張のための保証、証拠として扱われる時、直接的役割をする。証拠が主張を受け入れるために十分であるかについての判断に対して価値が影響するとき、価値は間接的役割をする。(p. 377)

間接的役割を正確に考えるために、仮説選択の例を考えてみよう。片方の仮説 A は毒性を低く推定し、もう片方の仮説 B は毒性を高く推定するとする。また、本当は毒性が低い場合を a、毒性が高い場合を b とする。このとき、仮説 A を選択して、実際に A が正しかった時(a)に起こる帰結 Aa、仮説 A を選択して、実際に A が正しくなかった時(b)に起こる帰結 Ab、仮説 B を選択して実際に B が正しかった時(b)の帰結 Bb、仮説 B を選択して実際に B が正しくなかった時(a)の帰結 Ba の四つの帰結が存在する。それぞれの帰結は例えば、次のようなものが挙げられるだろう。

	a（毒性低い）	b（毒性高い）
仮説 A（毒性低い）を選択	Aa:製薬会社の利益が上がる。 健康、環境を守ることができる。	Ab:製薬会社の利益が上がる。 健康、環境に対する被害がある。
仮説 B（毒性高い）を選択	Ba:製薬会社の利益が下がる。 本来不必要な健康、環境を守るコストがかかる。	Bb:製薬会社の利益が下がる。 健康、環境を守ることができる。

このとき、Elliott の帰結的解釈と論理的解釈はそれぞれ、間接的役割の異なった側面についての理解を与えている。まず帰結的解釈は上の表のうちのどの項を考慮するべきかについて述べている。つまり上の表のうち、Ab と Ba のみを考慮するべきであると言い換えられる。帰結的解釈では仮説が正しかったときの帰結を考慮せずに、仮説が誤った時に生じる帰結のみを考慮するという点で、価値の考慮の仕方に制限が加えられている。

一方で、論理的解釈は仮説を考慮した後で何を選択するかについて述べている。上の表の Aa、Ab、Ba、Bb を考慮したのち、仮説 A、仮説 B の選択ではなく、証拠の十分性の基準（例えば有意水準）を選択すると主張している。このとき、論理的解釈では、あらかじめ望ましい仮説を選択しておいて、それを選ぶために証拠の十分性基準を調整するようなケースは除外されている(Douglas 2016)。なぜなら、このようにすると結局、帰結がもつ価値によって仮説を選択することになり、価値が直接的役割として働くからである。このように、論理的解釈では、全ての帰結の持つ価値を考慮したのち、その考慮が特定の仮説を導くことになるかどうかについては不明の段階において証拠の十分性基準を選択するという点で、価値の考慮の仕方に制限が加えられている。

(3) 筆者の主張

Douglas の間接的役割に関して、前節で、論理的解釈と帰結的解釈があることを紹介した。本節では、PM2.5 の事例に当てはめて考えた上で、どちらの解釈も取りえないことを示す。

PM2.5 の事例

PM2.5 の基準は報告書内で最終的に 15μg/m^3 と提案されている。しかし一方で、閾値を見いだすことが難しいということも指摘されている。つまり 15μg/m^{3}以下でも考慮すべ

き程度のリスクが見られることを指摘している研究が存在しているのである。これらについてはどのように判断しているのだろうか。「微小粒子状物質環境基準専門委員会報告」によれば、このような研究はいくつかあるが、それらのうちカリフォルニア子供研究について以下のような理由で考慮しないことが述べられている。

> カリフォルニア子供研究では、対象 12 地域の全体的な傾向として、PM2.5 濃度と肺機能の成長の遅延に関連がみられているが、平均濃度が 10μg/m^3 を下回る複数の地域と、平均濃度が 15μg/m^3 付近の地域において肺機能の成長に差があると結論づけることは困難である。
>
> (「微小粒子状物質環境基準専門委員会報告」4.3.1.2)

しかし、この基準値が設定されるまでの議事録をみれば、次のようにカリフォルニア子供研究を考慮するべきではないかと意見されている場面もある。

> 資料２で、確かに年平均値の１５を提案した理由はわかるのですが、前回にも私言いましたが、１５よりもっと低いところに値があるのではないかということが最近話題になっていて、その根拠の多くの一つが肺機能の研究評価なんですね。この２２ページから見ましても、かなり低いところから影響が出ていて、しかもこの肺機能の研究も、いわゆるコホート調査でかなり信頼性が高いということで、年平均値を考慮するときに１５以下のところを考えた方がいいんじゃないかというような人も最近出てきていると思うのですが、これはここの結論では、やはり１５だというふうになっていますが、どうなんでしょうか。
>
> （微小粒子状物質環境基準専門委員会（第６回）議事録 香川委員の発言より引用）

Douglas は、どのデータ、どの実験を重要なものとみなすかに関しても、価値が間接的役割として働くと述べているので、このカリフォルニア子供研究を重要なものとみなすかどうかにおいても間接的役割が働くことになる(pp. 103-105)。

帰結的解釈の批判

	15μg/m^{3}で害がない	15μg/m^{3}で十分害がある
カリフォルニア子供研究を証拠としてみなさない →基準値 15μg/m^3	Aa:小さな排出制限。 健康を守ることができる。	Ab:小さな排出制限。 健康に害がある。
カリフォルニア子供研究を証拠としてみなす →基準値 12μg/m^3	Ba:大きな排出制限。 本来不必要な健康を守るコストがかかる。	Bb:大きな排出制限 健康を守ることができる。

カリフォルニア子供研究を証拠とみなすとすれば、基準値は 15μg/m^3 未満に設定され

ることになる。一方で、証拠とみなさないとすれば、基準値は 15μg/m^3 に設定されることになる。帰結的解釈に従えば、上の表で、Ab と Ba のみを考慮した上で、カリフォルニア子供研究を証拠とみなすかどうかを決定することになる。しかし、なぜこのとき、Aa、Bb を考慮してはいけないのだろうか。全ての項目の帰結を考慮した上で、一部の帰結を重視するという判断は許容されるが、あらかじめ、考慮すべき対象を一部の帰結のみに限定するという考えは、かなり奇妙な考え方である。おそらく、Douglas は、「意図した帰結」という言葉によって、科学者が所属する企業の利益など、個人的な利益のことのみを想定し、このことから Aa、Bb を考慮の対象から外したのだろうと考えられる。しかし実際には Aa、Bb にも「健康を守ることができる」や「排出制限を小さく抑えられる」など社会的意思決定をする上で重要な項目が含まれている。このことを踏まえれば帰結的解釈は不適切である。

論理的解釈の批判

証拠の十分性基準に対して価値を考慮するというのは、仮説選択の際の有意水準の例は非常にわかりやすいものであるが、他のケースでは理解しづらい。

カリフォルニア子供研究を重要な研究とみなすかどうかについての判断基準は、有意水準のように、あらかじめ基準を設定できるようなものではない。個々の研究を重要な研究としてみなすかどうかという判断は必ず、ケースバイケースの判断にならざるおえない。したがって、仮説選択における有意水準の例とは異なり、証拠の十分性基準を決定することと、重要な研究としてみなすかについての判断は分離できないのである。これはつまり、証拠の十分性基準を選択することに関して、帰結の価値を考慮することは、帰結を考慮して、重要な研究としてみなすかどうか決定するということになり、価値が直接的役割を果たすことになる。以上の理由から、Douglas の論理的解釈による間接的役割は支持することが難しい。

(4) 今後の展望

筆者はこのカリフォルニア子供研究を証拠とみなすかどうかという判断は、Douglas の間接的役割で処理することはできず、また、議事録のように、何も価値を考慮せずに一部の科学者の判断に任せることも望ましくないと考える。このような場合にはむしろ、カリフォルニア子供研究が示すような、15μg/m^3 以下の基準値の可能性も考慮した上で、その基準値を採用したときの帰結を考慮して決定するべきだと主張したい。実際に、この判断は帰結を考慮して選択するため、Douglas の枠組みでいえば、価値が直接的役割を果たすことになるものの、Douglas がいうように証拠を無視しているわけではない。帰結を考慮するべき仮説はある程度データによって支持されるという制約を満たしたものであるからである。また、価値が直接的役割として働くからといって、必ずしも価値中立的ではないということにはならない。価値中立性は価値を考慮しないということだけではなく、あらゆる価値観を持つ人による批判にさらされることによっても得られるからである（シュレーダー=フレチェット 2007）。リスクアセスメントはこの意味での価値中立性を目指すべきである。

---

[i] ここで認識的価値とは具体的には、経験的十分性、広範さ、内的整合性、外的整合性、単純性などの性質を指す。Lacey は科学において「信念の認識的価値の評価によって、ある信念が正しいかどうか判断する」と述べており、この意味で認識的価値は科学にとって重要な性質である(Lacey 1999, p. 45)。

[ii] データの分類といっても、必ず証拠の十分性について考慮されるわけではない。明らかに社会的価値判断のみによって基準が定められるケースがある。このような例として骨

粗しょう症の事例が Biddle(2016)によって挙げられている。

(5) 参考文献

Biddle, J. B. (2016). Inductive Risk, Epistemic Risk, and Overdiagnosis of Disease. *Perspectives on Science* 24 (2):192-205.

Douglas, H. (2000). Inductive Risk and Values in Science. *Philosophy of Science,* vol. 67, no. 4, pp. 559-579

—(2009). *Science, Policy, and the Value-Free Ideal.* University of Pittsburgh Press.

—(2016). Values in Science. In Humphreys, P. (Eds) *The Oxford Handbook of Philosophy of Science.* Oxford University Press, pp. 609-630.

Elliott, K. C. (2013). Douglas on Values: From Indirect Roles to Multiple Goals. *Studies in History and Philosophy of Science* Part A, vol. 44, no. 3, pp. 375-383.

Lacey, H. (1999). *Is Science Value Free?: Values and Scientific Understanding*. Routledge.

Shrader-Frechette, K. (1991). *Risk and Rationality*. University of California Press.（松田毅訳 [2007]『環境リスクと合理的意思決定: 市民参加の哲学』 昭和堂）

微小粒子状物質環境基準専門委員会（第 6 回）議事録 http://www.env.go.jp/council/former2013/07air/y078-06a.html（アクセス日 2019/12/9）

微小粒子状物質環境基準専門委員会「微小粒子状物質環境基準専門委員会報告」2/3 https://www.env.go.jp/council/toshin/t07-h2102/01-2.pdf(アクセス日 2019/12/9)

所属大学：京都大学

なぜ古典論理に計算的意味を与えるのか
—SLC にもとづく Dummett 的観点の拡張の試み—
Why giving computational interpretation to classical logic?
-an attempt at extending Dummett's conception of meaning based on SLC-

角田健太郎

Abstract

Recently, computational interpretations of propositional logical calculi have been extended from intuitionistic logic to classical logic. That seems to prompt us to rethink Dummett's notion of meaning and make it more generous so that not only intuitionistic logic but also (at least the elementary parts of) classical logic could be made sense of. For this purpose, it would be indispensable to introduce one new technical notion from the semantics of programming languages: the notion of *continuation*, which designates, roughly speaking, 'that part of the computational process that has not been executed yet'. I conclude this note by firstly surveying the computational system named SLC (*symmetric lambda calculus*) developed by Filinski and Asai et al., in which the notion of continuation is fully made use of, and secondly proposing several important improvements that resulted from my own research which is still under development.

1 研究テーマ

論理学とはすなわち妥当な推論形式の研究であるという見方は、哲学者たちの間に広く共有されている。つまり論理学とは、モデル等の道具立てによって一定の妥当性の基準を据えて、どのような推論形式が妥当であるのかを述べる研究であるという見方である。実際それは間違いではないであろう。しかし近年ますます進展しているのは、そうした推論形式の妥当性を述べることに留まらないような、より詳細な、論理的推論の各ステップに含まれている計算的な推論的原理の解明である。例えば BHK 解釈は、実効的演算や幾何学的作図などに典型的に含まれる構成 (construction) の概念—これはすなわち、素材となるものの構造を常に保存する形で新たな対象を作り出すような対象形成手続きを特徴づける概念である—によって直観主義論理の各推論規則を特徴づけることができ、そして直観主義論理の証明プロセスにより導出された論理式は「その論理式が諸仮定からどのような実効的手続きを経て導出されたか」という証明の構造をその表現内容とすると解釈できることを明らかにした [7]。また、この BHK 解釈のひとつの洗練と見ることができる Curry-Howard 対応は—以下で詳しく説明するように—直観主義論理の証明プロセスが λ 式と呼ばれるコンピュータプログラムの抽象的表現の構築プロセスと対応付けられることを明らかにした [1][6]。

そうした近年の論理学の展開は、哲学にどういうことを教えるだろうか。すでにこうした情勢と結びついている哲学のひとつとして、Dummett の直観主義哲学がある。彼は 1973 年の論文 [2][12] で、命題の意味内容はその命題の「使い方・使用 (use)」に訴えて説明されるべきであり、そして BHK 解釈は直観主義論理に対して「使い方」による意味説明を与えていると論じた。ここで彼は直観主義論理による証明こそが命題の意味の源泉であると主張し、他方で古典論理によって証明された命題は正当化を欠いた無意味な命題であるとする過激な主張を行った。というのも、任意の式 $\mathfrak{A}$ について $\mathfrak{A} \vee \neg\mathfrak{A}$ が無仮定で成り立つ古典論理では、どの選言文も選言肢の少なくとも一方が実効的手続きを経た上で導出されるという BHK 解釈の条件が崩れている—つまり古典論理は「使い方」による説明を超えている—と彼は考えたからである。

Dummett は、BHK 解釈の観点が論理結合子延いては言語の意味の問題でもあることを深く明らかにした点で参考になる。しかし、すでに様々な応用が行われている数学の定理を導出した古典論理を一切無意味として切り捨ててしまうのは奇異だと思われる。直観主義論理の証明プロセスが含む実効的手続きを経る導出の概念には帰着できずとも、しかし依然としてある種の合理化と言えるような部分が古典論理の証明プロセスには含まれており、Dummett はこれを見過ごしてしまっているのではなかろうか。本研究は、古典論理に特有なこの合理化概念を究明し古典論理的命題の意味を与え、そして Dummett の哲学の (忠実ではないかもしれないが) ひとつの健全な発展の方向として古典論理ベースの哲学を展開することをテーマとしている。

2　研究の背景・先行研究

Dummett が指摘するような古典論理のいわゆる非構成性を引き受けた上で、しかしながら近年の古典論理研究の進展の中では、古典論理の証明プロセスは導出される論理式の内容に対してある本質的な関わりを持っているということが明らかとなってきた。参考となるのは Curry-Howard 対応の古典論理への拡張の研究である。この研究は、古典論理の証明プロセスが、計算の順序を制御するある特別な演算子を含むようなプログラムの構築プロセスと対応付けられることを明らかにした。以下ではこのことを順を追って説明する。なお以下では紙幅の節約のため直観主義論理を IL と呼び、古典論理を CL と呼ぶ。

出発点となるのは、Gentzen の 1935 年の論文 [4] である。Gentzen はここで IL の体系として直観主義自然演繹 NJ を、CL の体系として古典自然演繹 NK を与えた。Gentzen の体系では、ある論理式が証明可能であるとき、かつその時に限って、推論プロセスの 1 ステップを表現する図式を有限木構造状に連鎖させた証明図と呼ばれる図式が構成可能である。例えば IL で証明可能ではな

く、かつ CL で証明可能なパースの法則と呼ばれる論理式 $((\mathfrak{A} \supset \mathfrak{B}) \supset \mathfrak{A}) \supset \mathfrak{A}$ をとろう。パースの法則を導出する NK-証明図は例えば次の形をしている。[1]

$$
\dfrac{\dfrac{\dfrac{}{\mathfrak{A} \supset \bot \vdash \mathfrak{A} \supset \bot}\,\text{公理} \qquad \dfrac{\dfrac{\dfrac{}{\mathfrak{A} \supset \bot \vdash \mathfrak{A} \supset \bot}\,\text{公理} \quad \dfrac{}{\mathfrak{A} \vdash \mathfrak{A}}\,\text{公理}}{\mathfrak{A} \supset \bot, \mathfrak{A} \vdash \bot}\,\supset\text{除去則}}{\mathfrak{A} \supset \bot, \mathfrak{A} \vdash \mathfrak{B}}\,\text{爆発則}}{\mathfrak{A} \supset \bot \vdash \mathfrak{A} \supset \mathfrak{B}}\,\supset\text{導入則} \qquad \dfrac{\dfrac{}{(\mathfrak{A} \supset \mathfrak{B}) \supset \mathfrak{A} \vdash (\mathfrak{A} \supset \mathfrak{B}) \supset \mathfrak{A}}\,\text{公理}}{(\mathfrak{A} \supset \mathfrak{B}) \supset \mathfrak{A}, \mathfrak{A} \supset \bot \vdash \mathfrak{A}}}{(\mathfrak{A} \supset \mathfrak{B}) \supset \mathfrak{A}, \mathfrak{A} \supset \bot \vdash \bot}\,\supset\text{除去則}}{\dfrac{(\mathfrak{A} \supset \mathfrak{B}) \supset \mathfrak{A} \vdash (\mathfrak{A} \supset \bot) \supset \bot}{(\mathfrak{A} \supset \mathfrak{B}) \supset \mathfrak{A} \vdash \mathfrak{A}}\,\text{二重否定除去則}}\,\supset\text{導入則}}{\vdash ((\mathfrak{A} \supset \mathfrak{B}) \supset \mathfrak{A}) \supset \mathfrak{A}}\,\supset\text{導入則}
$$

「$\mathfrak{A}_1, ..., \mathfrak{A}_n \vdash \mathfrak{B}$」(ただし $0 \leq n$) という表現はシーケントと呼ばれ、命題の集合「$\mathfrak{A}_1, ..., \mathfrak{A}_n$」を構成するすべての命題が仮定されたとき「$\mathfrak{B}$」という命題が結論されるという仮言的判断を意味する。各規則は、「水平線の上の判断が成り立つとき、水平線の横に表示された推論規則によって水平線の下の判断が成り立つ」という意味である。NJ と NK の違いは、推論規則に二重否定除去則を前者は含まない一方で後者は含む、という点のみにある。つまりパースの法則の NK-証明図中の二重否定除去則より上の証明図は NK-証明図でありかつ NJ-証明図である。ポイントは、シーケントたちが形成する木構造によって、証明図の一番下のシーケント (終式と呼ばれる) に対しては、それが表す判断へと至る証明プロセスが与えられるということである。

さて 1960 年代に明らかとなったのは、NJ と型付き λ 計算と呼ばれるプログラム構築体系との間に、ある同型性が与えられるという事実である [1][6]。この対応は Curry-Howard 対応 (以下 CH 対応) と呼ばれる。

まず λ 計算とは、計算機科学・情報科学の分野において計算機やプログラムの一般的性質に関する理論的研究を行うために、プログラムの構築とその実行という概念を記号列表現の形成とそれに対する変形操作という観点から抽象化した形式体系である。特に、各 λ 式に対して型の概念が与えられている λ 計算が型付き λ 計算と呼ばれる。型付き λ 計算の基本的な定義は次の通りである。変数記号として「$x, y, z, ...$」、束縛演算子として「λ」、型変数記号として「$A, B, C, ...$」、型演算子として「$\rightarrow$」、補助記号として「$), (, .$」を用いる。また以下で x は任意の変数記号ちょうど 1 文字を表し、$\mathcal{A}$ は任意の型変数記号ちょうど 1 文字を表す。

- λ 式の定義　$\mathsf{L}, \mathsf{M}, \mathsf{N} ::= \quad \mathsf{x} \quad | \quad \lambda\mathsf{x}.\mathsf{L} \quad | \quad \mathsf{MN}$
- 型の定義　$\alpha, \beta ::= \quad \mathcal{A} \quad | \quad \alpha \rightarrow \beta$
- 簡約規則の定義　$(\lambda\mathsf{x}.\mathsf{M})\mathsf{N} \rhd_\beta \mathsf{M}[\mathsf{N}/\mathsf{x}]$

  ある λ 式中の「$\rhd_\beta$」の左辺の形をした部分式を可約部と呼ぶ。また左辺の

形の λ 式を変形して右辺の形の λ 式を得ることを β 変換と呼ぶ。$\mathsf{M}[\mathsf{N}/\mathsf{x}]$ という表現は M 中の自由な x をすべて N で置き換えた表現を表す。

- λ 式型付け規則の定義

$$\frac{}{\mathsf{x}:\tau,\Gamma\vdash\mathsf{x}:\tau}\,id \qquad \frac{\Gamma\vdash\mathsf{L}:\tau}{\mathsf{x}:\alpha,\Gamma\vdash\mathsf{L}:\tau}\,wk \qquad \frac{\mathsf{y}:\alpha,\mathsf{x}:\alpha,\Gamma\vdash\mathsf{L}:\tau}{\mathsf{x}:\alpha,\Gamma\vdash\mathsf{L}[\mathsf{x}/\mathsf{y}]:\tau}\,ctr$$

$$\frac{\mathsf{x}:\alpha,\Gamma\vdash\mathsf{L}:\beta}{\Gamma\vdash\lambda\mathsf{x}.\mathsf{L}:\alpha\to\beta}\,FUN \qquad \frac{\Gamma\vdash\mathsf{M}:\alpha\to\beta \qquad \Delta\vdash\mathsf{N}:\alpha}{\Gamma,\Delta\vdash\mathsf{MN}:\beta}\,APP$$

「$\mathsf{x}_1:\alpha_1,...,\mathsf{x}_n:\alpha_n\vdash\mathsf{M}:\tau$」(ただし $0\leq n$) という表現は、「変数 $\mathsf{x}_1,...,\mathsf{x}_n$ の型がそれぞれ $\alpha_1,...,\alpha_n$ であるとき、λ 式 M の型は τ である」という型判断を意味する。各規則は、「水平線の上の型判断が成り立つとき、水平線の下の型判断が水平線の横に表示された λ 式型付け規則によって成り立つ」という意味である。また大文字のギリシア文字は $\mathsf{x}_1:\alpha_1,...,\mathsf{x}_n:\alpha_n$(ただし $0\leq n$) を表す。

ポイントは、定義により $\lambda\mathsf{x}.\mathsf{M}:\alpha\to\beta$ という λ 式は「型 α の λ 式が右から入力として結合したら (すなわち $\mathsf{N}:\alpha$ として、$(\lambda\mathsf{x}.\mathsf{M})\mathsf{N}$ という λ 式が形成されたら) 型 β の λ 式 $\mathsf{M}[\mathsf{N}/\mathsf{x}]$ を出力として返すような関数」を表現しているという点である。つまり、λ 式型付け規則によって型を与えられる λ 式は、まさにその型が表す仕様通りの振る舞いをする関数となっているということである。具体例をとろう。自然数および自然数上の後者関数を定め、後者関数の計算を表現することを考えよう。自然数を定める一つのやり方は、型 nat を $nat ::= (A\to A)\to(A\to A)$ で定め、さらに型 nat を持つ λ 式を、$y:A\to A$、$x:A$ として $0 ::\equiv \lambda y.\lambda x.x$、$1 ::\equiv \lambda y.\lambda x.yx$、$2 ::\equiv \lambda y.\lambda x.y(yx)$、...、$n ::\equiv \lambda y.\lambda x.y(\cdots(y(yx))\cdots)$ と定めるというものである [2]。さらにこれを踏まえ後者関数 suc を、$z:nat$、$y:A\to A$、$x:A$ として $suc ::\equiv \lambda z.\lambda y.\lambda x.y((zy)x)$ で定めれば、型付き λ 計算で後者関数の計算を実現できる。例えば $suc0$ から 1 を得る計算は、$suc0\equiv(\lambda z.\lambda y.\lambda x.y((zy)x))\lambda y.\lambda x.x\rhd_\beta\lambda y.\lambda x.y(((\lambda y.\lambda x.x)y)x)\rhd_\beta\lambda y.\lambda x.y((\lambda x.x)x)\rhd_\beta\lambda y.\lambda x.yx\equiv 1$ のようになる。重要な点は、nat を上のように定義したときに、後者関数 suc に当たる λ 式が意図した通り nat 上の 1 変数関数となっていることが次の λ 式型付け規則の連鎖的表現 (型付けの図式と呼ばれる) によって保証されるということである。

$$\begin{array}{c}
\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{y:A\to A\vdash y:A\to A}\,id \quad \dfrac{\dfrac{\dfrac{}{z:nat\vdash z:nat}\,id \quad \dfrac{}{y:A\to A\vdash y:A\to A}\,id}{z:nat,y:a\vdash zy:A\to A}\,APP \quad \dfrac{}{x:A\vdash x:A}\,id}{z:nat,y:a,x:A\vdash(zy)x:A}\,APP}{z:nat,y:a,x:A\vdash y((zy)x):A}\,APP}{z:nat,y:a\vdash\lambda x.y((zy)x):A\to A}\,FUN}{z:nat\vdash\lambda y.\lambda x.y((zy)x):nat}\,FUN}{\vdash\lambda z.\lambda y.\lambda x.y((zy)x):nat\to nat}\,FUN
\end{array}$$

型付けの図式は NJ-証明図とそっくりである。特に FUN と APP の図式中の型の配置は、それぞれ先ほどの証明図の $\supset$ 導入則 と $\supset$ 除去則 の図式の論理式の配置と (書体の違いを無視し型演算子 $\rightarrow$ と論理結合子 $\supset$ とを同一視すれば) 一致することがわかる。CH 対応とは、型付けの図式と NJ-証明図の間に見られる、この対応関係のことに他ならない（なお $\supset$ 以外の結合子についても特定の演算子を計算体系に導入し型付け規則を定めて対応を与えることができる）[1][6]。ただしこの対応は単なる類比ではない。CH 対応は、BHK 解釈によって NJ の推論規則に見出される構成の概念と、型付き λ 計算の型付け規則の λ 式形成手続き (すなわち素材となる λ 式の構造を保存しながら型の制約に則って新たな λ 式を構築する手続き) が持つ構成の概念とが同型だということを明らかにしているのである [1][6]。NJ-証明図と対応する型付け図式によって構築される λ 式は、まさにその証明図の終式の結論の BHK 解釈の具体的な形式を与えていると捉えられるのである。ところで CH 対応についてもう一つポイントがある。察しの通り、実は Gentzen の証明図は型付き λ 計算での β 変換と対応する証明図変形機構を備えている [1][4][6]。この変形は—紙幅の都合によりここでは詳細を述べられないが、ともかく—余剰な推論規則の適用箇所 (証明の回り道と呼ばれる) を除去し証明を正規化する操作にあたる。NJ-証明図と対応する型付け図式によって構築される λ 式は、BHK 解釈の体現であると同時に、証明の正規化の手順を与えているのである。[3]

さて近年、CH 対応の「証明に対してその正規化手順を与える」という側面を CL へと拡張する方法が明らかとなった。これを初めて定式化したのは 1990 年の Griffin の論文 [5] である。ここで明らかとなったのは、型付き λ 計算に call/cc と呼ばれる継続呼び出し演算子を導入した拡張体系で構築される式が、NK(詳しくは NK と同等な古典自然演繹) の正規化手順を与えるということであった。

継続 (continuation) とは、ひとつの λ 式のある 1 箇所または 0 箇所の可約部に焦点を当てたときにその可約部にとっての背景となる (いわば無視される)、その λ 式中のその可約部を除いた部分を指す概念である。「計算の残りの部分」とも呼ばれる。たとえば、先の $suc0$ を表す λ 式の計算過程で現れた λ 式 $\lambda y.\lambda x.y(((\lambda y.\lambda x.x)y)x)$ について、$\lambda y.\lambda x.y(\underbrace{((\lambda y.\lambda x.x)y)}_{\bigstar}x)$ の★で示した可約部に焦点を当てたとき、$\underbrace{\lambda y.\lambda x.y(}_{\heartsuit}\underbrace{(\lambda y.\lambda x.x)y}_{\bigstar}\underbrace{)x)}_{\heartsuit}$ の ♡ で示した部分が★にとっての継続である。$\lambda y.\lambda x.y(((\lambda y.\lambda x.x)y)x)$ について、0 箇所の可約部すなわち式全体に焦点を当てたときは、空な表現がその可約部にとっての継続である。簡約における可約部と継続の移り変わりは、先の $suc0$ についての簡

約を例にとれば、$\underbrace{}_{\text{継続}}\underbrace{(\lambda z.\lambda y.\lambda x.y((zy)x))\lambda y.\lambda x.x}_{\text{可約部}}\underbrace{}_{\text{継続}} \triangleright_\beta \underbrace{\lambda y.\lambda x.y(}_{\text{継続}}\underbrace{(\lambda y.\lambda x.x)y}_{\text{可約部}}\underbrace{)x)}_{\text{継続}} \triangleright_\beta$ $\underbrace{\lambda y.\lambda x.y(}_{\text{継続}}\underbrace{(\lambda x.x)x}_{\text{可約部}}\underbrace{)}_{\text{継続}} \triangleright_\beta \underbrace{}_{\text{継続}}\underbrace{\lambda y.\lambda x.yx}_{\text{(0 箇所の) 可約部}}\underbrace{}_{\text{継続}}$ のようになる。さて継続呼び出しとは、「焦点が当たっている可約部の内部の特定の継続表現をその可約部にとっての継続で置き換える操作」に当たる。β 変換が「ある λ 式の内部の特定の変数をその式に右から入力として結合している λ 式で置き換える」操作であることを踏まえると、継続呼び出しとは言わば β 変換の裏返しである。つまり継続呼び出しを含む計算とは、仕組みとしては、$\underbrace{(\lambda \mathsf{x}.\mathsf{L})}_{\text{継続}_A}\underbrace{((\lambda \mathsf{y}.\mathsf{M})(\underbrace{\ldots}_{\text{継続}_B}\underbrace{\mathsf{N}}_{\text{可約部}_B}\underbrace{\ldots}_{\text{継続}_B})}_{\text{可約部}_A}\underbrace{)}_{\text{継続}_A}$ という形の λ 式の 可約部$_A$ の内部の 継続$_B$ を 継続$_A$ で置き換えて $(\lambda \mathsf{y}.\mathsf{M})((\lambda \mathsf{x}.\mathsf{L})\mathsf{N})$ という λ 式を得るような「計算の順序を変更するような計算」である。

つまるところ Griffin により、CL の証明プロセスは、証明中の背景や文脈に相当する部分の補足・制御を含むような、IL よりも複雑な証明正規化手順を与える式についての—しかしそれが BHK 解釈の体現とはなっていない式についての—構築プロセスと対応付けられるということが明らかとなった。

3 筆者の主張

BHK 解釈の条件が崩れている CL に対して上のような CH 対応の拡張が与えられるという事実は、次の見方を導くと筆者は考える。それは、まず IL についての CH 対応においては証明によって構築される λ 式は、第一に BHK 解釈を体現し、第二に証明の正規化手順を与えるのであったが、実はこの二つのことが本来は別のことであり、そして CL の CH 対応においては後者のみが実現している、という見方である。CL によって導出される論理式に対しては、一方で確かに BHK 解釈を与えることはできないのであるが、しかし他方で構築される式が証明正規化手順を与えているという点では単なる二値性に帰着できないような証明の構造に関する概念 [4] をその表現内容としていると見ることができよう。

さて Griffin によって明らかとなったのは、CL で導出される論理式が表現内容としている事柄すなわち証明正規化手順の概念は、IL の証明正規化の際には扱われない証明の文脈・背景に対する操作を含んでいるという事実であったが、さらに近年鮮明となってきたのは、その正規化手順が (IL では失われているような) 対称性を備えているという事実である。特に、CL と CH 対応する体系である Wadler[8] の双対計算 (dual calculus) と Filinski[3]・浅井ら [11][9] の対称 λ 計算 (symmetric lambda calculus、以下 SLC) では、可約部とその操作を定める構文的道具立てと、継続とその操作を定める構文的道具立てとが

すべて互いに対称的な形で与えられ、証明正規化手順を表現する式はそれら対称な両者の結合表現として構築される。つまり、CL によって導出される論理式が表現内容とする証明正規化手順には、IL において扱われる証明に対する抽象／適用操作に加えて、これと対称的な証明の文脈・背景に対する抽象／適用操作が組み込まれているという事実が鮮明となった。この対称性を如何に解釈するかについては様々な余地があると考えられるが、一つのやり方としては、CL により導出される論理式は IL の BHK 解釈的な内容とその双対すなわち「ある論理式が反証についての諸仮定からどのような実効的手続きを経て反証されたか」という BHK 解釈の双対的内容を部分的に併せ持っていると解釈することができるはずだと筆者は考える。そして特にこの点で、$\supset$ とその双対の機能を持つ結合子とさらにそれらについての両義性の機能を持つ結合子の三者が組み込まれていることを特徴とする SLC[11][9] が構築する式は、CL の証明プロセスが構築する証明正規化手順が証明の構造と反証の構造の並存やそれらの間の相互作用をまさに表現していると筆者は考える。

4　今後の展望

SLC が構築する式の形を踏まえて、CL により導出される論理式が表現する証明の構造に関する内容が BHK 解釈的な意味とどのような隔たりを持ったものであるのかを明らかにし、そしてここで CL の命題が持つ本質的な概念とは何かを特定することが今後の課題である。また筆者による研究の途上ではあるが、SLC に対してある拡張—詳しくは〈継続 | 関数 | 項〉という構文的構造の入れ子を許すという拡張—を行った体系 入れ子対称 λ 計算 (nested symmetric lambda calculus) は、これまで与えられていなかった SLC の述語論理の CH 対応への拡張を与えられる体系であると見込まれる。この体系の整備を進め述語論理レベルの証明の分析も行う予定である。

注

[1] ここでは説明のしやすさの都合により、自然演繹を Gentzen[4] のオリジナルの記法ではなくシーケントスタイルと呼ばれる記法によって表している。

[2] 何故このような形の λ 式によって自然数を定めることができるかについては、[10] で論じられている。

[3]CH 対応に対しては「論理の内実を計算が明らかにしている」という見方と「計算の内実を論理が明らかにしている」という見方のいずれもが与えられると考えられるが、本稿では計算論理学の分野で広く認められる作業仮設に従って前者の見方のみを取り上げて議論を進めている。

[4]Frege や Frege を踏まえている Dummett であればこの概念を意義 (Sinn) と呼ぶだろう。

文献

[1] Curry,H.B. and Feys,R. (1958) *Combinatory Logic*, Studies in Logic and the Foundations of Mathematics, Vol.I, 1st edition, Amsterdam: North-Holland

[2] Dummett,M (1973) The Philosophical Basis of Intuitionistic Logic. *Truth and Other Enigmas*, Cambridge: Harvard UP, 1978, pp.215-247

[3] Filinski,A (1989) Declarative continuations and categorical duality. Master's thesis, DIKU Seport 89/11, University of Copenhagen

[4] Gentzen,G (1935) Investigations into Logical Deduction. *Mathematische Zeitschrift 39*, pp.176-210,405-431. Reprinted in M. E. Szabo, editor *The Collected Papers of Gerhard Gentzen*, North-Holland, 1969

[5] Griffin,T (1990) A Formulae-as-Types Notion of Control. *17'th symposium on Principles of Programming Languages (POPL'90)*, pp.47-58

[6] Howard,W.A. (1969) The Formulae-as-Types Notion of Construction. *To H.B.Curry: Essays on Combinatory Logic, Lambda Calculus and Formalism*, Academic Press, pp.479-490

[7] Troelstra,A.S. (1991) *History of constructivism in the twentieth century*, ITLI Prepublication series for Mathematical Logic and Foundations ML-91-05, University of Amsterdam

[8] Wadler,P (2003) Call-by-value is dual to call-by-name. *Proceedings of the eighth ACM SIGPLAN International Conference on Functional Programming (ICFP' 03)*, pp.189–201

[9] 上田 やよい、浅井 健一 (2010) 「型付き対称 λ 計算と古典論理」 第 12 回プログラミングおよびプログラミング言語ワークショップ論文集, pp.34-48

[10] 岡本 賢吾 (2003) 「命題を集合と同一視すること——包括原理からカリー＝ハワード対応へ」 『科学哲学』 第 36 巻 2 号, pp.103-118

[11] 阪上 紗里、浅井 健一 (2009) 「対称 λ 計算の基礎理論」 『コンピュータソフトウェア』 Vol.26 No. 2, pp.3-17

[12] マイケル・ダメット (1973) 「直観主義論理の哲学的基底」 藤田 晋吾 訳 (1986) 『真理という謎』 勁草書房 pp.212-266

（東京都立大学）

新進研究者 Research Note

存在論的多元主義の可能性
The possibility of ontological pluralism

稲田　進一

Abstract

This paper aims to clarify the stipulation and the position of K. McDaniel's 'ontological pluralism' in modern metametaphysics, and to indicate the better direction of it. Ontological pluralism is stipulated as the position that there are multiple analogical categories on the fundamental level of the world, distinguished from 'ontological monism', which admits only one category. My proposal is that the focal point of the disputes about ontological monism and pluralism is not on the linguistic level, but ontological level, and that ontological pluralism is not inferior than monism at present. Furthermore, I propose that ontological pluralists need to explicate the notion of 'analogy'.

(1) 研究テーマ

本研究のテーマは、現代メタ形而上学における一つの立場である、存在論的多元主義（ontological pluralism）の可能性を探ることである。現代メタ形而上学において存在論的多元主義と呼ばれる立場は複数あるが[注1]、本研究が取り上げる存在論的多元主義はK. McDanielによって提示されているものであり[6]、その存在論的見解としての要点を一言で表せば、「存在（being）の種類および基礎的な存在者の領域は複数ある」というものである。現代の分析形而上学では、「存在の種類および基礎的な存在者の領域は一つである」とする一元主義的立場が主流派の立場の一つの特徴であり[10][13]、存在論的多元主義はこれと対立する非主流派の立場であるといえる[7]。分析形而上学においては従来、客観的世界のあり方を明らかにすることが形而上学の目的である、とする見解が暗黙裡に共有されてきた感があり、近年T. Siderによって「客観的世界の基礎的構造の探究」としての形而上学観が強く打ち出されるとともに[11]、それはまた多くの批判も招いているが[7][12]、そのような主流派の立場からすれば「世界の基礎的な存在者の領域が単独か複数か」ということは世界の基礎的あり方に関わる大きな問題であり、十分に検討を要する問題であると言える。

ただし、現在のメタ形而上学の議論は非常に錯綜しており、上述の存在論的多元主義の規定・整理も実際にはもう少し複雑であり、さらなる説明・整理が必要である。したがって、本研究ではまず存在論的多元主義の規定をT. Sider[10]やJ. Schaffer[9]などの存在論的一元主義やE. Hirsh[4][5]の立場と比較しつつ整理し、その上で存在論的多元主義の課

題と可能性について考察する。

(2) 研究の背景・先行研究

近年、分析形而上学の主流派に対して厳しい批判が起こり、「形而上学とは何か？」を問い直すメタ形而上学的議論・整理が盛んになってきている[1][2][7][12]。現代メタ形而上学における多種多様な論点のうちの一つが、「基礎的な存在者の領域は単独か複数か」という点であり、T.SiderやJ. Schafferなどの主流派は「単独である」という一元主義的な立場をとり[9][10][13]、それに対してK. McDanielは「複数である」という多元主義的な立場（存在論的多元主義）をとる[6]。しかし、この一元主義と多元主義の対立と、それを取り巻く状況の説明にはもう少し整理が必要である。

まず、分析形而上学に一般的に見られる傾向として、存在の問題とそれを記述する言語・理論の問題をパラレルに考える、という傾向がある。特に、W. O. Quineの提出した存在論的コミットメントの基準（ある理論がある対象の存在にコミットする＝その対象がその理論の束縛変項の値となる）[8]に従って、最良理論における束縛変項の値となる、つまり量化子のドメインの中にあることを対象の「存在」の条件とする、という考えは主流派形而上学において根強い。そのような流儀に従って、上記の存在論的一元主義と存在論的多元主義の対立も「最良理論における基礎的な量化子の種類は単独か複数か」という対立に帰される形になっている[7][12]。

上述のように存在論的一元主義と存在論的多元主義の対立が論じられているのであるが、広義の存在論的多元主義にもいくつかのヴァリエーションがあるため、ここでMcDanielの多元主義[6]をE. Hirshの立場[4][5]と比較することは有益である。現代のメタ形而上学において、基礎的な量化子の意味が一義的（univocal）かそうでないか、という対立軸があるが[7][12]、主流派とされるSiderやvan Inwagen等の立場[9][13]は一義的であるとし、E. Hirshは一義的ではないとした[注2][4][5]。例えばSiderは、形而上学的な意味で単独最良の量化子の意味があるとし、それは実在の構造を正しく切り分けているとして、「何があるのか」「世界の基礎的なあり方はどのようであるのか」という問いに答えようとするのであるが、HirshはSiderのような形而上学的に特権的な単独の量化子の意味を認めず、複数の同等な量化子の意味がありうるとするのである。この「量化子に複数の意味を認める」という点においては、McDanielのような存在論的多元主義者とE. Hirshの立場は共通する。しかし、以下の二点において両者は異なる。第一に、Hirshは主流派のようにクワイン的存在論的コミットメントの基準に従って「最良理論における量化子のドメインが世界の基礎的レベルでの存在領域に対応する」そして「（基礎的なレベルで）何があるのかはその存在領域のうちにあるか否かで決まる」という立場を明示的にはとる訳ではないが、McDanielはSiderと同様にそのような立場をとる。第二に、Hirshは複数ありうる基礎的な量化子の意味について単に多義的である（あるいは言語ごとに異なる）としているように見えるが[4][5]、McDanielは多義的な量化子の複数の意味の間に、類比的な（analogical）意味連関を認める[6]。

筆者が再整理するところでは、McDanielの言う「類比」あるいは「多義（polysemous）」とは、多義的な意味連関を有する語のうちに成り立つ性格であり、総

称的な意味（generic sense）を有する語について言われるものである。つまり、同一の語として使用されてはいるが、基礎的なレベルでは互いに連関する多義性を有する語について言われるものである。McDanielは[6]においては「類比」あるいは「多義」という意味連関の内実については深く述べていないのが、これと区別されるものとして、「一からの異義（pros hen equivocal）」と、「曖昧（ambiguous）」を挙げている。「一からの異義」は、中核的な意味とそれに従属する周辺的意味に基づく多義的な意味連関を有する表現に用いられるものである。例えば、「〜は健康的である」という語が豆腐について言われる場合と、特定の人物について言われる場合があるが、これを特定の実体が健康的であるという意味が中心的であり、それとの関係において従属的に豆腐が健康的であるという意味が生まれていると考えるならば、これは「一からの異義」の例となる。対して「類比」は、上の「健康的」の例のように中心的な意味を措定せず、複雑な意味連関の網を考える場合がそれである。また他方、「曖昧」は、多義的ではあるがそれらの意味の間に密接な意味連関を有さない表現に用いられ（これは筆者の言葉で言えば、「完全な多義」ということになる。）、「類比」あるいは「多義」は密接な意味連関を有する表現に用いられるとされる。[注3]

ここでのポイントは、McDanielが「存在」述語あるいは存在量化子に対して「類比的」と言うことで、多元的な存在の意味に対して暗黙裡の連関を認めようとしている、ということであり、この点がHirshのような量化子に関する完全な多元主義者とは異なる点である。以上をまとめると、McDanielの存在論的多元主義は、Hirshの立場と比較すると①最良理論における量化子の、世界の基礎的構造に関する含意を認める点、②多元的な量化子に対して類比的な意味連関を認める、という二点において異なる。

また、Siderなどの一元主義とMcDanielの多元主義についてもう少し整理しておこう。両者はいずれも上述の「最良の理論における量化子の世界の基礎的構造に関する含意を認める」という点において共通するが、「世界の基礎的なレベルにおける存在領域を単独とするか複数とするか」という点において異なる。この対立点についても、先述の分析形而上学の流儀に倣って、言語レベルでの「量化子の種類を本質的に単独とするか複数とするか」という問題を通じて議論が展開されている[7][12]。

(3) 筆者の主張

以上のような議論状況に対しての筆者の主張としては、まず一見して分かるように、議論が非常に複雑化しており見通しが立ちにくい点が問題である、という点がある。既に述べたように、分析形而上学においては言語レベルでの論理的分析を通じて存在の問題を探求する、という流儀は伝統的でありスタンダードである。しかし実際のところ、この「存在論的一元主義か多元主義か」という問題に対しては本質的には量化子を持ち出さなくとも考察は可能であるように筆者には思われる。

例えば、J. Schafferは存在の基礎づけ関係（grounding relation）に注目する形而上学理解を示した[9]の中で、量化子の問題を導入することなく自身の存在論的一元主義の立場を規定している。すなわち、形而上学的な世界の構造として「平らな構造」、「分類された構造」、「順序づけられた構造」の三つの種類を導入し、「順序づけられた構造」で

かつ分類された構造でないものとして自身の立場を位置付けている。すなわち、平らな構造とは世界の内に基礎づけ的な構造を認めず、かつ一つのカテゴリー（ここでは、存在および存在者の分類的領域）しか認めない立場であり、分類された構造とは世界の内に基礎づけ的な構造を認めず、複数のカテゴリーを認める立場であり、順序づけられた構造は世界の内に基礎づけ関係の階層を認める立場である。このように、量化子の問題を導入せずとも存在論の問題自体は定式化可能であり、議論可能である。

また、量化子の意味についての対立にしても、Siderは単独最良の量化子の意味があるとする立場、McDanielの類比的な量化子の意味を認める立場、Hirshの完全に多義的な量化子の意味という複数の立場があるが、それらの形而上学的な領域への含意を問題にするかぎり、それらの間の優劣を決定するのは最終的に存在論的な考察であろうと思われる。例えば、「ある抽象的な数が存在する」という文を論理式の形に表す時に、一方では、パラフレーズする先の理論のうちに基礎的な量化子としては無制限量化子「∃x」のみを認め、他の量化子を認めずに、それと「xは抽象的である」を意味する述語「Ax」、「xは数である」を意味する「Nx」を用いて「∃x(Ax & Nx)」と表す仕方がある。他方では、基礎的な量化子としては全ての抽象物かつ抽象物のみそのドメインに含む原始的な制限量化子「∃*x」と全ての具体物かつ具体物のみそのドメインに含む原始的な制限量化子「∃'x」のみを認め、前者を用いて「∃*xNx」と表す仕方がある。このとき、両者は同じく「ある抽象的な数が存在する」を事実を表しているが、前者のパラフレーズを選ぶ場合、Siderのように単独最良の量化子の意味に応じた単一の存在領域があるとする立場と親和的となり、後者のパラフレーズを選ぶ場合、McDanielのように複数の量化子の意味に応じた複数種類の存在領域があるとする立場と親和的となる。従って、事実を記述する理論の選択肢はパラフレーズの仕方に応じて複数あり、存在論的探究のためにいずれを選択するかと言う事に関しては、何らかの基準が必要であると思われる。そして、そのような選択に関しての基準を与えるのは、結局それが存在のあり方をよく捉えているかどうか、ということになると筆者には思われる。そうであるとすれば、むしろダイレクトに存在論的な議論として、存在のあり方、世界の構造というものを考察した方が見通しが得やすいと思われる。

以上のことから、以下は存在論的多元主義に関して、量化子についての問題は扱わず、存在のあり方レベルでの議論に限る事とする。その上で、J. Schafferの提示した世界の構造の区別を土台として議論を進めたい。まず、Schafferの三図式を前提とすると、存在論的多元主義は「分類された、順序づけられた構造」ということになろう。つまり、Schafferの図式でいうと「分類された構造」と「順序づけられた構造」は両立可能であり、順序づけられた構造のうち、基礎的なレベルでのあり方において複数のカテゴリーがあると見なすものが存在論的多元主義、ということになる。これに対して、基礎的なレベルでのあり方を単一カテゴリーのものと見なすものが存在論的一元主義、ということになる。

Schaffer自身は、順序づけられた構造について、基本的には一元主義的なものしか検討はしない[9]。というのも、Schafferによれば多元主義的な「分類された、順序づけられた構造」は、さらなる基礎づけ関係の導入により一元主義的な「順序づけられた構造」に帰することができるからである。つまり、世界の基礎的なレベルでの複数カテゴリーを、よ

り基礎的なレベルでの単一カテゴリーからの複数の基礎づけ関係による派生と見なすことができるだろう、ということである。

しかし、この点についてはSchafferは単に一元主義的な構造が正しいという自身の立場を述べているだけであるのか、多元主義的な構造であってもつねに一元主義的な構造からの基礎づけ関係による派生型という形に還元できると主張しているのか、明らかではない。前者であれば、筆者が思うに、それは単純にSchafferの独断であり、後者であれば、多元的な基礎的カテゴリーの構造をより根源的な一者からの派生と見なせる、という主張自体が検討を要する問題であろう。後者のように「基礎的な」レベルでの多元的カテゴリー構造をSchafferがいうような形で一元主義的構造に還元するのであれば、少なくともその「派生」を実現する基礎づけ関係が複数種類なければならない。そうでなければ、McDanielがそもそも主張しているカテゴリーの「多元性」自体の本質を取りのがしていることになってしまうであろう。しかし、そのような複数種類の基礎付け関係を導入することは、McDanielのように複数の基礎的なカテゴリーを認める事と本質的に何が異なるのか、明らかではないように思われる。少なくとも、新たに導入された根源的一者と、複数種類の基礎づけ関係の内実が明らかにされねばならないだろう。従って、Schafferのように一元主義的な構造を自明と見なす事は現時点では妥当ではない。世界の基礎的なレベルでのあり方に関する存在論的一元主義と存在論的多元主義は、少なくとも現時点では同等の候補と見なされるべきであろうと思われる。

また、本稿の整理で言えば、存在論的多元主義においてもMcDanielのように基礎的あり方について「類比的」であると見る立場と、それぞれのカテゴリーが独立しており完全に「多元的」であると見る立場、McDanielが「一からの異義」と呼ぶ関係により捉える多元主義的立場がある。この「一からの異義」的立場は、SiderやSchafferのような存在論的一元主義者とMcDanielのような「類比」的多元主義者の中間のような立場となるだろう。つまり、世界の基礎的レベルでのカテゴリーの一元性・統一性が強い方から順に、SiderやSchafferのような一元主義、「一からの異義」的な多元主義、McDaniel流の「類比」的な多元主義、完全な多元主義、という並びとなる。このように整理すれば、現在の存在論的一元主義と存在論的多元主義の対立も、かなり見通しが良いものとなろう。

その上で、問題となるのはMcDanielのいう「類比」の中身である。McDanielはこの「類比」の内実については深く述べてはいないが、「一からの異義」のように中心的な意味が措定されるものではない仕方で、何らかの連関があるものとしている[6]。この「類比」の内実、および世界の基礎的なレベルでそれが成り立っているかどうかが、存在論的一元主義と存在論的多元主義の優劣を決定するところであり、また探求の実質的な争点であろうと思われる。したがって、まずは「類比」概念の解明という哲学的な仕事が重要である、ということは言えるだろう。

(4) 今後の展望

今後の展望としては、まず既に述べたようにMcDanielの言うような「類比」概念の解明という課題がある。McDanielの「類比」あるいは「総称的意味」概念は、トマス・アクィナスやハイデガーの哲学からインスパイアされたものとなっており、また定義不可能な意味連関という点はウィトゲンシュタインの「家族的類似性」を思わせる。要するに、

哲学史上ですでに近い議論が何度も現れている概念である。この「類比」概念の内実をより明らかにしておく事が、存在論的多元主義にとっては重要な課題であると思われる。

注1：主に中期カルナップの立場を指すもの、新カルナップ主義者によるもの、K. McDanielによるもの等がある。
注2：Hirshの立場は量化子変動（quantifier variance）と呼ばれる。
注3：正確には、McDanielは[6]のにおいて総称的意味をもつ場合を「類比」とし、そのうち複数の意味をもつ場合を「曖昧」とし、曖昧のうちで複数の意味の間の密接な意味連関がある場合を「多義（polysemous）」とし、多義のうちで中心的な意味が存在する場合を「一からの異義」とする、というようにも整理している。他方、ある表現が「一からの異義」かつ「類比的」であることはないと述べるなど、やや整理に不明な点が見受けられる。ここでの筆者の再整理は、McDanielが念頭に置いていると思われる区別を分かりやすくする為に、上の意味での「類比」と「多義」を同一視した上で、類比、一からの異義、曖昧を相互排他的なものとして捉え直したものである。これは[6]に忠実であろうとすればややミスリーディングということになるが、正直なところ筆者には上述のMcDanielの整理の利点が理解し難く、あまりに煩雑と思われたので、このように整理し直した。

（東京大学）

(5) 参考文献

[1] Berto, F. and Plebani, M. (2015). *Ontology and Metaontology: A contemporary Guide*, London: Bloomsbury.
[2] Chalmers, D., Manley, D. and Wasserman, R. (ed.) (2009), *Metametaphysics, new essays on the foundations of ontology*, Oxford University Press.
[3] Chalmers, D. (2009). 'Ontological Anti-Realism', in Chalmers et al. (ed.) (2009).
[4] Hirsh, E. (2003). 'Quantifier Variance and Realism', Philosophical Issues 12 (1): 51-73.
[5] Hirsh, E. (2009). 'Ontology and Alternative Language', in Chalmers, Manley & Wasserman (ed.) (2009).
[6] McDaniel, K. (2009). 'Ways of being', in Chalmers et al. (ed.) (2009).
[7] Manley, D. (2009). Introduction: A Guided Tour of Metaphysics, in Chalmers, Manley & Wasserman (ed.) (2009).
[8] Quine, W. V. (1948). 'On What There Is', Review of Metaphysics 2:21-38.
[9] Schaffer, J. (2009). 'On What Grounds What', in Chalmers et. al. (ed.) (2009).
[10] Sider, T. (2009). 'Ontological Realism', in Chalmers et. al. (ed.) (2009).
[11] Sider, T. (2011). *Writing the Book of the World*, Oxford University Press.
[12] Tahko, T. E. (2015). *An Introduction to Metametaphysics*, Cambridge

University Press.
[13] van Inwagen, P. (2009). 'Being, Existence, and Ontological Commitment', in Chalmers et al. (ed.) (2009).