応用倫理―理論と実践の架橋―

Vol. 12
2021 年 3 月

北海道大学大学院文学研究院
応用倫理・応用哲学研究教育センター

目　次

アルゴリズムの判断はいつ差別になるのか ── COMPAS 事例を参照して

前田春香（東京大学大学院、理研 AIP センター）

要旨

本論文の目的は、Correctional Offender Management Profiling for Alternative Sanctions（以下COMPAS）事例においてアルゴリズムが人間と似た方法で差別ができると示すことにある。技術発展とともにアルゴリズムによる差別の事例が増加しているが、何を根拠に差別だといえるかは明らかではない。今回使用する COMPAS 事例は、人種間格差が問題になっているにもかかわらず、そのアルゴリズムが公平であるかどうかについて未だ論争的な事例であり、さらには差別の観点からは説明されていない。本論文では、「どのような差異の取り扱いが間違っているのか」を説明する差別の規範理論を使って COMPAS 事例を分析する。より具体的には、差別の規範理論の中から「ふるまい」による不正さを指摘する Hellman 説を適切なものとして選び、アルゴリズムに適用できるよう改良したうえで COMPAS 事例が差別的であるかどうか分析をおこなう。この作業によって、差別的行為を「ふるまい」の問題として独立させ、一見差別の理論が適用できなさそうなアルゴリズムによる差別性の指摘が可能になる。

Abstract

This paper aims to explain how algorithms can discriminate wrongly against humans in a human-like way. With technological developments, discrimination by algorithmic systems has become an issue. However, it is not clear what the wrongness of discrimination by algorithms is. It is also true for a famous case of Correctional Offender Management Profiling for Alternative Sanctions (hereafter COMPAS). The case has been controversial because of the fairness of the algorithm. Besides, the question of whether it is morally wrong discrimination or not remains to be seen. In this paper, I will give an analysis of the case of COMPAS by focusing on the point of whether this algorithm discriminates in a wrong way or not. First, More concretely, I will choose Hellman's account, which can detect the wrongness of discrimination based on the behavior of the discriminator, as an appropriate one for algorithmic discrimination. Secondly, Hellman's account assumes that humans as a discriminator will be improved in terms of power for this account to apply to algorithms. After that, I will offer an explanation of the case of COMPAS from an ethical viewpoint. This paper may provide a way to explain the wrongness of discrimination by algorithms based on their behavior even if it seems to be impossible for a theory of discrimination to apply to algorithms.

はじめに

本論文の目的は、COMPAS 事例を用いて、アルゴリズムが差別を擬似的に ――つまり人に似たやり方で ――行えると示すことにある。

2008 年、Hewlett-Packard（以下 HP 社）のノートパソコンの画像認識プログラムが黒人を認識しないという形で「差別」したとして、投稿者に批判の意図はなかったにもかかわらず、YouTube に投稿された動画[1]が炎上した。HP 社は自社ホームページに掲載した謝罪で、「照明の問題だ」と説明した（Hing 2009）。つまり、これは技術的な問題にすぎず、差別の問題ではないのだというわけだ。

以来、アルゴリズム[2]によって人間が差別されている（とされる）事例は増え続けている。Google の検索エンジンは黒人女性の問題のある表象を提供し（Noble 2016）、Google アドセンス広告は黒人らしい名前には、白人らしい名前より高い頻度で、逮捕の経歴を示唆する広告を返す（Sweeney 2007）。今では雇用やローンといった、より社会的影響が大きい分野においてもアルゴリズムは重大な問題を引き起こし、人々の間の格差を広げているという（O'Neil 2016）。このような偏ったアルゴリズムによる問題全般を指して、アルゴリズムの偏見（algorithmic bias）と呼ばれている（Koene et al. 2018）。

アルゴリズムによってもたらされるこのような現象は、直観のレベルで全く差別的でないと言い切ることは難しいだろう。実際にそのような現象が炎上によって知られたり、その後も話題にされたりしていることは、われわれがこの問題を不正なものと認識していることの証左であるといえよう。しかしさらに一歩踏み込んだ理論的なレベルでは、どのような偏りが倫理的に問題があるのか、つまり道徳的に不正な差別とよべるのかは説明がなされていない。実在の事例は多々あるものの、個別の事例の分析は不十分であることが指摘されている（Sandvig et al. 2016）。本論文では、アルゴリズムによって提示されるどのような偏りが差別的であるといえるのかを検討するために、Correctional Offender Management Profiling for Alternative Sanctions（以下 COMPAS）の事例を分析する。

COMPAS とは、インプットデータに質問紙調査の結果を使い、ブートストラップ法や生存率曲線を用いて再犯リスクを診断する統計的自動判断プログラムである（Brennan et al. 2009）。2016 年 5 月、公益を目的とした独立系報道機関である Propublica が、当該プログラムの判断には人種間格差があると告発し、大きな議論を巻き起こした。当該事例を参照する理由は二つある。第一に、COMPAS が偏ったアルゴリズムの典型と目されていることが挙げられる。実際にアルゴリズムの出力は特定の集団に対して偏っており、さらに当該の偏りは数理的に根拠があるものである。かといって、数理的な根拠があれば差別して良いのだろうか。このように、当該の事例はそもそも公平、ひいては差別とは何か、という問題に関連するものだといえる。次に、COMPAS に関する論争は、調査者側は偽陽性と偽陰性に基づく不公平を主張したのに対して、制作側はキャリブレーションを元に批判に応答した（Dressel and Farid 2018）。つまり制作側は、問題になっているリ

1　https://www.youtube.com/watch?v=t4DT3tQqgRM
2　本論文におけるアルゴリズムとは、一定のタスクを遂行する際の計算式をいう。本論文では、アルゴリズム、プログラム、ソフトウェアは区別せずに扱う。

スクスコアが同一の両人種は、それぞれが同じ割合で再犯しているため、人種差が存在せず、問題ないという認識を示したというわけである。この二つの基準はどちらも数理的な公平性基準であるが、同時に満たすことができないことが数学的に明らかになっている（Kleinberg et al. 2017）。このことは、数理的なこの二つの基準と、道徳的に問題がある差別の基準との関係を探究する必要性を示唆する。

このような問いを扱う理論が、差別の規範理論である。差別の規範理論では、異なる処遇のうちどのようなものが道徳的に不正であるかを論ずる。例えば、性別によってトイレを分けることは男女のどちらも差別していない。その一方で、性別によって雇用の処遇を変えることはおそらく差別の誹りを免れないであろう。これら二つを分かつものは何だろうか。

本論文ではまず、差別の規範理論の中でどのような説がアルゴリズムによる差別に適しているのかを確認した後、COMPAS 事例の検討に移る。先取り的に言ってしまうと、Hellman 説の適切さを示した後に、一見人間にしか満たせないような要件をアルゴリズムも満たすことができると示し、修正した Hellman 説を COMPAS 事例に適用する。

この作業によって、以下のような意義が見込まれる。これまでのアルゴリズムによる差別の研究では、格差の拡大など社会に及ぼす影響をもって悪を判断する説明が主であった（Sandvig et al. 2016）。本論文では、判断そのものを不正なものと考えられる基準を提供するとともに、その適用例を示すことで、より早期にその差別性を察知することを目指す。というのは、現状ではアルゴリズムの判断が引き起こす帰結としての格差が観察されてから対処することになるが、ふるまいによって不正であるかどうかが検討できるならば、より早い対処が可能になるからである。また、COMPAS 事例は有名ではあるが、不正な差別であるかどうかは未だ明確にされていない。不正さの根拠を示すことによって、他の事例への適用への道が開かれる。

1. 事例の説明とこれまで付されてきた説明

本節では、COMPAS 事例の分析に先立ち、事例の概要を提示し、いかなる道徳的不正さに関係する解釈が提供されてきたかを参照する。当該の解釈には正確性や個人の尊重原理、公平性基準といったものがあるが、COMPAS の場合にはいずれも不十分なものであるため、新たに差別の不正さにかんする道具立てが必要であることを示す。

1）COMPAS 事例の概要

まずは、COMPAS による事例がどんなものであるかを確認しておこう。

2016 年 5 月、Propublica は “Machine Bias” という記事[3] で、COMPAS にまつわる以下の問題を告発した。その報告によれば、被告に対して実際に再犯したかどうかを追跡調査したところ、黒人[4] は実際には再犯していないのに高リスク群としてラベル付けされた頻度が白人の 2 倍であり、

3　当該記事には、再犯リスクが低いと判断された白人男性のインタビューが掲載されている。その白人男性は万引で逮捕され、犯罪履歴には加害暴行や複数回にわたる窃盗、重罪である麻薬密売があったが、再犯リスクが低いと判断された。インタビューで彼は、提示されたリスクが「あまりにも低いので驚いた」と述べている。その後一年以内に、彼は 1000 ドルの万引を含む二つの重罪によって再び告発された（Angwin et al. 2016: pr. 42-44）。

4　ここにおける「黒人」「白人」とは、保安官事務所に保管される書類上に記載されているものを指している。

白人が実際には再犯しているのに低リスク群としてラベル付けされる頻度は黒人よりも高かったという（Angwin et al. 2016）。つまり、全体として黒人が不利に判断されているというのだ。

具体的な調査内容は以下の通りである。COMPAS と同様のアルゴリズムを入手し、フロリダ州ブロワード郡で 2013 年と 2014 年に 7000 人以上に割り当てられた再犯に関するリスクスコアを算出、さらにその再犯予測が現実と合致しているかどうか、その後の 2 年間にわたり追跡調査を行った（Angwin et al. 2016）。一般犯罪の再犯予測が合致している確率は 61%、暴力的な再犯予測が合致している確率はおよそ 20%だった（Larson et al. 2016）。さらにリスクと照合する形で人種や年齢といった要素のロジスティック回帰分析が行われており、当該の分析によって高リスク判定と属性に関連性があることが明らかになった[5]。

この分析結果は、結果を提供した Propublica と制作企業 Northpointe 社の論争を招いた。Angwin ら Propublica 側にとって、格差を評価する基準となっているのは偽陽性と偽陰性の割合が同程度でないことである。つまり、黒人は本来よりも「危険」であると判断される可能性が高く、白人は本来よりも「危険でない」とされる可能性が高い。一方で Northpointe 社は、最も標準的に用いられる公平性の別の基準であるキャリブレーションをもって反論した。反論の根拠としては、第一に再犯予測は高リスク群の黒人にも白人にも同様の正確性を有していること（predictive parity）、第二に黒人でも白人でも再犯者と非再犯者を同様に見分けることができること（accuracy parity）、第三に再犯可能性を算出するために与えられたスコアに人種差がないこと（calibration）が挙げられている（Dieterich et al. 2016）。言い換えれば、Northpointe 社は、どちらの人種に対しても COMPAS は再犯するかどうかを同程度正確に予測できる、したがってキャリブレーションの条件を満たしており公平である、というのだ。実際、Sumpter（2018=2019:82）の検証によれば、ハイリスクを割り当てられた黒人のうち 6 割が再犯し、同様にハイリスクを割り当てられた白人のうち 6 割が再犯しているといい、確かにキャリブレーションの観点からは問題がない。

現在もこの問題は多数の論者を巻き込んで、その予測がどれほど公平であるかどうかなどについて論争が続けられており、未だ決着をみていない。

2）「公平さ」の評価はうまくいっているか

以上では COMPAS という事例の特徴と、その特徴が公平性の基準と関連していることを指摘した。本項では、公平性の基準がうまく差別を指摘できていない上、一般に言及される個人の尊重原理によってもここでの道徳的問題を捉えることができず、したがって別の基準が必要であることを示す。

先述の論争は、どちらかが本当のことを言っていて、どちらかが誤っているというものではない。このように対立する根本的理由は、Propublica が挙げた基準である分類均一性（classification parity）と、Northpointe 社が挙げた基準であるキャリブレーション（calibration）の基準を両立することができないことにある（Kleinberg et al. 2017）。少なくとも今回のケースでは、数学的に差別であるかどうか、公平であるかどうかを判断することはできないのだ。そしてそう考えるならば、Propublica と Northpointe 社の論争は自然なものであったとすら言えるだろう。

5　詳しくは Larson ら（2016）を参照。

もちろん実効性の観点からして、ある公平性基準をみたせないならば差別としてもよいのではないか、という反論がありうる。しかし、どちらの公平性基準も固有の限界を有すると指摘されている（Goel et al. 2018）。分類均一性については、偽陽性率はグループ全体の再犯率に応じて機械的に増加する傾向が指摘されており、当該の傾向は infra-marginality 問題として知られている（Ayres 2002 など）。つまり、再犯率が実際に白人よりも黒人の方が高い場合に、当然の帰結として、黒人の方が再犯率が高いと判定することになるのである。これが差別であるといってよいかは疑わしい。一方でキャリブレーションについては、この基準を満たしたまま差別を行うことが可能である（Corbett-davis and Goel 2018）。例えばローン貸与の文脈で、白人と黒人で同程度のデフォルト率を想定し、住所とデフォルト率に関連性をもたせたとする。しかし住んでいる地域と経済的水準、そして経済的水準と人種に密接な関係性がある場合には、合法的にこの方法でハイリスクな人種を排除することができる[6]（Corbett-davis and Goel 2018: 1）。このことによって、ハイリスクとされる人種はローンを供与されないことになる。結果として、さらなる格差につながりかねない。Binns（2018）によって公平を数理的に定義することが一般に困難だと指摘されているように、差別も数理的に定義することは難しいのである。

ならば次の問いは、どのように差別を特徴づけすることが適切なのかというものになる。どのような集団も異なっていて当然の中、われわれはある差異の抜き出し方が間違っていて、差別的であると感じることがある。では特定の仕方で間違っていると何に基づいて主張できるのだろうか。

一つの仕方は、法律上保障されるべき人権に基づいて説明することだろう。COMPAS の事例では、このような自動判断プログラムを量刑判断に使用することは人権の一つであるデュープロセスの権利を侵害しているとして、実際の被告である Eric Loomis によって裁判がおこされている。被告の反論内容は、1. COMPAS のアルゴリズムの内容が不明であることは、正確な情報に基づいて判断されるという被告の権利の侵害である、2. 個別に判断される権利を侵害している、3. 評価に不適切なジェンダー差がある、の三つであった（State *v.* Loomis）。

これらの論点は正確性・信頼性に関するもの（1 と 3）と個人を尊重しているか否か（2）に分けることができる。州最高裁はこの請求に対して、確かにジェンダーによって異なる尺度が用いられていたことを認めるものの、その異なる尺度はプログラムが非公開であるため[7]検証できず、ただちに不適切だとは判断しかねるとする（State *v.* Loomis）。もっとも、州最高裁もプログラムが不正確である可能性や偏見については指摘しているが、当該プログラムが差別的な影響力を持つかどうかについてまでは明確な判断を行っていない（山本・尾崎 2018）。

仮に正確であったとしても、まだ残っている問題がある。それは、自動判断プログラムを使用して個人の量刑判断を行うことは、本人を尊重しているとはいえない、というものである。

これは COMPAS のみならず、統計的手法を用いて個人の評価を行うツールに共通する問題である。個人評価にツールを使用することは個人を尊重しているとはいえないのだろうか。そもそも人を尊重して判断するということはどういう行為であり、それは可能なのだろうか。Schauer

6　当該の慣行はレッドライニングとして知られている。
7　ここでいう「内容が不明」とは、プログラムの中身が技術的にどうなっているかわからないといういわゆるブラックボックス性を指すのではなく、当該プログラムが企業の販売品、すなわち営業秘密として保護の対象となることを指している。

(2003) は、個人の尊重をあらゆる不完全な一般化に基づいた取り扱いを排除するものとして理解する場合、そもそも個人を尊重して判断することは不可能であるという。一見個々人に向き合った取り扱いでも、人間はある程度の統計的一般化を行いながら日常生活を遂行しているため、究極的には統計的一般化に基づいているといえるからだ。

確かに、ツールを使用して判断するとき、ツールで計測できない変数は棄却される。例えば被告が酩酊状態だったことや、生活できないほど貧困だったことなどが考慮できないかもしれない。しかし、ツールや感情に流されずに判断するために裁判官がいるのであり、最高裁もツールに全面的に依拠しないよう注意喚起をしている[8]。以上の理由から、プログラムを量刑判断に使用することは合憲であるとされ、この請求は棄却されている。付け加えるならば、COMPAS は、再犯リスクの説明変数として、例えば犯罪心理学で犯罪と関連すると明らかにされている変数を計測する 137 項目をリスク判定の材料としている（Propublica 2016）。これは、少なくとも人間が行う慣行であり人種だけを根拠とする人種プロファイリング[9]よりは、個人をより多角的に判断しているということもできるだろう。実際に Loomis の「ジェンダーが不適切に考慮された」との主張に対して、ジェンダーのみに基づいて量刑判断が行われたわけではなく、量刑判断において多角的な判断がなされていたこと、現在告訴されている犯罪の内容と被告人の犯歴が重視されていたと裁判所は判示している。このことは、裁判所が COMPAS が多角的に当該個人を考慮しており、個人の尊重原理に反するものではないと考えていることを裏付ける（山本・尾崎 2018）。

以上より、COMPAS の道徳的問題の分析のためには、正確性、もしくは個人の尊重原理からのアプローチは不十分である。しかし「自動判断プログラムを量刑判断に使用することはデュープロセスの権利の侵害ではない、したがって米国憲法に定める平等違反ではない」というこの判断には法学の立場から多くの批判がある（Harvard Law Review 2017; Beriain 2018; Freeman 2016; Liu et al. 2019; Washington 2019）。本稿ではこれらとは別の角度から、すなわち道徳的に不正な差別かどうかという観点から批判を加えたい。

第一に、法学的に規定される差別と、道徳的に不正とされる差別が侵害する権利には違いがある。後者のほうがより広い範囲を包含しているのだ。ここでいう法学的に規定される差別は、米国でいう米国市民権法第七編であり、日本での男女雇用機会均等法でいう差別をさす。両者が禁じている間接差別とは、一見中立的な処遇によって属性によって偏った不利な影響を生じさせるような取り扱いのことである。この偏った不利益は明確な比較対象を持って算出する必要があり（Khaitan 2019）、影響が明らかでなく計算できない場合には効果が薄いと考えられる。

第二に、これがより重要な理由であるが、法学的な議論では、当該プログラムの差別性については結局のところ判断されていないということである（山本・尾崎 2018）。裁判所は確かにプログラムを使うことに対して注意喚起を行ったが、これはプログラムを使う人間だけに着目したものであり、プログラムの影響を無視している。とはいえ、判断を効率的もしくは正確にするような影響がないならばプログラムを使用する意味はない。人がプログラムを使用することを原則とし

8　これは GDPR が掲げる、人間は自動判断だけで判断されない権利を有し、非倫理的な判断を防止するために人間を配置する必要があるという方針と一致する。
9　ここでは、人種を根拠にして街頭でのパトロールで聞き取りをするかや捜査をするかどうかを決めることを指す。詳細は後述。

ながら、いかに共同の判断を「よく」すればよいか、すなわちその使い方がここでの問題なのだ。ならば、まずはプログラムがアウトプットするどのような判断が道徳的に許容可能かという問いに答える必要がある。

2. 道徳的側面からの分析

以上では、COMPAS の道徳的問題の有無を解釈するには新たな道具立てが必要であることを述べた。本節では、新たな道具立てとして差別の規範理論が適切であることを示す。次に、差別の規範理論の説の中で適切なものを選び、当該の説を詳細に説明する。

1）規範理論からの検討

どのような判断が道徳的に許容可能であり、どのようなデータの偏りが許容できるのか、その問いに答えるために使用するのが差別の規範理論である。差別の規範理論とは、差別がなぜ不正なのかを探求する哲学的理論である。

差別の不正さの由来には、大きく二つの説がある（堀田 2014）。一つは、被差別者に与える害を基準とする害ベース説、もう一つは基準を尊厳の侵害に求める尊厳ベース説である。この二つは絶対的原理（bedrock principle）として採用しているものによって分けられているため、害ベース説と尊厳ベースが同じ基準を含みうる。例えば、害ベース説でも個人の尊重に言及することはあるし（Knight 2013）、全ての尊厳ベース説が害に無頓着であるというわけでもない（Hellman 2008=2018: 3）。

ではどちらの説がアルゴリズムによる差別に照らしてより適切なのだろうか。尊厳ベース説ではアルゴリズムが社会に及ぼした影響や不利益を不問にしたまま、当該の判断が社会的文脈に照らしてどうか、を判定する。すなわち、仮に実際のサービスに何ら影響を与えなかったとしても、研究室の中で画像認識プログラムが肌色の暗い人を認識しなかった場合には、やはり差別的だと考えることができるだろう。尊厳ベース説ならばこのような含意をくみ取ることができる。

もちろん、害ベース説でもその悪質さを説明・予見することはこの COMPAS 事例においても可能である。人種によって不均等なリスク判定はそのまま、例えば不均等な量刑判断を導きうるし、それはすなわち害に直結するだろう。人種プロファイリングにおいても、害を予見することによってその悪質さを批判することができる（Eidelson 2015）。アルゴリズムが人間の意思決定を真似るものである以上、このような司法だけではないあらゆる場面で、特定の人種が不利に扱われることが予見されるため、予防もできるかもしれない。

しかし、尊厳ベース説にも害ベース説に比して、無意識的な差別や構造的差別を扱えるという利点がある（Lippert-Rasmussen 2006: 21）。この利点はアルゴリズムによる差別においてはより重要になりうる。なぜならアルゴリズムの使用は、われわれの差別意識を顕在化させたり、もしくは差別の件数を増加させたり、その効力を増幅させる可能性があるためである。あるいは、差別であるとわれわれが気付かないケースの可能性もありうる。例えば、われわれ全員が属性 B よりも属性 A が優れていると信じ切っている場合、差別的待遇を不正なものとして受け取らないかもしれない。つまり、害をいかに認識するかについての問題がある。人間がどう考えるかがアル

ゴリズムによって再現され、それに基づく判断が実行されれば、もはや人々がどう思っているか、その評価にたいしてどう思うかは重要でなくなる。アルゴリズムによる差別は、構造的な差別の一部なのである。

2）いかなる説が適切か

尊厳ベース説の中でも、何を根拠にして尊厳を侵害していると判断するかは諸説ある。次は、アルゴリズムによる差別においてどれが最も有効な基準なのかを特定しなければならない。

まず、Alexender（1992）は、行為者の偏見や悪意によって差別行為の不正さを指摘する心的状態説を提唱しているが、これはあまり望ましくない。悪しき心的状態になくても、可能な差別的行為が統計的差別である。統計的差別[10]とは、多くの人々から特定の目的に沿った人を選出しようとするときに、特定の目的を十分に代替すると思われる属性を持つ人々を選ぶことが集団間に異なった影響をもたらす場合に使用される言葉である（Arrow 1971; Phelps 1972 など）。この場合、必ずしも選出者は悪しき心的状態によって特定の属性を選んでいるわけではない。

次に、行為の表現内容によって尊厳を侵害しているかどうかを指摘できる表現説がある（Hellman 2008=2018; Scanlon 2008）。例えば、ジェスチャーによって侮辱を表現することには問題があると解される。両者の説には意図を考慮するか、考慮しないかで重要な違いがあり、その点でもっともラディカルになりうる説といえる。

最後に、心的状態を考慮に入れるものの、差別者の行為や態度を総合的に斟酌するものとして熟慮説がある（Eidelson 2015）。

アルゴリズムによる差別には、心的状態によってその不正さを問うことは不可能である（Binns 2019）。このことに照らせば、もっともアルゴリズムによる差別の不正さを問題化するためにふさわしいのは表現説、とりわけ差別者の心的状態を考慮しない Hellman 説であるということになる。

3）Hellman 説の詳細

Hellman 説における不正な差別が、個人の尊厳を尊重しないようなものであることは先述した。Hellman は当該の不正な差別を貶価と呼んで、その成立条件を二つに分けている。一つは被差別者の尊厳を毀損するような表現内容がなされるという表現条件、もう一つは当該表現を押し付ける権力性を意味するヒエラルキー条件である。ただ個人を侮辱するだけでなく、侮辱によって表現された劣等性を押し付けることによってはじめて貶価になるのだ。

これでは人前で上司が部下を侮辱することと差別行為の区別がつかないので、もう少し説明が必要である。上記の二条件を満たすような上司が部下を侮辱する場合であったとしても、その侮辱の理由が国籍や人種、性別といった攻撃を受けている人が所属しているカテゴリーに関連したものであれば、われわれには差別的だと感じられるであろう。さらに、その上司は当該カテゴリーに所属していない人に対しては侮辱を行わず、対等に接するものとする。この場合、侮辱の理由は当該カテゴリーに由来しているものとみることができ、そのカテゴリーに属しているかいない

10　本来 Arrow や Phelps がさす統計的差別には、雇用者と被雇用者の選択が合致してしまうことにより状況が固定化されることを問題視するが、ここでは本文中にあるような意味で取り扱う。

かによって、上司は差異をつけた処遇をしているとみられる。このような差異処遇は道徳的に不正だといえるだろう。

両者は何が違うのだろうか。Hellman によれば、当該の属性の背負ってきた歴史を含む社会的状況によって、当該の属性を基準に分類することそのものが不当になる可能性をはらむ（Hellman 2008=2018:35-42）。ここで着目したいのは、「分類することそのもの」という言葉である。上の例でいえば、上司が部下を属性によって侮辱する場合に、当該属性によって不利な状況が生じるからではなく、当該の差異処遇 ── ここでは侮辱すること自体が、道徳的に不正になりうるというのだ。

とはいえ、侮辱自体が道徳的に問題であることに疑問の余地はない。上司が人前で部下を侮辱しているとき、その侮辱が人種に基づいたものであろうとなかろうと、その行為は同様に不正である。では誰に対してなされても一般的に行為自体が不正であるとは考えられないような、次のような事例を考えよう。校長がバスの前方座席に白人生徒を座らせ、後部座席に黒人生徒を座らせることには道徳的に問題があるだろうか（Hellman 2008=2018:37）。Hellman によれば、この差異処遇には問題がある。なぜなら、私たちの[11]文化的理解では、公共交通機関の座席を人種ごとに区分することは劣等性を意味すると考えられるからである（Hellman 2008=2018:37）。しかも、人種によって人々を分離することは特定の人種の劣等性を表現するために、これまでも行われてきたことである。また、校長の命令は一般的な生徒にとって逆らえるようなものではない。当該の命令は、仮に被差別者がその意図に気づかず、差別されたことによる害を感じなかったとしても道徳的に問題があると考えることができる。

したがって、当該の表現行為が不正であるかどうかは表現内容を検証し、強制力を持つかどうかで判定すればよい。その基準になるのは、当該行為をおこなうもの、当該行為がおこなわれた文脈、当該行為において用いられる言葉である（Hellman 2008=2018:90）。これらはそれぞれ、話者がヒエラルキー条件、文脈と用いられる言葉が表現条件に大まかに対応しているとみることができるだろう。表現の内容の検証には、当該の表現がどれほど特定の属性の劣等性を示す典型的な手段と類似しているか検討することも含まれる（Hellman 2008=2018:61）。これを以下では慣習条件と呼ぶ。

アルゴリズムもこの条件を人間と同様にみたせるのであろうか。文脈と用いられる言葉が問題になりうることはアルゴリズムでも同様といえよう。実際、われわれは多くのアルゴリズムによる差別的な表現を道徳的に問題があるものとみなしてきた。Google Photo のアプリが肌の色が暗い人にゴリラとタグ付けしたことで炎上し（Hing 2009）、Tay によるヘイトスピーチもわれわれは確かに問題であると感じるのはその典型的な例だ。しかし、行為者についてはどうであろうか。アルゴリズムは一般に道徳的行為者たりえず、その立場を人間と同一視することはできない。それにもかかわらず、Hellman 説の図式において、アルゴリズムは人間同様に道徳的問題をはらんでいると言えるのだろうか。

11　ここでいう私たちの文化とはどこの文化なのかは疑問の余地がある。もっとも狭くとらえるならば Hellman のいるアメリカの文化ということになるだろう。ただし、日本においても歴史の教科書でこの状況がよく知られていると考えるとき、同じように人種による座席の分離が劣等性を意味すると考えられても不思議はない。

3. 理論および COMPAS 事例の批判的検討

2 節では、本論文で使用する説について説明した上で、依然として課題が残っていることを示した。その課題は道徳的行為者性に関係するものである。本節では、Hellman 説の図式においてアルゴリズムが行っていることが差別であると言えるかについて、理論的側面の検討、および COMPAS 事例の分析を通して検討をおこなう。

1）Hellman 説の理論的検討 ── ヒエラルキー条件

2 節の最後では、アルゴリズムが道徳的行為者でないために、差別行為を行うことができず、したがって Hellman 説の問題にならない可能性に言及した。この問題を追及するために、以下では差異処遇に話を戻して、まず Hellman 説の 2 条件のうち、当該の問題はヒエラルキー条件の問題であることを指摘する。次に、アルゴリズムがヒエラルキー条件をみたすかどうか理論的検討をおこなう。

差異処遇は、必ずしも人間が行うわけではない。バスの例でいえば、特定の人物が分かれて座るよう人間が直接言語をもって命令している場合ばかりではないだろう。他にも、間接的な命令にあたる事例として、法制度や政策などが考えられる。これらの共通点は、人間に指示内容を受け入れさせるような強制力を持ち、人々を一定の方角へ方向付けることである。この 2 つは Hellman の説の条件のうち、ヒエラルキー条件が取り扱うものである。

Hellman は、最もわかりやすいところでは、話者が権力を持っていることは強制性を構成する大きな内容になりうると記述している（Hellman 2008=2018: 91-94）。例えば、先述の上司が部下を侮辱する例がそれにあたるだろう。侮辱を通じて表現された部下の劣等性は、第三者にそれを受け入れさせる契機となりうる。次に、命令者が厳密には人間ではない事例であるが、法律や政策もその強制性が大きいと考えることができる。アパルトヘイトは、たとえ完全には法律が守られなかったとしても、人種によってエリアを分けるよう人々に命令する [12]。

こう考えるならば、人間でなくても強制力を持ちうることになる。このことがここでは重要である。われわれは、直接明示的に命令を下されなくとも一定の方向に向かって動くことがある。チェスやスポーツをするとき、われわれは反則をせず、なおかつ勝つように方向付けられる。このような仕組みを強制することができるならば、スポーツなどの自発的に何かをするという局面から脱して、われわれが否応なく営む日常生活のような局面でもその効力を発揮し得る。

では、アルゴリズムはそのような強制性をもつといえるだろうか。物質的な空間でも、例えば入場・出場がアルゴリズムによって管理されている場合にはあてはまるだろう。駅に設置されている自動改札機は、電子マネーの残高が十分であるかどうかによってふるまいを変え、残高が十分でない場合にはそのゲートを閉じる。この場合、一定の方向付けがプログラムによってなされているのだ。もし仮に、この基準が「十分に電子マネーの残高が残っているか」ではなくて人種であったとしたら、われわれはそれを差別的だと感じるはずであるし、当該の場合でも Hellman

12 次に問題になるのは、第三者が存在しない場合にも Hellman 説が成立するかということである（石田 2019: 4-5）。人間が行為している場合には成立しうるが、アルゴリズムが判断している場合に成立するか否かは議論の余地がある。

説は効力を発揮するだろう。われわれはそうと意識しない場合でも、日常生活を営む上で強制性にしたがっているのだ。ローンを貸すかどうかアルゴリズムだけが判断するなどの場合にも、同様の理由で強制性が発揮されているといえる。

無論、これだけでは（特に今回の事例において）アルゴリズムが強制性を持っていると述べるには十分ではない。自動改札機はアルゴリズムがわれわれを物質世界において従わせるもっともシンプルな例であるし、まして現実における多くの事例ではアルゴリズムだけが判断を下しているわけではない。より重要なのは、アルゴリズムが「真実」を産出し、われわれに当該の「真実」を信じさせることで、一定の方向付けをおこなうという機能の方である。

アルゴリズムは、いかに当該人物が扱われるかを決定するだけでなく、いかに扱われうるかという範囲を狭める役割を担っている（Beer 2017: 8-9）。一定の方向付けがどれほど、そしてどのように決定的となっているのかが強制力との関連では重要である。Beer は、真実の産出を行う方法について二種類を挙げている。

第一に、人間に選択肢を直接提示することによるものである（Beer 2017: 12-13）。アルゴリズムが人間の取れる行動範囲を提案するとしよう。結果、アルゴリズムのアウトプットは、人間を通して真実になっていくのだ。Google が私たちのウェブ上での行動をトラッキングし、次に買うもの、見るものの範囲をコントロールすることで、アルゴリズムが挙げた選択肢が後に現実に存在するようになってゆくのである。もちろん、人間にはレコメンデーションに従わない自由もあるが、本論で述べる強制力を有することがここで確認できるだろう。レコメンデーションとならぶ同様の方法として、ナッジを挙げることも可能である（cf. Yeung 2016）。

第二のものは、アルゴリズムが提示する規範に従わせる方法である（Beer 2017: 13）。Wii スポーツがわれわれの運動データを参照し、健康年齢を算出するとき、当該の健康年齢は無論実際の年齢ではなく、運動データと関連付けられたいわば架空の数値である（Chenny-Lippold 2017=2018: 150）。しかし、おそらく実際の年齢より若ければ喜ぶ人も多いだろう。この場合直接行動をサジェストされるわけではないが、それにもかかわらずわれわれの行動は、望ましい方へ、つまりここでは健康へ向かうように、一定の方向付けを得る。このはたらきは規範に自らを添わせようとする権力の働きの一種なのだ（Chenny-Lippold 2017=2018: 151-152）。

この場合、アルゴリズムが規範を提示するといっても、何もないところからそれが生み出されているわけではない。あくまで、現実に存在する価値と結びついて、アルゴリズムがその権力を行使しているのだ。そのような事情から、第二のアルゴリズムの方向付けの説明には、Foucault の知と権力の関係が参照されることもある（Rouvroy and Berns 2013; Introna 2016; Beer 2017 など）。Foucault は知と権力の関係について、権力は知識や、知識を運用する機構なくしては機能しないものと述べている（Foucault 2004: 34）。アルゴリズムは知を運用する機構となっているのであり、この場合の知とは、すなわちわれわれについての知識である（Chenny-Lippold 2017=2018: 55）。アルゴリズムは観念を具体化し、人間は割り当てられた観念、すなわち自分に対するアルゴリズムの解釈を認識することによって新たな行動を生み出すのだ[13]。こうしてアルゴリズムそのも

13　この観念を割り当てるはたらきがどう結果として表れるかは当然、アルゴリズムのアウトプットによって異なる。本人の認識を通じて本人の行動を変えることもあるが、より重要かつ潜在的な効果として、資源の配分を管理することがあることが指摘されている（Chenny-Lippold 2017=2018: 200）。

のが価値観を指し示すものとなるのである（Beer 2017: 14）。

Ziewitz（2016）が指摘するように、このようなアルゴリズムの働きは、決定的な影響力をもつものとして捉えられている。コンピュータの言語は、ただ世界に存在するにとどまらず、「世界を創る」ものでもあるのだ（Introna 2016: 12）。これらの記述は第一、第二の方法とも、アルゴリズムの強制力と、その強制力に従わざるを得ない構造下におかれる人間の状況を反映したものである[14]。アルゴリズムにおいては双方が結びついて、人間に命令を下さないまでも、やはりその影響下においているということができる。

かくして、こうして狭められた範囲・もしくは決定、そしてその狭められた範囲に反映されたアルゴリズムが提示する規範にも人間は従っているといえる。必ずしも決定や規範が実現されずとも、ここで見る限り十分に「一定の方向付け」をしており、したがって強制力をもっているといえるだろう。

次項では、COMPAS 事例に Hellman 説を適用し、その不正さを分析する。順に慣習条件、表現条件、ヒエラルキー条件を検討する。

2）COMPAS 事例の Hellman 説による批判的検討

2 節 3 項では、差別かどうかの条件が表現条件とヒエラルキー条件の二つから成り立っていることを説明し、なおかつ前項ではヒエラルキー条件をアルゴリズムがみたせる見込みがあることを説明した。したがって、Hellman 説においては、アルゴリズムの差別の不正さをその判断そのものに着目して指摘することができる可能性がある。

それでもなお、次のような反論があるかもしれない。それは、ヒエラルキー条件をアルゴリズムが一般的にみたせるとしても、アルゴリズムはやはり道徳的に不正な差別をおこなうことができないというものだ。なぜなら一つには、差別も一つの行為である以上、アルゴリズムが人間と同様に差別という行為をおこなえるわけではないから、という根拠が考えられる。確かにアルゴリズムは人間と全く同様の行為をおこなうことはできないが、この反論が成立するのは、アルゴリズムのみを「行為者」と考えた場合である。実際にはアルゴリズムと人間は多くの場合協同して判断しており、その場合には人間とアルゴリズムのエージェンシーが重なっている（mesh）場合の問題として捉えることができよう（Beer 2017: 7）。本論ではアルゴリズムに行為者性を認めず、アルゴリズムの影響力を主に考慮することを目的としている。そのため、この反論は、本論文の立場からして不十分な反論である。

以下では、この不十分さを具体的に示すために、いかにアルゴリズムは道徳的に不正な表現をおこない、その内容に人間を従わせることができるか、COMPAS 事例を用いて検討する。具体的には、Hellman が提示する条件のうち、慣習条件、表現条件、そしてヒエラルキー条件をいかにみたしているかを順に検討する。これらの条件を検討することで、アルゴリズムは道徳的に不正な差別をおこなえないという反論に対する再反論とする。そのためにはまず、これまで人種、とりわけ黒人という属性が犯罪の文脈でどのような意味を持ってきたか、そして人種プロファイリ

14　ただし強制力は実際に命令の結果が実現するかどうかとは区別される（Hellman 2008=2018: 51）。ここで重要なのはあくまで、アルゴリズムがその影響下に人間をおくことができるかどうかである。

ングと比較して、COMPAS は何がどう異なるのかを述べる必要があるだろう。最終的には本項は、COMPAS が表現しているのはどのような意味なのか、それは平等な人格的価値の毀損にあたるのかを確認する。

慣習条件

慣習条件とは、問題となっている行為が、歴史上において劣っていると烙印を押すような行為とどれほど類似しているかというものであった。COMPAS の事例がそのような行為と類似しているかを検討するために、以下では、犯罪という文脈で黒人という属性がいかなる意味を持っているのかを確認する。

黒人というステレオタイプは、長らく悪いイメージを持たれてきたのは周知の通りである。例えばアメリカでは、アフリカ系アメリカ人や有色民族と思われる人を、その民族的ルーツを根拠に捜査したり、身体検査や職務質問（stop and frisk）をかけたりする人種プロファイリング（racial profiling）が行われているといわれている。他の特徴に基づいたプロファイリングと比較して、人種が特有の表現内容を持つとして問題になる可能性があるのは、当該人種カテゴリーが、他の性別や年齢といった特徴ではありえないような重要性を持つという意味で、特有のネガティブさを伴っているからである（Hellman 2005: 236）。例えば、ここでいう黒人というカテゴリーは、単に黒人とされる人々を指し示すだけでなく、アメリカで黒人であるとはどういうことなのか、どういう人だと考えられるのかというアメリカの歴史に根ざした一種の文化的理解をも指し示すものだといえるのだ。

これは黒人といった人種だけでなく、9.11 テロ以後のアラブ人や、2021 年現在におけるアジア人も同様である。9.11 テロによってアラブ人という人種にはテロリストとしてのイメージが付き纏い、それ以降重点的に取り締まられるようになった（Burkeman 2002）。また、現在はコロナウイルスの流行により、アジア人も「危険」の烙印を押されている。この二つは、社会的状況との関連で「他の状況の場合にはない」「社会的重要性」が付与されてしまった例である。

この「社会的重要性」が平等な人格的価値を否定するかどうかによって、統計的一般化に基づいたプロファイリングが道徳的に問題になるかどうかが決まる。犯罪予測の文脈で言えば、犯人としての黒人は一つのステレオタイプであり、しかもそのことは人種プロファイリングによって侮辱的に行使され、黒人を貶価する（Hellman 2005: 237）。黒人が犯人である可能性が高い、もしくは今から犯罪を行う可能性が高いとするのは軽蔑的な意味を当人に付与し、平等な人格の尊重を行っていないといえる。

COMPAS と人種プロファイリングは、人種が問題になっているという点で共通点があるものの以下の点で異なる。第一に用途の違いが挙げられる。人種プロファイリングは主に警察が行うものであるため、それが活用されるのは捜査や逮捕、身体検査などのプロセスに限られる。一方で COMPAS は、量刑判断、リハビリ、仮釈放判断など、主に逮捕以後のプロセスにおいて使用される。COMPAS が量刑判断に使用されることが問題化されることが多いが、リハビリテーションに使用することはほとんど言及されない。このことはリハビリテーションに使用されることは問題がないが、量刑判断に使用することは問題含みであると示唆するものである（Barabas et al. 2018）。つまり、使用が許容される目的と、使用が許容されない目的があると考えられているのだ。

第二に、差別の量・質が変わる可能性があることである。

確かに、COMPAS と人種プロファイリングを比較するとき、考慮するのが人種だけでない COMPAS の方が人種プロファイリングよりも多角的に当該人物を評価しているといえるかもしれない。しかしここで着目すべき点は、当該の統計的一般化が侮辱的であるかどうかであり、多角的に判断しているかどうかには関係がない。多角的に判断していたとしても、差別は差別であるというわけである。COMPAS はもともと人種を変数として考慮するソフトウェアではなく、その質問紙の内容からして本人の経歴や社会的環境に目を向けている（Propublica 2016）。人種だけによって犯罪者かどうかを判断されるよりはまだ良いという反論もありうるが、考慮する項目を増やしたとしても、その項目に侮辱的な内容が含まれているならば、それもやはり問題だといえるのではないだろうか。

念を押しておきたいのは、ここでの「侮辱的」とは、意図とは関係なく、貶価するようなものをさすということである。侮辱する意図なしに他者を侮辱することは可能である。特にこの場合、「科学的な隠れ蓑」を使っていて、意図説で処理しようとするのは賢明でない。例えば女性の勤続年数が短い傾向にあるから女性の雇控えをする場合、差別者の意図が侮辱的であるから不正であるわけではない。質問紙を作った人も、おそらくは COMPAS が最善の働きをするように作っていることが想定され、必ずしも特定の人種を侮辱する意図はないだろう。問題は、人種プロファイリングと同じように、貶価する点にあるといえるのだ。

しかしながら相違点としては、COMPAS のようなアルゴリズムの導入によって、差別が増幅されたり、差別に新たな、もしくはより強い「根拠」が付加されることが考えられる。しかもその「根拠」は、これまで使用されていたものよりも強固かつ説得的でありうる。貶価する点で人種プロファイリングと大まかな構造は同一かもしれないが、その力は大いに異なっているだろう。その力というのは、例えば裁判官などといった、しばしば実在の人間を通して行使される。

第三に、人種変数を考慮しているか否かの違いが挙げられる。COMPAS では検証の結果人種との関連性が認められており、しかも当該人種パラメータのうち、統計的に優位であったのは通常犯罪における再犯では黒人、ヒスパニック、ネイティブアメリカンの三種類の「人種」であったが、暴力犯罪の再犯では黒人だけであった（Larson et al. 2016）。

これらの違いは、COMPAS が直接表現している内容そのものに影響を与えるものではない。COMPAS は直接人種を考慮しておらず、したがって人種ゆえに誰かが再犯をするだろう、と「告げて」いるわけではないが、むしろ別の形で侮辱的に働いているといえるだろう。暗黙裡に人種が考慮されているために、当該の人種と犯罪がいかに結びついてきたかという慣習条件も引き継がれているといえる。

表現条件

これまで、犯罪予測において人種がどのような意味を持っているかを記述してきた。これは Hellman 説における慣習条件にあたるものである。COMPAS はそのような慣習を引き継いだ上で、いかなる表現を行っているといえるのだろうか。アルゴリズムによる表現は、基本的にはいかにわれわれの目に当該の属性が意味づけされて見えるかという観点から表すことができる。

特定の人物の再犯リスク判定を提示するということそのものの意味をまず考えよう。当該リスク判断は、リスクレベルを従属変数として、質問紙で尋ねられる以下の説明変数が設定されている。本人の犯罪歴、法律不遵守、被調査者の家族の犯罪、友人関係、住居、社会環境、教育、職業、

趣味、社会的孤立、犯罪的人格、怒り、犯罪への態度の合計 13 セクション、合計質問数は 137 問である（Propublica 2016）。

この判断は生まれもった環境を考慮に入れている。これをリスク判定に持ち込むことは裁判所のルールに反する。裁判で重要なのは、「その人が誰であるか」ではなく、「その人が何をしたか」であるからだ。当該リスク判断は、量刑判断が行われる法廷で使用する場合には、「その人が誰であるか」を不当に裁判に持ち込むものであるといえる。もちろんリスク判断ではその人が何をしたかも考慮されるが、これが未来の判断にも適用されることによって、「その人が何をしたか」よりも「その人が誰であるか」の比重が大きくなっているといえるのだ。彼らはこれまで危険だった人間と共通の属性を持つがゆえにこれからも危険なのである。したがって、仮に人種との関連性がなくとも、この表現は道徳的に問題になりうるといえる。

もちろん、当該表現は人種との関連でより一層問題となっている。これが問題となるのは、量刑判断の文脈で言えば裁判所が、社会的カテゴリーとしての人種ごとの差異を認めるような印象を与えるからだ。認められた差異が量刑判断やリハビリテーションなどの個々の文脈に持ち込まれるため、裁判所以外でも「その人が誰であるか」に比重が移るのである。

白人であれば再犯する確率がより低く、黒人であれば再犯する確率が高いとみなすことは、「何をしたか」ではなく「誰であるか」に考慮の比重が移ることによって、白人と黒人の評価基準が結果的に変化してしまう。誰であるかを非対称に評価することは、その非対称性はここでみてきたように、平等な人格的価値を毀損するような仕方で、である。よって、COMPAS は表現条件をみたすと考えられる。そして、もちろん結果として、黒人のステレオタイプを強化すると考えることができる。

ヒエラルキー条件

以上、本節では Hellman 説における道徳的に不正な差別の条件である慣習条件と表現条件について検討してきた。最後に COMPAS の事例におけるヒエラルキー条件について検討する。判断を下すアルゴリズムは誰のどのような立場を代理しているのだろうか。

量刑判断において裁判官の判断を補助するために COMPAS が使用されるということは、当該ソフトウェアがいわば「国のお墨付きを得た」ソフトであることを意味する。この判断が法廷に持ち込まれる場合と、リハビリテーションプログラムの決定のために持ち込まれる場合では、当該判断の役割は異なっているだろう。法廷は COMPAS が、個人を尊重した形で量刑判断を行い因果的責任を持つことを保証している（Rubel et al. 2019: 16）。したがって COMPAS は、部分的には裁判官、ひいては国家を代理して判断するといえる。

当該リスク判断は当然、裁判官の判断に影響を及ぼすことが望まれているが、問題は裁判官の判断に悪影響を与える場合である。ここでいう悪影響というのは、差別的な判断を行うようアルゴリズムが仕向ける場合を指す。

Hellman 説の先述の検討からすれば、代理していても Hellman 説は満たせるといえるだろう。その上、裁判官はプログラムを精査する技量や能力に欠けており、算出されたリスク判断を目にしてなお、公正に判断することは不可能である（Harvard Law Review 2016; Liu et al. 2019）。この点からして、アルゴリズムによる影響は悪影響である可能性が高く、しかも修正が不可能であ

る[15]。以上より、COMPAS はヒエラルキー条件をみたしている。

むすび

本論文では、以下の作業を行った。2 節で COMPAS の特殊性とこれまでなされてきた説明が不十分であり、差別の問題として依然検討に値するものであることを示した。3 節では、分析の道具として Hellman 説が適切であることを論じ、Hellman 説の詳細について説明した。4 節では、Hellman 説の理論的検討を行い、そのうちの条件を補った。次に、当初の想定よりもその適用範囲を拡張した Hellman 説を用いて、COMPAS の道徳的不正さを説明した。

この作業によって得られた結論は以下のとおりである。COMPAS がおこなっているのは道徳的に不正な差別である。その不正さの根拠は、Hellman 説が基盤とする個人の尊厳の不尊重による。また、アルゴリズムによる不尊重は個人の尊厳に反するような表現を提示することによって行われており、したがって Hellman 説が規定する道徳的に不正な差別であることを示した。

一方で課題も残されている。今回立証したのはアルゴリズムによる差別的な表現の不正さであり、人間がどのように判断を下しているかを考慮していない。アルゴリズムだけが判断を下し、実際の対応につながる場合には[16]、このような議論が成り立つかもしれない。しかしながら、現状の一般的なアルゴリズムを取り入れた判断が完全に自動化されているとは考えにくい。そのうえ、自動化されたプロセスだけに基づいた決定に人間が従わない権利を有する旨が記述されている GDPR22 条の文言を考えても、今後中心的な問題となる可能性は低い。

むしろ必要なのは、人間とアルゴリズム双方が協力して判断を下す場合にいかに意思決定が行われるかという視点であるといえよう。とりわけ機械学習の分野では、誤った意思決定を防ぎ、意思決定の結果をより公平にするために、アルゴリズムだけに判断させるのではなく、判断の過程に人間も入れる人間参加型（human-in-the-loop）機械学習という取り組みが着目されている。今回用いた事例である COMPAS は機械学習を使用していない統計的プログラムであるが、自動的に判断を下すプログラムであるという共通点から、どのように人間がアルゴリズムに補助されて意思決定を行っているかに今後着目することは有益であると考えられる。

謝　辞

この論文の執筆にあたり、東京大学大学院教授佐倉統先生、同大学院の水上拓哉氏にはさまざまな助言をいただいた。また、この論文の初期のバージョンは修士論文の形で公開されている。修士論文執筆にあたっては上記のお二人に加え、東京大学大学院の戸田聡一郎助教、東京大学大学院所属の石田柊氏に貴重なコメントをいただいた。感謝の意を表する。

15　本稿の射程を超えるが、この点はさらなる論争を喚起するだろう。というのは、これまでの法廷での判決がすべて公正だったとは言えない可能性が高く、比較すればやはりアルゴリズムの判断のほうが公正なのだという結論もありうるからである。また、バイアスの度合いに仮にそれほど違いがなかったとしても、修正可能性は人間よりもアルゴリズムの方が高いという可能性も否定できない。

16　例えばアルゴリズムがローンの審査を行い、貸与の可否までアルゴリズムが一貫して決定を行う場合などが想定される。

引用文献

Alexender, Larry, 1992, “What Makes Wrongful Discrimination Wrong?: Biases, Preferences, Stereotypes, and Proxies”, *University of Pensilvania Review*, 141: 149-219.

Angwin, Julia, Jeff Larson, Surya Mattu and Lauren Kirchner, 2016, “*Machine Bias*”, (Retrieved December 18, 2020, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing).

Arrow, Kenneth J., 1971, *The Theory of Discrimination*, Prinston University.

Ayres, Ian, 2002, “Outcome Tests of Racial Disparities in Police Practices”, *Justice Research and Policy*, 4: 131-42.

Barabas, Chelsea, Karthik Dinakar, Joichi Ito, Madars Virza and Jonathan Zittrain, 2018, “Intervention over Predictions: Reframing the Ethical Debate for Actual Risk Assessment”, *Proceedings of Machine Learning Research*, 81: 1-15.

Beer, David, 2017, “The social power of algorithms”, *Information, Communication & Society*, 20 (1): 1-13.

Beriain, Iñigo De Miguel, 2018, “Does the use of risk assessments in sentences respect the right to due process? A critical analysis of the Wisconsin v. Loomis ruling”, *Law, Probability and Risk*, 17 (1): 45-53.

Binns, Rauben, 2018, “Fareness in Machine Learning: Lessons from Political Philosophy”, in *Journal of Machine Learning Research*, 81: 1-11.

Brennan, Tim, William Dieterich and Beate Ehret, 2009, “Evaluating the predictive validity of the COMPAS risk and needs assessment system”, *Criminal Justice and Behavior*, 36 (1): 21-40.

Burkeman, Oliver, 2002, “Visa detainees allege beatings”, The Guardian, (Retrieved January 1, 2020, https://www. theguardian. com/world/2002/may/24/usa. afghanistan).

Chenny-Lippold, John, 2017, *We Are Data*：*Algorithms and the Making of Our Digital Selves*, New York: New York University Press.（= 2018, 高取芳彦訳,『We are Data――アルゴリズムが「私」を決める』日経 BP 社.）

Corbett-Davies, Sam and Sharad Goel, 2018, “The measure and mismeasure of fairness: A critical review of fair machine learning”, *ArXiv*, 1808.00023v2[cs.CY], (Retrieved December 18, 2020, https://arxiv.org/abs/1808.00023).

Dieterich, William, Christina Mendoza and Tim Brennan, 2016, “*COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity*”, (Retrieved December 18, 2020, http://go.volarisgroup.com/rs/430-MBX-989/images/ProPublica_Commentary_Final_070616.pdf).

Dressel, Julia and Hany Farid, 2018, “The Accuracy, Fairness, and Limits of Predicting Recidivism”, *Science Advances*, 4 (1): 1-5.

Eidelson, Benjamin, 2015, *Discrimination and Disrespect*, Oxford University Press.

Foucault, Michele, 2004, *Society must be defended: Lectures at the Collège de France, 1975-76*, London: Penguin.

Freeman, Katherine, 2016, “Algorithmic Injustice: How the Wisconsin Supreme Court Failed to Protect Due Process Rights in State v. Loomis”, *North Calorina Journal of Law& Technology*, 18 (5): 75-106.

Goel, Sharad, Ravi Shroff, Jennifer L. Skeem and Christopher Slobogin, 2018, “The Accuracy, Equity, and Jurisprudence of Criminal Risk Assessment”, *SSRN Electronic Journal*, 1-21.

Harvard Law Review, 2017, “Criminal Law — Sentencing Guideline — Wisconsin Supreme Court Requires Warning before Use of Algorithmic Risk Assessments in Sentencing. — State v. Loomis, 881 N.W.2d 749 (Wis. 2016).”, *Harvard Law Review*, 130: 1530-7.

Hellman, Deborah, 2005, “Racial Profiling and the Meaning of Racial Categories”, Andrew I. Cohen and Christpher H. Wellman eds., *Contemporary Debates in Applied Ethics*, Hoboken: Wiley-Blackwell, 232-44.

――― , 2008, *When is Discriminarion Wrong?*, Cambridge: Harvard University Press.

（= 2018, 池田喬・堀田義太郎訳, 『差別はいつ悪質になるのか』法政大学出版局. ）

Hing, Julianne, 2009, *HP Face-Tracker Software Can't See Black People*, (Retrieved December 18, 2020, https://www.colorlines.com/articles/hp-face-tracker-software-cant-see-black-people).

Introna, Lucas D., 2016, “Algorithms, Governance, and Governmentality: On Governing Academic Writing”, *Science, Technology & Human Values*, 41 (1): 17-49.

石田柊, 2019,「差別と危害 : 帰結主義的差別論の擁護」『社会と倫理』34: 1-12.

Khaitan, Tarunabh, 2018, “Indirect Discrimination”, in *The Routledge Handbook of the Ethics of Discrimination*, London: Routledge, 30-41.

Kleinberg, John, Sendhil Mullainahan and Manish Raghavan, 2017, “Inherent Trade-offs in the Fair Determination of Risk Scores”, *Computer Science, Mathematics*, 67:1-23.

Knight, Cirl, 2013, “The Injustice of Discrimination”, *South African Journal of Philosophy*, 32 (1): 47-59.

Koene, Ansgar, Liz Dorthwaite and Suchana Seth, 2018, “IEEE P7003™ standard for algorithmic bias considerations”, *IEEE*, (Retrieved December 18, 2020, http://fairware.cs.umass.edu/papers/Koene.pdf).

Larson, Jeff, Surya Mattu, Lauren Kirchner and Julia Angwin, 2016, *How We Analysed the COMPAS Redicisism Algorithm*, (Retrieved December 18, 2020, https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm).

Lippert-Rasmussen, Kasper, 2006, “The Badness of Discrimination”, *Ethical Theory and Moral Practice*, 9 (2): 167-185.

Liu, Han-Wei, Ching-Fu Lin and Yu-Jie Chen, 2019, “Beyond State v Loomis: artificial intelligence, government algorithmization and accountability”, *International Journal of Law and Information Technology*, 27 (2): 122-41.

Noble, Safiya U., 2018, *Algorithms of Oppression*, New York: New York University Press.

O'Neil, Cathy, 2016, *Weapons of Math Destruction: How Big Data Increases Inequality and Threaten Democracy*, Phoenix: Crown.

Phelps, Edmund S., 1972, “The Statistical Theory of Racism and Sexism”, *American Economic Review*, 62 (4): 659-61.

Propublica, 2016, “Sample-COMPAS-Risk-Assessment-COMPAS-“CORE”,” documentcloud, (Retrieved December 18, 2020, https://www. documentcloud. org/documents/2702103-Sample-Risk-Assessment-COMPAS-CORE. html).

Rouvroy, Antoinette and Thomas Berns, 2013, *Algorithmic Governmentality and Prospects of Emancipation: Disparateness as a Precognition for Individuation through Relationships?*, 177 (1): 163-96.

Rubel, Alan, Clinton Castro and Adam Pham, 2018, “Algorithm, Bias, and the Importance of Agency”, Jo

Bates, Paul D. Clough, Robert Jäschke, and Jahna Otterbacher eds., *Proceedings of the International Workshop on Bias in Information, Algorithms, and Systems*, 9-13.

Sandvig, Christian, Kevin Hamiton, Karrie Karahalios and Cedric Langbort, 2016, "When the Algorithm Itself is a Racist: Diagnosing Ethical Harm in the Basic Components of Software", *International Journal of Communication*, 10: 4972-90.

Scanlon, Thomas M., 2008, *Moral Dimensions: Permissibility, Meaning, Blame*, Cambridge: Belknap Press.

Schauer, Frederick, 2003, *Profiles, Probabilities, and Stereptypes*, Cambridge: Harvard University Press.

State *v.* Loomis, 881 N.W.2d 749, 767（Wis. 2016）.

Sumpter, David, 2018, *Outnumbered: from Facebook and Google to Fake News and Filter-Bubbles: The Algorithms That Control Our Lives*, London: Bloomsbury.（= 2019, 千葉敏生・橋本篤史訳,『アルゴリズムはどれほど人を支配しているのか? ── あなたを分析し、操作するブラックボックスの真実』光文社.）

Sweeney, Latanya, 2013, "Discrimination in Online Ad Delivery", *ArXiv:1301.6822 [Cs],*（Retrieved December 18, 2020, http://arxiv.org/abs/1301.6822）.

Washington, Anne L, 2019, "How to Argue with an Algorithm: Lessons FROM THE COMPAS-PROPUBLICA DEBATE", *The Colorado Technology Law Journal*, 17 (1): 1-37.

山本龍彦・尾崎愛美, 2018,「アルゴリズムと公正 ── State v. Loomis 判決を素材に」『科学技術社会論研究』16: 96-107.

Yeung, Karen, 2017, " 'Hypernudge': Big Data as a Mode of Regulation by Design", *Information, Communication& Society*, 20 (1): 1-19.

Ziewitz, Malte, 2016, "Governing Algorhthms: Myth, Mess, and Methods", *Science, Technology, & Human Values*, 41 (1): 3-16.

企業の道徳的行為者性をめぐる企業の意図の問題 ── 推論主義に基づく検討

西本優樹（北海道大学大学院文学研究科）

要旨

本稿では、ビジネス倫理で「企業の道徳的行為者性」（corporate moral agency）をめぐって中心的な論点となる企業の意図の問題を、推論主義（Brandom 1994）と呼ばれる言語行為論の一類型を援用して検討する。

企業に行為の意図を認めることができるかという問題は、従来から心の哲学の心理主義と機能主義の対立を反映する形で議論されてきた。すなわち、意図に関して心理主義を支持するレンネガードとヴェラスキーズ（2017）が、心を持たない企業が意図を持つことはありえないと主張するのに対し、機能主義を支持する論者は、企業に意図の機能的特徴を見出すことができると主張する。

本稿では、この対立を概観した後、意図に関して言語論的な機能主義を支持する推論主義を援用することで、レンネガード、ヴェラスキーズの議論に反論を提起し、この議論で推論主義が適切であることを示す。この作業の後、本稿では、条件付きではあるが推論主義から企業の意図および企業の道徳的行為者性が正当化できることを示し、そうした議論から帰結する問題点を指摘する。

Abstract

This paper shows that the problem of corporate intentions, discussed in the issue of the “corporate moral agency” in business ethics, can be successfully explained from inferentialism (Brandom 1994), an argument from the philosophy of language. The question of whether corporations have intentions has traditionally evolved in a way that reflects the debate between psychologism and functionalism in the philosophy of mind. That is, while Rönnegard and Velasquez (2017) who support psychologism argue that corporations cannot have intentions without minds, proponents of functionalism argue that functional features of intentions can be found in corporate activities. After reviewing this debate, this paper raises a counterargument to Rönnegard and Velasquez by supporting inferentialism that is linguistic functionalism concerning intentions, and shows that inferentialism is appropriate. After this discussion, this paper shows that corporate intentions and corporate moral agency can be justified from inferentialism, albeit conditionally, and points out the problems that result from such arguments.

はじめに

企業が社会的・道徳的責任を問われる存在であることは、一般に受け入れられた見解である。しかし、企業がそれらの責任を負うことができると考える哲学的な妥当性に関しては、現在も議論が続いている。ビジネス倫理の「企業の道徳的行為者性」（corporate moral agency）は、そのような問題を扱う議論である。この議論では、企業それ自体が道徳的責任を負うことのできる行為者かを問うことで、企業不祥事や事故の道徳的責任のあり方を検討する。企業が道徳的行為者だといえる場合、企業内の個人に責任を帰属しきれない場面や、企業内の誰にも責任を帰属できない場面で、企業それ自体に責任を帰属させることが正当化される[1]。その場合、企業内の個人だけでなく企業それ自体もまた、非難や処罰の対象になる。反対に、企業が道徳的行為者といえない場合、責任はあくまで関係する個人の問題となる。この議論の帰趨は、企業の処遇に関する法学上の議論の基礎になるとも考えられている[2]。

この議論で中心となる問題の一つは、企業活動を意図する主体は誰かという点である[3]。道徳的責任は、一般に出来事や行為を意図的に引き起こしたことに基づいて帰属される[4]。基本的に、意図は行為者の心的状態として理解されるため、心のない企業が意図を持つことは不可能に見える。しかし、企業の道徳的行為者性の擁護者は、例えばフレンチの「何かが意図的に行為すると述べることは、それが行動を動機づける目的や計画、ゴール、関心を持つと述べること」であり「その状態［意図］の主要な要素である計画は、企業の意思決定に典型的に見られる」（French 1995:10-12）のように、意図の機能的特徴に注目することで、企業に意図を認めることを主張してきた。

確かに、そうした仕方で企業に意図を認めることが可能なら、複雑化する企業活動の実態に即した責任の帰属が期待できるかもしれない。しかし、企業が心的な意図を持つことができないという点は、解決済みの問題というわけではない。レンネガードやヴェラスキーズら企業の道徳的行為者性の批判者は、意図とは心的状態のことであり、企業がそれを持つことはあり得ないと繰り返し主張する（Rönnegard 2013, 2015; Velasquez 1983, 2003）[5]。これに対し、企業の道徳的行為者性の擁護者は、例えば、アーノルドの「もし［意図に関する］直観が別の場所にあるなら、もし志向性[6]を示す存在のクラスが心的状態を持つものに限定されないという可能性を受け入れるの

1　直接不正を犯した社員だけに責任を問う場合、いわゆる「トカゲのしっぽ切り」が生じて企業全体の責任を検討することが困難になる（杉本 2008:43）。また、個人への過度の責任追及は問題の構造を不明確にし、再発防止を妨げることになる（Sanders 1993:74）。この議論で取りあげられる事例は多岐にわたる。例えば杉本（2019）は、企業それ自体を処罰すべきと指摘のあった事例として、2005 年の JR 福知山線の脱線事故をあげる。他に、企業内の誰にも責任を問えない事例としては、チャレンジャー号爆発事故（Sanders 1993）や NZ 航空エレバス山墜落事故（Phillips 1995）などがあげられる。

2　例えば、米国の連邦最高裁で出された、企業に政治的表現の自由を認める判決（Citizens United v Federal Election Commission 558 U.S. 310. 2010）や宗教の自由を認める判決（Burwell v Hobby Lobby Stores, Inc. 134 S Ct 2751. 2014）を受けて、企業の道徳的行為者性との関係が議論されている（Friedman 2020; Hussain and Sandberg 2017）。

3　本稿で主体が意図（intention）を持つという場合、それが次の二種類の意図を持ち得ることを指す。すなわち、行為主体が行為の遂行中に持つ意図と、行為に先立って持つ意図である。以下で見る議論では、前者はブラットマンの現在指向的意図、サールの行為内意図に相当する。後者は、ブラットマンの未来指向的意図、サールの先行意図に相当する。

4　周知の通り、道徳的責任は伝統的に自由意志と決定論の問題として議論される。企業の自由意志に関しては、これを擁護したヘス（Hess 2013）の議論がある。

5　同様の指摘は、ダンリー（Danley 1980）、ドナルドソン（Donaldson 1980）にも見られる。

6　志向性（intentionality）は、心的状態の「ついて性」（aboutness）を特徴づける概念である。例えば、誰かが「水を飲みたい」という欲求を持つ場合、その欲求は水についてのものである。一般に、信念や欲求、意図など、多く

を厭わないなら、企業の志向性の議論が提供されるだろう」（Arnold 2006:284）のように、意図を心的状態から説明する立場を否定することなく、それを他の仕方でも説明できると主張するに留まっている。そのため、レンネガードやヴェラスキーズの批判はなお有効に見えるし、実際に二人はそう考えている。

この批判に正面から応じるには、まず意図は心的状態だという立場を批判しなくてはならない。意図を機能的特徴から説明するのは、その後の作業である。そうした議論は、意図に関して心理主義を主張するレンネガードやヴェラスキーズに対してだけでなく、個人の意図に関して心理主義を許容しながら、企業の意図に関して機能主義[7]を主張する論者に対しても、批判を投げかけることになる[8]。言い換えれば、その議論は、個人と企業のどちらの意図についても、心理主義を否定した上で機能主義を支持する議論ということである[9]。本稿では、このように個人と企業の意図が機能的側面から等価なものとして扱われる場合に、企業の道徳的行為者性が成立すると主張する。この作業を通じて、機能主義から企業の道徳的行為者性が正当化される事態の内実を明らかにすることが、本稿の主たる狙いである[10]。

本稿の議論は以下のように進む。まず、意図に関して心理主義を主張するレンネガードとヴェラスキーズの議論を、二人が援用するサール（Searle 1992, 1995, 1997, 1998, 2010）の心の哲学および社会存在論に言及しながら概観する（第一節）。次に、意図に関して言語論的な機能主義を支持するブランダム（Brandom 1994）の推論主義を取りあげ、企業活動の道徳的責任が問題の場合、意図に関する心理主義が妥当性を欠き、推論主義が支持されることを示す（第二節）。その後に、推論主義を採用する場合、一定の条件下で企業の道徳的行為者性の正当化が可能であること、しかしその条件下で、いくつかの検討すべき課題が残ることを指摘する（第三節）。

1. レンネガードとヴェラスキーズの企業の道徳的行為者性に対する批判

本節では、レンネガードとヴェラスキーズによる企業の意図の理解、およびそれを通じた企業の道徳的行為者性への批判を概観する。ヴェラスキーズは、企業の道徳的行為者性をめぐる論争の当初から今日まで、これを否定し続けてきた論者である（Velasquez 1983, 2003）。レンネガードは、2010年代から、ヴェラスキーズの議論に新たな論点を加えた批判を提起したことで注目を集める論者である（Rönnegard 2013, 2015）。二人は、企業の道徳的行為者性をめぐる2017年の論集

の心的状態は志向性を持つと考えられる。ただし、漠然とした不安のように、全ての心的状態が志向的ではないとする見解もある（Searle 1983:1-2）。

7　機能主義は、信念や意図など特定のタイプの心的状態を、それが機能する仕方、あるいはそれが果たす役割から説明する心の哲学の立場である。機能主義は行動主義と異なり、（コンピュータのプログラムのような）対象の内部状態も考慮するが、本稿ではより広く、内部状態を考慮しないデネットの志向システム論（Dennet 1987）のような議論も機能主義として扱う。志向システム論については、脚注18を参照。

8　例えば、個人の意図は心的状態だが、企業の意図はそれと異なる機能的状態だと考えるワーヘイン（Werhane 1985）の議論は、その一例である。

9　機能主義は、脚注7の通り特定のタイプの心的状態を説明する立場であり、基本的に心理主義を含意する。これは、心の多型実現、例えば機能的特徴からコンピュータのプログラムも心的状態を実現できると主張する場合も同様である。他方で、本稿で提示しようとするのは、これらと異なり心理主義を含意しない機能主義である。第二節でそうした立場を検討する。

10　この作業では、企業の道徳的行為者性に向けられる一つの懸念を念頭に置いている。すなわち、企業を道徳的行為者だと認めることで、企業の非倫理的行為がかえって促進されるという懸念である（Ashman and Winstanley 2007; 杉本 2019）。

（Orts and Smith 2017）で、企業の道徳的行為者性に対する批判をまとめた共著論文（Rönnegard and Velasquez 2017）を発表している。意図の問題はそこで提示される論点の一つである[11]。他方で、道徳的行為者であるために心的状態が必要だと考える二人の姿勢は、他の論点でも共通している。以下では、意図の理解を中心に、道徳的行為者であるために心的状態が必要だとする二人の議論を概観する。

企業の意図の問題

まず、前提となる道徳的責任を確認しておこう。ここでは、ヴェラスキーズの定義を参照する（Velasquez 2003:532）[12]。企業の道徳的行為者性で問題になるのは、誰かあるいは何かがすでに引き起こした行為や出来事に向けられる責任である。ヴェラスキーズは、この責任を「因果的責任」と呼ぶ。因果的責任には二つの区別がある。一つ目は、例えばハリケーンが街を破壊したというような、自然の行為者に帰属される責任である。二つ目は、意図的な行為者（intentional agents）に帰属される責任である。問題となる道徳的責任は後者、すなわち「人間のような意図的行為者が何らかの過去の出来事を引き起こし（あるいは引き起こすのを助け）、そしてそれを意図的に行った場合に、我々が彼らに帰属する種類の因果的責任」（ibid.）である。

この定義に従えば、企業の道徳的行為者性を主張するためには、企業がそれ自体で行為を引き起こすこと、そして行為の意図を持つことの二点を示す必要がある[13]。本稿で焦点を当てるのは後者である。企業が行為を引き起こすことができるかについては、企業の意図に関する議論を経た後に補足的に言及することとする。

意図に関する機能主義

次に、企業の道徳的行為者性の擁護者による、企業に意図を認めることができるという主張を確認しよう。本稿ではフレンチの議論をとりあげる。レンネガードとヴェラスキーズが主に批判の対象としてきたのが、フレンチの議論だからである。次の例から考えよう（French 1995:12）。ある日、フレンチの家に、ニューヨーク州バッファローにある自動車会社Mのクレジット部門から通知が届いた。通知によれば、フレンチはM社からリースした車に対する前年分の財産税の一部を滞納している。通知にはSという名前のサインがあり、滞納分を納めるかSに連絡するよう記してある。フレンチは通知が誤りだと思ったので、記載された番号に電話をかけた。Sが電話に出てフレンチの質問に答えた。Sによれば、M社には、フレンチの滞納を示す記録が確かにある。フレンチはその点を何度も確認したが、Sの答えは変わらない。二人はやり取りの末に、問題を解決する合意できる結論に達した。

フレンチによれば、この事例で督促の意図を持つのは、SでなくM社である。この議論は、ブ

11　レンネガードとヴェラスキーズが提示する論点は次の六つである。(1) 企業の道徳的行為者性は反直観的である、(2) 企業は心的な意図が持てない、(3) 企業は行為を引き起こせない、(4) 企業は感情を持てない、(5) 企業に責任を認めると不公平な責任帰属が生じる、(6) 企業は自律性を持てない。(4) は本節、(3) は第三節、(6) は脚注 24 で触れる。

12　ムーアは、1999 年までの論争を総括した論文で、ヴェラスキーズの定義を、企業の道徳的行為者性をめぐる議論で一般に受け入れられたものと紹介している（Moore 1999:300）。このことは、現在の議論でも同様と思われる。

13　本稿では、基本的に道徳的行為者性を意図的行為者であることとして議論を進める。意図的行為者であることに加えて要請される条件については、第三節で言及する。

ラットマン（Bratman [1987] 1999）の意図の計画理論を援用することで正当化される。ブラットマンによれば、意図の特徴は、例えば、「明日の飛行機で新千歳空港に飛ぶ」のように、未来指向的に形成される点にある（ibid. 8）。この点で、未来指向的意図は計画の一部である。つまり、未来指向的意図は、人間が計画を形成し、それを保持し、場合によって結合したり、修正したりする中で本質的な役割を果たす。この理解は、意図に関する伝統的な議論と対照をなす。伝統的な議論は、意図的行為の遂行中に見られる、現在指向的意図に焦点を当てる。その場合、意図はその時点で行為者が持つ信念と欲求の組み合わせと同一視されるため、計画に関わる未来的指向的な要素は考慮されないことになる（ibid. 6-9）。しかし、そうした意図の理解は犬や猫にこそ適切かもしれないが、計画する生き物としての人間にとってはそうではない。ブラットマンの目的は、意図を信念と欲求の組み合わせと考える信念欲求モデルに替えて、意図の計画理論を提示することである。

フレンチは、この議論が企業の意図を正当化すると考える（French 1995:10-27）。もし意図が信念欲求モデルから説明されるなら、心を持たない企業は信じたり欲求したりすることができないため、意図を持つことは不可能である。しかし、意図が計画や計画することに関わる機能的側面から特徴づけられるなら、企業にそれを見出すことは可能である。つまり、意図に関して機能主義を採用するなら、「その状態［意図］の主要な要素である計画は、企業の意思決定に典型的に見られる」（ibid. 12）のである[14]。

このような仕方で企業の意図を説明する場合、企業の意図と企業内の個人のそれとの区別が問題になる。というのも、M 社の督促は、M 社の意図的行為としても、例えば、「S がフレンチ宛の督促にサインした」のような、S の意図的行為としても記述できるからである（French 1995:23）。このとき、督促が M 社の意図的行為であることを確証するのは、全ての企業が持つ意志決定構造である（ibid. 25-26）。フレンチはこれを、企業内決定構造（Corporation's Inner Decision Structure，CIDS）と呼ぶ。

CIDS は二つの要素から構成される。それは、企業内の地位とレベルを記した組織フローチャートと、企業内の決定や行為が企業のものであることを承認する規則（通常は企業憲章を含む）である。組織フローチャートは、企業の決定がどのレベルで、誰によってなされるべきかを定める。さらに、特定の誰かによってなされた決定や行為は承認規則を参照することで、企業の決定や行為とされる。すなわち、企業内の誰かによる決定や行為を CIDS から見るならば、それは企業の意図や意図的行為として適切に記述される。M 社の督促についていえば、CIDS に照らして適切なものである限り、それは M 社の意図的行為として理解される。

意図に関する心理主義

このように、フレンチは意図に関する機能主義と CIDS を組み合わせることで、企業に意図を認め

14　フレンチは、このように心的状態なしに意図を持つことは可能だと議論するが、ブラットマン自身は、意図を心的状態として議論を進めており、ブラットマンの議論からフレンチの議論が導かれるかは疑問が残る（Bratman [1987] 1999:10）。他方で、ブラットマンは近年、自ら企業の道徳的行為者性をめぐる議論に参入し、特定の共有された手続きを経てなされた集団の決定は、心的状態でなく、また構成員の共有意図（Bratman 2014）とも異なる集団の意図だと主張している。この議論でブラットマンは、集団の意図が心的状態でないことについて、フレンチよりも詳細な検討を行っている（Bratman 2017）。

ることを主張する。企業に意図を認める他の論者も、何らかの仕方で、意図に関する機能主義を支持している（Rönnegard and Velasquez 2017:128）[15]。しかし、レンネガードとヴェラスキーズによれば、意図に関する機能主義は誤りである（Rönnegard and Velasquez 2017:134-37; Velasquez 2003:558 n.40）[16]。端的に言えば、二人の議論は、企業の道徳的行為者性で問題になる意図とは人間の持つ心的な意図のことであり、企業にそれを持つことはできないと主張するものである。二人によれば、意図に関する機能主義は、機能的側面から理解した意図を企業に帰属できることをもって、企業に意図を認めるという議論である。その際の意図の帰属の仕方には次の二通りがあるが、どちらも企業に心的な意図があることを示すものではない[17]。したがって企業の道徳的行為者性は誤りということになる。

企業に意図を帰属する仕方の一つ目は、我々が企業に意図があると比喩的に語る場合や、それによって企業の振る舞いを予測する場合に行う帰属である。M 社の督促でいえば、M 社に督促の意図があると語ったり、それによって M 社の振る舞いを予測したりする場合がそうである[18]。しかし、我々がそのように企業に意図を帰属させるとしても、企業に人間と同様の意図があるわけではない。この批判は、サール（Searle 1992）の志向性論を援用する（Rönnegard 2013:86-76; 2015:24-25; Velasquez 2003:546）。サールは、人間の心の内に実在する志向的状態と、心を持たない対象に帰属されるだけの志向的状態を区別する。例えば、誰かが「私は喉が渇いている」と発話する場合、発話者の心の内に、渇きの感情と飲みたいという欲求がある（Searle 1992:128-29）。これに対し、誰かが「企業は督促を意図している」と発話する場合、企業に同様の仕方で意図があるわけではない。サールは、人間の心に実在する志向的状態を「本来的志向性」（intrinsic-intentionality）、比喩的に語られる志向的状態を「あたかも志向性」（as-if intentionality）と呼ぶ。この区別に従うなら、企業の意図はあたかも志向性に過ぎない[19]。

意図の帰属の二つ目は、企業を意図的な行為者として扱うよう指令的（prescriptive）に、企業に意図を帰属させるものである。このタイプの帰属は、あたかも志向性の帰属が記述的帰属と呼ばれるのに対し、指令的帰属と呼ばれる（Velasquez 2003:547）。指令的帰属は、CIDS に言及する上記のフレンチの議論に見られる。つまり、M 社の督促でいえば、我々は CIDS に基づき、督促を M 社の意図的行為として扱うよう指令的に、M 社に意図を帰属しているということになる。フレンチの「CIDS は規範的な役割を遂行する必要がある。つまり、それは指令的であって単に記述

15 レンネガードとヴェラスキーズがあげるのは、本節のフレンチ（French 1995）の他、フレンチ（French 1979, 1984, 1992）、ペティット（Pettit 2007）の議論である（Rönnegard 2013:85; 2015:17-29; Rönnegard and Velasquez 2017:128-9）。他に、スミス（Smythe 1985）、ワーヘイン（Werhane 1985）、ウィーバー（Weaver 1998）、アーノルド（Arnold 2006）、杉本（2008）、ブラットマン（Bratman 2017）も、意図に関して機能主義を採用する。

16 以下で取りあげるレンネガードとヴェラスキーズの意図に関する議論は、その詳細をそれぞれの先行する議論に負っている。よって以下では、必要に応じてそれぞれの個別の議論も参照する。

17 意図の帰属にはもう一つ、企業のメンバーの全てあるいは多くが同一の意図を持つことの省略表現として、企業に意図を帰属させるという仕方がある（Rönnegard and Velasquez 2017:130; Velasquez 2003:545）。企業の意図の問題は、基本的に企業内の特定のメンバーに帰属できない意図を問題にするものであるため、本稿では省略表現としての帰属を扱わない。

18 2017 年の共著論文では、ペティット（Pettit 2017）の議論がこの帰属の例にあげられている。それ以前の議論では、デネットの志向システム論（Dennett 1987）があげられる（Rönnegard 2013:85-87; 2015:21-25）。志向システム論は、対象が信念や意図など志向的状態を持つと解釈することで、その振る舞いを予測できる場合、対象が当該の志向的状態を持つと考える議論である。

19 あたかも志向性を本来的志向性と混同する議論として、サールはしばしばデネットの志向システム論をあげる（Searle 1992:82）。

的なものではない」（French 1995:31）という言葉に照らしていうなら、意図の指令的帰属は、企業を意図的な行為者として扱うことを、規範的に要請するものだといえる。

付言しておくと、こうした規範的な要請は、企業活動だけに見られる特別なものではない。このことは、指令的帰属がサール（Searle 1995）の制度的事実（institutional facts）の亜種とされていることから読み取れる（Rönnegard and Velasquez 2017:129; Velasquez 2003:559 n.46）。制度的事実とは、特定の集団が企業であるとか、特定の人物がCEOであることなど、人間の制度的世界を構成する事実の総称である。この事実は、宣言（declarations）という言語行為が遂行され、かつ関係者がそれを適切なものと認識する場合に創り出される（Searle 2010:12）。例えば、我々は一定の条件（具体的には法定要件）を充たし、「ここに企業がある」と宣言することで、特定の企業があるという事実を創り出すことができる[20]。この事実は、関係者にその内容に相応しい仕方で行為することを要請する、義務論的力（deontic powers）と呼ばれる規範的な力を持つ（ibid. 8-9）。人間の制度上の行為は一般に、制度的事実の義務論的力に従う形で遂行される。企業を意図的な行為者とする指令的帰属の規範性も、制度的事実の義務論的力ということになるだろう[21]。

しかし、我々が制度的事実の内容に従って行為することを規範的に要請されるとしても、企業に心的な意図が生じるわけではない。そこには、本来的志向性の意味での意図がない。ヴェラスキーズは言う。「手続きもポリシーも、集団の心的状態や集合的心性を創造しない。フレンチも他の論者も、手続きとポリシーに従うことで集団がそれ以前になかった本当の志向性を創り出すという証明を提出していない。そうした論証がないなら、集団の意図は比喩的なものだという直観的でもっともらしい見方を放棄する理由はない」（Velasquez 2003:546）。

ここまでの議論が正しいならば、意図に関する機能主義は、いかなる帰属の仕方をもっても、企業に心的な意図があることを示すことはできない。さらに、他の論点でも、レンネガードとヴェラスキーズは、道徳的行為者であるために、行為に関する心的な知識や気づき（awareness）、感情など、心的な要素が必要だと主張する（Rönnegard and Velasquez 2017:128,131,137）。二人に言わせれば、心的状態に言及しない仕方で意図や他の要素を説明する機能主義的な議論は、いずれも誤りなのである。

レンネガードとヴェラスキーズの議論の検討

この議論に対して、どのような応答が可能だろう。セピンウォール（Sepinwall 2016）は、機能主義を採用することで企業の道徳的行為者性を擁護する論者と、心的な要素に言及することでそれを批判する論者の論争を次のようにまとめる[22]。すなわち、企業の道徳的行為者性の擁護者は、二つの課題のうち一つに直面することになる（ibid. 11）。一つ目は、道徳的行為者性に必要とされる心的な要素の類似物を企業に見つけ出すことである。意図やその他の心的状態を機能的特徴から

20 サールは、カリフォルニア州の会社法を例にあげる（Searle 2010:97-98）。このとき既存の法律が制度的事実であるように、制度的事実は他の制度的事実に依存して成立する場合もある。

21 しかし、サールのいう制度的事実に企業の意図が含まれるかどうかは疑問が残る。M社から届いた紙片が督促状であることは制度的事実だが、企業が督促の意図を持つというような事実を、サールは議論していない。

22 セピンウォールの議論は、企業の道徳的行為者性をめぐる対立が目下のところ解消できないこと、およびその対立のポイントがどこにあるかを論じるものであり、対立する立場のいずれが優勢かを示すものではない。セピンウォールによれば、対立のポイントは「企業の競合する概念ではなく、道徳的行為者性が何を要求するかについての意見の相違」（Sepinwall 2016:3）にある。

説明する議論が、これに当たるだろう。しかし、そのような類似物は、すでに見たように「粗末な代用品」（ibid. 11）として否定されることになる。

二つ目は、心的な要素が道徳的行為者性に必要な要素ではないと主張することである。例えば、アーノルドの「なぜこのような志向性の理解に同意しなければならないのか。ヴェラスキーズは何の議論もせずに、この立場の直観的な魅力なるものに頼っている」（Arnold 2006:284）[23] とか、ヘスの「これらの［心的な要素の］仮定は、議論されても正当化されてもいないし、それ自体が道徳的行為者性にとって必要不可欠なものではない」（Hess 2010:61）などの言明が、こうした主張に当たるだろう。つまり、道徳的行為者であるために心的な意図が必要だという仮定が正当化されない限り、機能主義を擁する論者は、レンネガードとヴェラスキーズの議論にかかずらう必要はないというわけである。

確かに、この仮定に関して、レンネガードとヴェラスキーズはサールの志向性論を援用して意図は心的なものだと主張する他に、特段の正当化を行っていない [24]。好意的に理解すれば、二人の議論は、サールに従い志向的状態は心的状態でしかないと考えることで、道徳的行為者性に必要とされる意図も心的状態でしかないと主張するものといえるだろう。基本的に、志向的状態（およびそこに想定される志向性）を心的なものと考える見解は、心の哲学の伝統に沿ったものであり、多くの議論の前提を構成してきたものである。この点を鑑みれば、機能主義を擁する側にこそ、機能的特徴を示せば心的な要素を考慮しなくてもよいと考えることの正当性や、そうした扱いをした場合の帰結を提示する責任があるように見える。それが企業の道徳的行為者性の当否を分ける論点なら尚更である [25]。

加えて、サールはその志向性論、すなわち本来的志向性とあたかも志向性を区別し、心的状態である前者のみを志向的状態と認める議論に関して、それ以上の正当化は必要ないことを強調している。その理由は、志向的状態が心的状態であるという事実が、社会的世界を構成する制度的事実が我々の認識や行為に依存して成立するのと対照的に、我々の行為や認識から独立した事実だからである（Searle 1998:9-10, 95）[26]。この事実は、例えば、我々は意識を持たないと主張する場合ですら、相手が意識的であることを前提せざるを得ないように、人間の認識や行為の前提をなすものである [27]。言い換えれば、この事実は正当化を要する理論や見解ではないのである。これに対し、機能主義は、志向的であるという志向的状態の特徴を機能的関係に置き換えることで、その特徴に関して何らの説明も与えないどころか、世界の多くの事物が人間と同様の仕方で志向的だと論じる誤った議論である（ibid. 50）。志向性に関して必要なのは、それがあるという事実を認

23 アーノルドの批判はヴェラスキーズ（Velasquez 2003）に向けたものだが、レンネガードとヴェラスキーズ（Rönnegard and Velasquez 2017）に向けたものと考えて論旨に影響はない。

24 意図以外の点でも、このことは同様に見える。行為に関する知識を心的なものと考える点に特段の正当化は見られない（Rönnegard and Velasquez 2017:127-8）。また、行為に関する気付きを心的なものと考える点に関しては、サール（Searle 1980）の中国語の部屋の議論で強調される意味の心的な理解が、これに相当するとされる（Rönnegard 2013:86-7; 2015:25）。加えて、この気づきの理解は、二人が考える自律性の条件にも適用される（ibid.）。レンネガードとヴェラスキーズの議論は、このように主要な論点を、サールの議論に負っている。本稿の以下でサールの議論を中心的な問題とするのは、この理由による。

25 脚注 14 でも言及したが、ブラットマンは、彼の議論を援用してきたフレンチやアーノルドより、企業に認められる意図が心的状態でないことにより注意を払っている（Bratman 2017:49-50）。

26 正確には、志向的状態が心的状態であることは、個人の主観によって捉えられるという点で、人間の主観に依存した事実である。しかしそれは、自分以外の観察に依存しないという点で独立だといわれる（Searle 1998:94）。

27 この例は、デネットに向けたサールの皮肉から構成した（Searle 1997:130）。サールは、外的世界の実在論に関する超越論的論証として、この形の議論を提示している（Searle 1995:177-98）。

めることであり、それを別の仕方で説明することではない。

このように見れば、レンネガードとヴェラスキーズは、これらの理由をもって、意図に関する心理主義に正当化は必要ないと主張できるだろう。それに対し、「もし［意図に関する］直観が別の場所にあるなら、もし志向性を示す存在のクラスが心的状態を持つものに限定されないという可能性を受け入れるのを厭わないなら、企業の志向性の議論が提供されるだろう」（Arnold 2006:284）というように、意図に関して異なる直観を持ち出しても、有効な反論とはならない。こうした仕方で異なる意図の理解を提示することはできても、レンネガードとヴェラスキーズもなお、「集団の意図は比喩的なものだという直観的でもっともらしい見方」（Velasquez 2003:546）のように、自分達の直観を繰り返すことが可能であり、議論を取り下げることはないからである。

したがって、意図に関する心理主義に正当化がなされていないとか、意図に関して異なる直観を支持すると主張するだけでは、レンネガードとヴェラスキーズへの有効な反論にならない。意図に関する心理主義それ自体が論駁されない限り、機能主義の方こそ、レンネガードとヴェラスキーズがかかずらう必要のない議論ということになるだろう。この点で、意図に関する心理主義が誤りであることの「証明の重荷は、まさに集団主義者（collectivist）の肩に課せられる」（ibid. 549）のである。セピンウォールのまとめに戻れば、その課題が果たされない限り、「我々に馴染みある現象学的な構成要素を欠いた道徳的責任の説明は、承認できないほど異質なもの」（Sepinwall 2016:11）ということでしかないだろう。

もっとも、この議論は、レンネガードとヴェラスキーズの主張を好意的に補足したものである。二人の議論に関しても、なぜ企業活動の道徳的責任が問題である場合に、心的な意図と同等の機能的特徴を示す事物を同じ仕方で扱うことが適切でないのか、より積極的な議論が求められるだろう。しかしながら、上で見たように、機能主義を擁する論者も、心という個人と企業の違いを捨象して機能的特徴から企業活動の道徳的責任を考えることの正当性、またその帰結を示すことが求められる。まとめれば、これらの点を示さない限り、双方とも相手を論駁するのに十分な議論を提示するには至らないのである。そこで次節では、意図に関する心理主義を正面から検討することで、意図をめぐる心理主義と機能主義の対立に答えを与えることを試みる。その議論は、意図に関する心理主義が企業活動の道徳的責任を論じる場面で適切でないことを示すだけでなく、意図に関する機能主義に関しても、これまで心理主義を批判することなくその見解を提示してきたため、機能主義に立つ場合の意図の理解を十分に議論できていないことを示すものとなる。

2. 企業活動と推論主義

本節では、レンネガードとヴェラスキーズによる意図に関する心理主義を検討するため、ブランダム（Brandom 1994, 2000）の推論主義を取りあげる。推論主義は、主張を中心とした言語表現の推論的な役割から、文の内容やそこに含まれる真理や指示などの概念を包括的に説明する、意味論における機能主義の一種である。さらに推論主義は、以下で見るように言語表現の推論的な役割から信念や意図など志向的状態を説明する点で、心の哲学における機能主義の一種でもある（Brandom 1994:154）。もちろん、（レンネガードとヴェラスキーズの依拠する）サールの議論に従えば、そうした議論は、前節で見たように志向的状態が心的状態であることを説明していない

と退けられるだろう。しかし、ブランダムをはじめ推論主義の支持者には、志向的状態を心的状態ではなく言語使用から理解すべきと主張することで、心理主義を否定する者もある（González de Prado Salas and Zamora-Bonilla 2015; Heath 2008; Salis 2017）。つまり、推論主義は心の哲学における機能主義の一種であると同時に、心理主義を含意しないのである。したがってその議論は、意図に関する心理主義を否定し、機能主義を支持する議論を提供すると見込まれる。以下では、推論主義による志向的状態の説明を概観した後、サールの心理主義と対比する形で、いずれの立場が適切かを検討する。

理由を与え求めるゲーム

まず、推論主義の基本的な道具立てを確認しよう。ブランダムは、合理的行為者と他の存在を分かつ特徴として、概念の理解に注目する。これは合理主義の一般的な特徴である。その中で、推論主義の独自性は、概念の理解を「理由を与え求めるゲーム」と呼ばれる言語実践から説明する点にある（Brandom 2000:48-9）。この実践は、参加者が主張を基本的な手番として、なされた主張を理由として次の主張を行ったり、なされた主張の理由を尋ねたり答えたりすることで進行する。

オウム（動物）と人間の違いに注目して、この実践を見てみよう。オウムと人間は共に、赤い物を見て「これは赤い」と発話する、弁別的に反応する信頼できる傾向性（reliable dispositions to respond differentially）を持つ。他方で、オウムと人間を分けるのは、この発話における「赤」の概念の理解である。例えば、人間は「これは赤い」に続き「色がある」とか「青くない」など、さらなる発話を行うことができる[28]。また人間は、なぜそのような発話を行ったのかを尋ねられた場合、その理由を説明することもできる。これに対しオウムは、「これは赤い」と発話することはできても、その発話を理由としたさらなる発話も、相手に発話の理由を尋ねられた場合に答えたりすることもできない。このような、理由を与える求める実践で適切に概念を使用できるかどうかが、オウムと人間の違いである。

このアイデアを、サールのそれと比較しておこう。言語行為論の点から見る場合、サールは、統語論的に正しく言語を扱う対象があったとしても、その対象が言語の意味を理解しているわけではないと考える（Searle 1980:422）。この議論は、中国語の部屋の議論として知られている。その要点は、プログラム通りに正しく言語を出力するコンピュータがあったとしても、コンピュータがその言語の意味を理解していることにはならないということである。意味の理解は、プログラムを入力したりその出力を解釈したりする人間の心の内にある。対照的に、推論主義はブランダムの「意味論は語用論に答えなくてはならない」（Brandom 1994:83）という言葉の通り、理由を与え求める実践でなされる言語使用から意味の内容を説明する。意味の理解は、このような実践に習熟していることとされる（ibid. 5）。つまり、推論主義は、意味の理解を主体の心の内に前提するサールの議論と逆の出発点をとるのである。意図を含む志向的状態も同様に、理由を与え求める実践でなされる言語使用から説明されることになる。

---

28 ブランダムは、このように我々が言語実践で日常的に用いる推論を実質的推論（material inference）と呼び、形式論理学の推論と区別する（Brandom 1994:97-8）。推論主義による意味や志向的状態の分析は、この推論の質的な良さを基に進められる。本稿で「推論的」という場合、この実質的推論の推論関係を指す。

信念的コミットメントとしての信念

具体例として、志向的状態の代表である信念の説明を確認しよう。行為の意図も、基本的に同様の道具立てで説明される。信念について、ブランダムは次のように言う。「信じることの状態または地位は、本質的に、単に偶然にではなく、主張するという言語的パフォーマンスに関連している。信念は、本質的に主張を行うことによって表現され得る種類の事柄である」（ibid. 153）。この見解と対照的に、サールは信念を、言語行為なしに行為者の内に実在する心的状態と考える。サールにとって、主張の内容は、信念の内容に基づき引き出される派生的なものである（ibid. 147, 671 n.8）。いずれの立場が適切かは後に検討する。引き続きブランダムの議論を確認しよう。

信念と主張が本質的に関連するというとき、両者の関係とはいかなるものだろう。ブランダムによれば、「信念は、主張を行うことで引き受けあるいは承認される、推論的に分節化される種類のコミットメントにおいてモデル化され得る」（ibid. 157）。コミットメントとは、言語実践の参加者が獲得する実践上の地位である。この地位は、規範的地位（normative status）、あるいは義務論的地位と呼ばれる（ibid. 142）。規範的地位には、コミットメントと、それと対をなすエンタイトルメントの二種類がある。ここでは差し当たり、コミットメントを、発話者が主張を行うことで引き受ける、当該の主張のエンタイトルメントを示す責任、エンタイトルメントを、主張を行うために発話者に帰属されていることが求められる権威と考える（ibid. 161）[29]。

例えば、誰かが「ここに火がある」と主張するとしよう。このとき、発話者は主張を行うことで、尋ねられた場合に当該の主張を行うエンタイトルメントのあること（例えば「煙が上がっているからだ」のような元の主張を行う権威のあること）を示す責任を引き受ける。言い換えれば、発話者は主張を行うことで、当該の主張を正当化することへのコミットメントを引き受ける。他方、聞き手の側では、同様のコミットメント（つまり「火がある」と主張したエンタイトルメントを示す責任）を発話者に帰属させる。このように、ブランダムの考える言語実践は、参加者が規範的地位を自ら引き受けたり帰属させ合ったりすることで進行する[30]。

さらに、上の引用にある「推論的に分節化される種類のコミットメント」とは、行為者のコミットメントやエンタイトルメントから何が帰結し、また何が両立不可能なものとして除外されるかに関する推論的な帰結関係に基づき（ibid. 168-70）、内容が詳細に規定されていくコミットメントをいう。例えば、発話者が「火がある」と主張することで引き受けるコミットメントは、「それは熱い」という主張へのコミットメントを帰結する一方、「それは水である」という主張へのコミットメントを両立不可能なものとして除外する。言語実践の参加者は、こうした帰結関係に従い次の主張や行為を行うことを通じて、当初は暗黙のうちにあったコミットメントの内容を明示的にしていくと考えられる。

信念は、このような仕方で主張を通じて引き受けたり帰属させ合ったり、またそれを通じて内

29 注意しておくと、ここで規範的あるいは義務論的な語彙を用いるからといって、ブランダムは道徳の話をしているのではない。むしろブランダムは、道徳的なものであるかどうかに関わらず、我々は合理的存在として、認識や行為において理由に拘束されると考える。ブランダムは言う。「この［理由の］力はある種の規範的力、合理的な「すべき（ought）」である。合理的であることは、これらの規範によって拘束ないし制約される存在であること、理由の権威に従う存在であることである」（ibid. 5）。

30 規範的地位を引き受ける、帰属させるという二種類の態度は、規範的態度（normative attitudes）と呼ばれる（ibid. 162）。引き受けには、下位クラスとして承認するという態度がある。脚注 33 を参照。

容が分節化されるコミットメントにおいてモデル化される[31]。こうしたコミットメントは、信念的コミットメント（doxastic commitments）と呼ばれる（ibid. 157）。これ以上の詳細には立ち入らないが、言語実践の参加者が信念的コミットメントを引き受けたり帰属させ合ったりする場合、参加者の規範的地位は上記以外にも様々に変化する。そこで重要な点は、発話者が主張を通じて信念的コミットメントを一人称的に引き受けるだけでなく、聞き手が発話者に対して、コミットメントを三人称的に帰属させる点である。例えば、もし発話者が「火がある」と言いながらそこで火傷してしまったとすれば、聞き手は発話者に、「彼は火がないと信じていた」のような、異なる信念的コミットメントを、三人称的に帰属させるかもしれない。このように、信念的コミットメントにおいては、一人称と三人称の両方の視点が本質的となる（ibid. 158）。別の言い方をすれば、「信念的にコミットすることは、特定の社会的地位を持つこと」（ibid. 142）なのである。このことは、意図の理解でも重要な点となる。

実践的コミットメントとしての意図

行為の意図も、基本的には信念と同様、言語実践における規範的地位の観点から説明される。行為の意図は二種類に分けられる。先行意図（prior intentions）と行為内意図（intentions in action）である[32]。双方の意図とも、行為者による行為へのコミットメントの承認（acknowledgements）[33]の点から説明される（ibid. 256）。

まず、先行意図は、行為に先立ち形成される意図である。この場合、行為者は「だろう」（will）とか「しよう」（shall）などの語を用いて、行為へのコミットメントを明示的に承認する。例えば、誰かが「私は公園へ行くだろう」と発話する場合、発話者は公園へ行くことへのコミットメントを明示的に承認する。このとき、聞き手は、同様のコミットメントを発話者に対して帰属させることになる。次に、行為内意図は、行為の遂行中に行為者が持つ意図である。この場合、行為者は端的に行為することで、行為へのコミットメントを承認する。このとき、実践の参加者は、行為者に同様のコミットメントを帰属させる。このような、行為に関わるコミットメントは、実践的コミットメント（practical commitments）と呼ばれる（ibid.）。

実践的コミットメントの場合も、信念的コミットメントと同様に、一人称と三人称の双方の視点が重要となる（ibid. 267-71）。例えば、「私は公園に行くだろう」という発話の場合、聞き手は「べき」（should）の語を用いて、発話者に対して、「彼女は公園へ行くべきだ」のような三人称的なコミットメントを帰属させることもある。あるいは、発話者もまた、「私は公園へ行くべきだ」のように、三人称的なコミットメントを自らに帰属させるかもしれない。

このように、推論主義は、信念や意図など志向的状態を言語実践の規範的地位の点から説明する。この議論が意味するのは、志向的状態の内容は、主体の内的な状態から一意に決まるものではないということである。このように見るとき、問題は、この議論がサールの心理主義に対する反論

31　もちろん、この説明だけでは、信念は主張と本質的に関連するかもしれないが、それは心的状態であると言うことが可能である。この点は、意図の説明を見た後に検討する。
32　二種類の意図はサール（Searle 1983）の区別による（Brandom 1994:256）。
33　承認は引き受けの下位クラスである。実践的コミットメントが問題の場合、両者の違いに注意する必要がある。例えば、行為者は推論的帰結として行為へのコミットメント引き受けながら、それを承認していない場合もあるとされる（ibid. 269）。以下では簡便さのため、実践的コミットメントに関して、承認にのみ言及して議論を進める。

となるかどうかとなる。

心理主義に対する批判

本節の冒頭で触れたように、ブランダムをはじめ推論主義の支持者は、志向的状態の内容を心的状態からではなく、言語実践から理解することが適切な立場だと主張する。ブランダムの議論は、基本的に特定の立場（本稿の関心でいえば心理主義）を明確に論駁するより、哲学史的な解釈も示しながら自身の見解を肯定的に説明するという形をとることが多い[34]。ここでは、ブランダムの議論を支持し、かつ心理主義を明確に否定する議論として、ヒースの議論を参照する（Heath 2008:103-11）。本稿は、ヒースの議論を、心理主義に対するブランダムの言及を正確に再現するわけではないが、心理主義に対するブランダムの問題意識を共有した上で、それを再構成しているものと考える。ヒースの論点に関して、ブランダムが同様の指摘を行っていることを脚注で示す。

ヒースの議論は、二つのステップから構成される。まず、(1)志向的状態は意味論と統語論により支配された命題の構造、言い換えれば文の形を持つ（ibid. 103）[35]。これは、志向的状態が、文的な構造を持つと考えなくては理解できない多くの特徴を持つことによる。いくつかの例をあげておこう。例えば、信念に関して、それが時制（明日雨が降る）、様相（雨が降るかもしれない）、否定（雨は降っていない）を含む場合、それらは心的なイメージではなく文の構造を持つと考えなくては理解できないだろう。また、同一指示表現を含む信念に関して、当該の表現を置換すると信念の同一性が保証されない場合がある（例えば、A と犯人が同一人物であっても、「A は指紋を残した」という信念と「犯人は指紋を残した」という信念の同一性は保証されない）ことも、信念が文の構造を持つことの帰結だと考えられる。他にも、信念に言表様相（de dicto）と事象様相（de re）で表現される区別がある（例えば、我々が「A は「窃盗犯が逃げた」と信じている」と言う場合と、「A は窃盗犯について「彼が逃げた」と信じている」と言う場合、我々は後者の場合に、窃盗犯の存在によりコミットしている）ことも、信念が文の構造を持つと考えなくては説明が難しい。以下では、志向的状態は本質的に文的なものではないとする見解も検討するが、このように見る限り、多くの志向的状態が文の構造を持つと考えるのは自然である。

次に、(2)志向的状態を文的なものと考える場合、問題は、文的な志向的状態を、サールのように心的状態として説明できるかどうかとなる。ヒースによれば、文的な志向的状態を心的状態として説明することはできない[36]。その理由は、文的な志向的状態を心的状態と考える場合、文に関わる規範性を説明できないからである。ヒースによれば、文はまず推論に用いられるが、推論は正しくなされたり誤ってなされたりするものである。また、推論の前提や帰結となる信念は、事態を正しく表象しているかどうかによって真または偽となる。さらに、文を構成する各概念もまた、それが指示しようとする対象や性質を指示できているかどうかで成功したり失敗したりする。

---

34 加えて、ブランダムが中心的な問題として扱うのは、心理主義そのものより、心理主義と一般に結びつく心的表象のアイデア、つまり心的状態に表象する心と表象される対象のそれ以上は説明不可能な関係が想定される点である（この立場をとる論者としてサールもあげられる）（Brandom 1994:68-70）。

35 ブランダムも、主張を中心とした言語実践から志向的状態の内容を考察するというように、志向的状態が文の構造を持つと考える（ibid. 5）。ブランダムはこの見解を、統覚が判断の形式（つまり文の形式）を持つと考えたカントの功績としている（ibid. 78-9）。

36 ヒースが念頭に置いているのは、心に意味論と統語論を備えた「思考の言語」があると考えるフォーダーの議論である（Heath 2008:106）。サールに対するブランダムの批判は脚注 43 を参照。

これらは、推論における文間的（intersententional）、文における文的（sententional）、文の構成要素における部分文的（subsententional）なそれぞれのレベルで、正しさと誤りという規範的基準のあることを示している（ibid. 108）[37]。そのため、文的な志向的状態を心的状態だと考える場合、心的状態からこうした規範性が説明できることを示さなくてはならない。

ではなぜ、規範性は心的状態から説明できないのか。それは、規範性が本質的に公的な性格のものだからである。ヒースは、ウィトゲンシュタイン（Wittgenstein 1958）の私的言語論に依拠して次のように言う。「ウィトゲンシュタインの私的言語論の要点はまさに、規範性を個人と世界の間の純粋に私的な関係として説明することができないということにある」（Heath 2008:109）。

次の例から考えよう。まず、A が無人島に漂着したとする。A は、島での滞在期間を記録しようと毎日灌木に刻みを入れようと決める。しかしこのとき、A には、自分がその日の刻みを入れたかどうか判断する手段がない。言い換えれば、A は一人で自分の行為の正しさを決める基準を持たないのである。このことは、刻みを入れたことの確認として木の側に小石を置くなど、さらなる手立てを講じても同様である。この例は、一人では、実際に正しいことと正しく思われることの区別がつかないことを示している。つまり、ある事柄が正しかったり誤っていたりするためには、他者や共同体による公的な基準が必要なのである。ヒースは言う、「誤りという規範的概念を意味あるものにするのは、まさにこの個人間の次元なのである」（ibid. 110）。

この議論に従えば、規範性は本質的に公的な性格を持つものである。そのため、文に関わる規範性もまた、同様に公的なものだと考えられる。つまり、我々が推論や文、概念を正しく使用できているかどうかの答えは、「他の人々が我々を理解できるかどうかに依存しなければならない。個人は自分だけで、自分が言うこと（あるいは考えること）が意味をなすかを決定する能力を持たない」（ibid. 113）ということになる。したがって、規範性を説明する資源が個人の心的状態に見出せない点で、規範性を持つ文の構造を持った志向的状態を、個人の心的状態から説明することはできないのである。

これに対して、推論主義は、すでに見たように志向的状態を言語実践の規範的地位から説明する点で、上記(1)(2)の基準を充たす。すなわち、(1)推論主義は主張、つまり文を基本的な単位として、(2)行為者がコミットメントやエンタイトルメントを引き受けたり帰属させ合ったりする規範的実践から、信念や意図を説明する。注意しておくべきは、こうした意図の理解が、従来の意図に関する機能主義とも異なる見解を提示する点である。すなわち、ブランダムが「原初的な（original）、独立した、あるいは非派生的な志向性は言語的な事柄（affair）である」（Brandom 1994:143）というように、意図は本質的に、行為者間の言語的、また社会的な地位の点から理解されるのである[38]。

---

37 ヒースは、この構造化をブランダム（Brandom 1994）に負っているとしている（Heath 2008:180）。ブランダムの議論でも、推論、表象、指示のそれぞれが正しさと誤りの規範的基準を持つことは繰り返し指摘される。これらに説明を与える際、ブランダムは推論の実質的な良さ（つまり推論レベル）から分析をはじめ、それに基づき表象や指示の適切さを説明する方針をとる。文的な志向的状態が規範的性格を持つという見解は、判断能力を規則に従う能力だと考えるブランダムのカント解釈に基づく（Brandom 1994:10-1, 30-1）

38 ヒースはより明確に次のように述べる。「何よりもまず重要なことは、志向的状態は心的状態でないということである。志向的状態は、理由を与えたり求めたりするゲームで書き留められる「標識」である。このゲームは最初、我々が習得する公的な実践である。内面化を通して、このゲームの「仮想的な」手番のシミュレーションを行う能力を獲得するのは、したがってこのゲームを用いて自分自身の計画能力を増幅する能力を獲得するのは、後になってからのことである」（Heath 2008:130）。

企業活動から見る心理主義と推論主義

サールはこの議論に同意するだろうか。実際、サールは上記の(1)(2)のいずれも否定する。しかし、本稿ではこの二点が、さらにこの二点を充たす推論主義が、企業活動の道徳的責任に関わる意図の分析に適切だと主張する。順に見ていこう。

(1)サールは、動物や幼児などの言語を持たない主体も、「水を飲みたい」とか「外へ出たい」のような志向的状態を持つため、志向的状態は必ずしも文的なものではないと主張する（Searle 1983:5）。しかし、企業活動の道徳的責任に関わる意図は、基本的に文的なものだと考えられる。フレンチのあげたM社の督促状の例でいえば、それは文の形で表現されるし、フレンチとSの対話も文を通じて交わされる。また、企業憲章をはじめ倫理綱領や定款、各種の契約や計画など、企業活動を構成する要素もまた文の形を持つ。これらの点を考えれば、企業活動を実現する意思決定や各種の行為（報告、命令、依頼や質問など）が、文の構造を持つ志向的状態に基づき遂行されると考えるのは自然である。

もちろん、企業活動は督促状にサインする際に「指を動かす」というような、基本的な身体動作を含む。その意図が文的なものかどうか判断は難しい。他方で、問題が企業活動の道徳的責任である場合、争点になるのは、誰がそのような身体動作を意図したかではなく、誰が督促を意図したかという点である。この場合の意図は、「誰々に督促を送付する」のような文的なものとなるだろう。加えて、企業活動ではグラフや画像、動画など、文を含まない媒体も使用される他、最終的な生産物がそうした媒体の場合もある。しかし、企業活動はそうした文を含まない媒体だけで進むわけではなく、必ずそれを用いる意図やそれを用いた帰結が、文の形で表現されることで進むはずである。さらに、最終的な生産物が文を含まない媒体であったとしてさえ、それを生産したり、使用したり、公開したりしたことで問題が生じた場合、発話であれ文書であれ文的な形でその意図が問われ、文的な形でそれに答えることが要請されるだろう。こうした点から見れば、企業活動の道徳的責任で問題になる意図が、文の構造をしていると考えるのは理に適っている。これに対し、サールのいうような文的でない意図（のようなもの）を持ち出しても、議論の目的に照らして適切な反論にはならない。むしろ、推論主義のいう主張を中心とした理由を与え求める実践こそが、文の形を持つ意図の分析に適切だと考えられる。

(2)サールは、志向的状態が本質的に規範的なものであることを認める一方、規範性が公的な基準を要請することを否定する（Searle 2001:182-3）。サールによれば、動物は目の前にある食べ物や障害物に関する信念を持つが、それは、正しかったり誤っていたりする点で規範性を持つ。この点で規範性は公的な基準を必要とするものではなく、自然の中にありふれたものと考えられる。しかし、企業活動の道徳的責任が問題である場合、問題になるだろう意図は、組織内や社会で共有された公的な規範によって理解されると考えるのは自然である。M社の督促でいえば、Sが意図的にフレンチ宛の督促にサインしたのか、誤って意図しない通知にサインしたかにかかわらず、それはM社の督促として理解される（French 1995:23）。つまり、M社の督促は、Sの心的状態や自己理解ではなく、公的な基準に従って理解されるのである。この例によらずとも、企業の構成員が、関係者の主張や行為を好きな仕方で解釈してはならないのは自明である。このように見れば、動物の志向的状態（のようなもの）を持ち出し規範性が公的なものであることを否定する議論は、より基本的な規範性の理解に関してなお妥当する余地があるかもしれないが、企業活動の道徳的

責任が問題である場合、適切な反論にはならない[39]。

このように、企業活動の道徳的責任に焦点を当てる場合、(1) 問題となる意図が文的な構造を持ち、(2) 公的な規範の基準に服すと考えることは理に適っている。反対に、意図に関する心理主義は、これらの点を充たさないため、適切な議論を提供することに失敗している。したがって、レンネガードとヴェラスキーズは、有効な議論を提起できていない。これに対し、推論主義は (1) 主張、すなわち文の使用を基本単位とし、かつ (2) その規範的な使用から志向的状態を説明する議論であり、企業活動の道徳的責任を問うという目的に照らして適切なものだといえる。この点で、問題を企業活動の道徳的責任に限定する場合、推論主義は適切な意図の理解を提供する議論といえる[40]。レンネガードとヴェラスキーズの機能主義への批判に照らしていうならば、言語実践に基礎を置く機能主義こそ、企業の道徳的行為者性を論じる際に適切な立場だということになる[41]。

3. 推論主義から見る企業の道徳的行為者性

では、企業活動の道徳的責任の問題に推論主義が適切だと考える場合、企業の道徳的行為者性はどのように理解されるだろう。推論主義は、志向的状態の内容を言語実践の役割から説明する点で、機能主義の一種である。他方でこの議論は、企業や人間を含む誰の意図をも機能的側面から説明する点で、企業の意図に関して機能主義を、人間の意図に関して心理主義を採用する、企業に意図を認める従来の議論とは異なる方針を取る。さらに、前節の議論に従う限り、推論主義を採用できるのは、議論の対象を文の構造を持つ志向的状態に限定する場合である。この制約が、企業の道徳的行為者性の議論にどのような含意を持つかは明らかではない。

そこで本節では、企業の道徳的行為者性を推論主義から検討することで、本稿としての評価を示す。まず、企業の意図の問題を検討して、その上で企業の道徳的行為者性を論じる。

企業の意図

企業の意図の問題から考えよう。推論主義は企業の意図を認めるだろうか。前節の議論に従うならば、その答えは、企業が実践的コミットメントとしての意図を自ら承認したり、互いに帰属したりされたりすることができるかによって決まる。さらに、行為の意図は、信念的コミットメントをはじめとする他の規範的地位との関係で理解される。したがって問題は、企業を各種の規範的地位の適切な主体として理解できるかどうかということになる。

ブランダムはこの点、その答えは寛容であるべきと述べるに留まる（Brandom 1994:644）。他方で、ゴンザレスとザモラは、推論主義に基づき集団の行為者性を論じた論文で、次のように答え

39 この点、推論主義を採用する場合、志向的状態の内容を説明する言語実践の規範性が、そもそもどうやって成立するかが問題となる（Brandom 1994:18-55）。しかし、企業のようなフォーマルな組織が問題の場合、安定したコミュニケーションを成り立たせる程度の規範性は所与としてよいと思われる。ブランダムは、規範的地位の説明で提示される権威と責任の関係の説明において、自身の提示する言語実践が組織行動の研究と親和的であると指摘している（ibid. 673 n.24）。

40 しかしこのことは、志向的状態の説明に推論主義が必要であることを意味しない。志向性に想定される表象関係を、別の仕方で説明できる可能性もあるからである（白川 2017:14）。

41 他方で、推論主義には、議論が言語実践の内部での話に終始しているために、言語を超えた実在的世界との関係が説明できないという批判がある（白川 2015）。この点は、言語行為論から制度的事実の成立を議論し、同時に外的世界の実在論や真理の対応説を主張するサール（Searle 1995, 1998）の議論と対照的である。

る。すなわち、その答えは「人々に受け入れられている言説的な規範に依存する」（González de Prado Salas and Zamora-Bonilla 2015:15）[42]。二人によれば、この答えはトリビアルなものだが、この基準から見ると、スポーツチーム、企業、組織、政府など特定の集団は、推論主義的な意味での行為者として理解することができる。例えば、サッカーチームは、ボールが適切な仕方でラインを通過した回数に応じて点数、つまり規範的地位を持つ。「これらの存在は、明らかに義務論的地位、すなわち世界の何らかの出来事をそれら存在の行為…として我々に解釈させる義務論的地位を伴う」（ibid.）のである。

しかし、二人の議論は、推論主義の重視する言語実践への参加を、単に何らかの規範的実践に参加することに矮小化している。少なくとも、サッカーの試合は（問題が生じない限り）明確な言語実践なく進行するのに対し、企業活動はそうではない。この点で、両者を同一視することはできない。

さらに、この議論は、企業が何らかの規範的地位を帰属される（あるいは規範的地位を引き受けていると解釈できる）ことを示すだけで、企業もまた実践の参加者に規範的地位を帰属させる可能性を検討していない。これは、企業を何らかのゲームの参加者と捉える従来の議論に共通する特徴である（French 1995; Ladd 1970; 杉本 2008）。これらの議論は、ゴンザレスとザモラのものも含めて、企業活動を外側から眺めて、企業を規範的実践の参加者と解釈しているに過ぎない。

これに対し、ブランダムが重視するのは、解釈者によって解釈される行為者と、解釈を行う行為者の違いである（Brandom 1994:55-65）。前者が持つ志向的状態は、後者が持つ志向的状態に依存する。ブランダムが説明しようとするのは、基本的に後者の志向的状態である[43]。後者は、前節で見たように、行為者が言語実践で引き受けたり帰属したりされたりする規範的地位の点から説明される。この点からいえば、志向的状態は「彼ら自身［実践の共同体］の活動の産物であって、その活動を解釈する理論家の産物ではない」（ibid. 61）のである。したがって、企業が規範的地位の適切な主体だと主張するためには、企業に規範的地位を帰属できるとか、企業がそれを引き受けていると解釈できることだけでなく、企業もまた実践の参加者に規範的地位を帰属できることを示さなくてはならない。

M 社の例から考えよう。フレンチは、M 社の督促について、M 社をその主体として対応していた。これは、フレンチによる M 社への、督促の実践的コミットメントの帰属である。この督促は同時に、M 社による督促の実践的コミットメントの承認とも考えられる。また、フレンチが問い合わせを行った際、S は M 社の記録に言及して、督促が適切であることを説明している。これは、M 社による督促のエンタイトルメントの提示と考えられる。さらに、この対話で、フレンチは督促が誤りだという信念的コミットメントを自ら引き受け、また同様のコミットメントを帰属されることにもなる。この帰属は、M 社によるフレンチへのコミットメントの帰属と考えられるだろう。もちろん、M 社の発話や行為は、S を通じて実行されるものである。次項で見るように、それら

42 推論主義ではないが、フセインとサンドバーグ（Hussain and Sandberg 2017）も、企業が行為する規範次第で企業の行為者性の評価が変わるという、多元論的機能主義を主張する。

43 ブランダムはこれを原初的志向性（original intentionality）と呼ぶ（Brandom 1994:60-61）。これは、サールの本来的志向性に相当する。ブランダムはサールと同様に、解釈される主体と解釈を行う主体に区別を設けないデネットの議論を問題視する。他方でブランダムは、原初的志向性（本来的志向性）を解釈者の内在的な信念や意図と考えるサールの議論に関して、信念や意図の内容を後退なく説明できない点で欠陥があるとしている（ibid.）。

の発話や行為が、Sの私的なものである可能性は常にある。他方で、フレンチがSの発話や行為をSの私的なものとして対応するなら、それが特異なものでない限り、不適切なのはフレンチの方ということになるだろう。そのような不適切さは、企業の側から理由を尋ねられたり、場合によって修正されたりすると思われる。

このように見れば、企業は規範的地位を引き受けたり帰属されたりするだけでなく、実践の参加者に帰属させもする主体だと考えられる。M社の督促を、フレンチは人間と企業のコミュニケーションと表現したが（French 1995:13）、推論主義の語彙で言い直せば、それは人間と企業による理由を与え求める実践ということになるだろう。

全てが企業の意図なのか

とはいえ、この議論は、企業がその活動を構成する全ての意図の主体だと主張するものではない。少なくとも、次の二点に注意する必要がある。一つ目に、企業活動を構成する意図の中には、「督促状をポストまで運ぶ」とか「電話の受話器を取る」のように、企業がその主体となることができないものもある。しかし、このような意図が企業活動に含まれること自体は、企業活動に企業が主体となる意図があることを否定しない。特に、基本的に責任の所在で問題になるのは、誰が督促状を集荷場所まで運ぶことを意図したかとか、誰が督促の電話で受話器を持ち上げたかというような、身体運動に言及する意図ではなく、誰が督促を意図したかのような、企業がその主体となり得る意図であるように思われる[44]。いずれにせよ、企業が全ての意図の主体になるわけではないことは、企業が道徳的責任の所在で問われる意図の主体として排除されることを意味しない。

二つ目に、問題を企業が主体となることのできる意図に限定しても、その意図が企業のものかどうか、常に議論の余地は残る。これは、フレンチによる企業の意図の議論と同様である。企業の意図と個人のそれを区別する基準の一つは、フレンチの言う通りCIDSだろう。実践の参加者は、問題の決定や行為が企業のものかどうかを尋ねることができるし、問われた相手はCIDSに言及して、それが企業の意図であることを説明することができる。また、なされた決定が企業のどのレベルのものか、例えば督促がM社それ自体のものか、あるいはM社のクレジット部門のものかといった点も、CIDSの組織フローチャートに照らして説明が可能である。このように見れば、CIDSは、問題の決定や行為が企業それ自体や企業内の特定の組織のものであることを正当化するエンタイトルメントの役割を果たすといえる。もちろん、特定の決定や行為が、企業内の個人に帰属されるべき場面もある。例えば、故意による企業の倒産、合併や分裂、社名の変更など、企業が道具主義的に扱われる場合、その決定は個人に帰属されるべきだろう（杉本 2008:52）。個々の事例で問題の決定や行為が誰のものであるかは、その都度の実践の様態に照らして判断される事柄である。重要な点は、企業活動を意図する主体が誰であるかは、常に実践で問われる余地があるということである。

以上、このように注意すべき点はあるが、企業が理由を与え求める実践の参加者であり、規範的地位を引き受けたり帰属したりされたりすると考えることは、理に適っていると思われる。こ

44 もちろん、企業内の個人は、身体運動をもって督促を発送したことや、督促の電話をかけたことの責任を問われる場合もあるだろう。この点については、次項で言及する。

の点で、推論主義から企業の意図を正当化することは可能である。

企業の道徳的行為者性

次に、企業の道徳的行為者性は上記の議論から正当化されるかを検討する。第一節で見た道徳的責任の理解に照らせば、企業が道徳的行為者だというためには、企業が意図を持つことに加え、行為を引き起こすことを示さなくてはならない。本稿では、意図が行為の原因となるかどうかを問う、いわゆる行為の因果性の問題には立ち入らない[45]。差し当たり、この点に関するブランダムの基本的な考えを示すに留める（Brandom 1994:259-61）。

まず、意図と行為の関係は、第一に規範的な関係として理解される。すなわち、行為者が特定の行為への実践的コミットメントを承認するなら、行為者はそのコミットメントを理由として行為すべきという関係である。他方で、ブランダムによれば、特定の場合にそのコミットメント（理由）は、行為の原因とも考えられる。ここで、特定の場合や原因の内実など、議論の詳細には立ち入らない[46]。企業が行為を引き起こすかどうかという点に関していえば、次項で見る、行為を引き起こす事態の内実に関する、企業と個人の相違が重要である。今の段階で確認しておくべきは、ブランダムの説明を額面通りに受け取るなら、企業の意図もまた、特定の場合に行為の原因として理解できるだろうことである。そうであれば、企業が意図を理由として行為を引き起こすと考えることは可能である。

付言しておくと、本稿では、道徳的行為者性の条件を、意図的に行為を引き起こす行為者であることとして議論を進めてきた。この条件に対しては、例えば、自律性（Rönnegard and Velasquez 2017）、合理性（List and Pettit 2011）、会話可能性（Pettit 2017）など、様々な追加条件が提案されている。これらの条件を推論主義がどう説明するか、さらなる検討が必要であることは言うまでもない。他方で、本稿で設定した、意図的に行為を引き起こす行為者であるという条件を企業が充たすのであれば、その限りで、企業は道徳的行為者だと考えられる。

個人と企業の関係をめぐる三つの課題

しかし、ここまでの仕方で企業の道徳的行為者性を正当化する場合、なお検討すべき問題は残る。本稿の最後に、推論主義から企業の道徳的行為者性を論じることで生じる問題として、次の三点を指摘しておく。

一つ目に、推論主義に基づき、意図が行為を引き起こすことを説明できるとしても、その内実は、個人と企業で大きく異なる。企業の行為は、個人の行為なしには引き起こされない点で、その成否を個人に依存する（Werhane 1985:52-5）。そうであれば、企業の行為に関わる個人は、部分

45　行為を引き起こすことの内実に関して、ヴェラスキーズは1983年の論文で、行為者を行為の創始者と考える行為者因果説の立場を示している（Velasquez 1983:13-4）。他方で、2003年の論文では、「事物に関する思考あるいは事物の知覚（またはその事物が内包的対象である何らかの他の志向的状態）がそれ自体でリアクションや反応を引き起こす時に、その事物が提示する因果性」として、志向的因果性（intentional causality）という概念を導入しており、志向的状態を行為の原因と考える立場も示している（Velasquez 2003:540, 555 n.28）。以下では、志向的状態が行為の原因となるかを問う後者の意味で議論を進めるが、仮に前者の行為者因果が問題になる場合、以下で言及する企業活動が個人の行為に依存するという点がより大きな問題となると思われる。

46　ブランダムによれば、意図が行為の原因とされるのは、行為者が行為へのコミットメントを承認し、かつエンタイトルメントを帰属される場合である（Brandom 1994:262）。この議論で注意すべきは、こうした意図と行為の関係が、あくまで規範的関係（実践の参加者の規範的態度）を通じて理解される点である（ibid. 14-17, 254）。

的にせよ責任を問われる余地があるように見える。ヴェラスキーズは、「企業の行為のある場面には、必ずそれを因果的に引き起こしたことに責任のある個人がいる」（Velasquez 2003:543）と主張するが、企業それ自体に責任を問う場合も、この議論は有効である。企業に道徳的責任を帰属させる場合の、個人責任のあり方を検討する必要がある[47]。

二つ目に、推論主義は企業と個人を意図的な行為者として同等に扱うが、そうした議論が、企業の非倫理的行為を促進するのではないかという懸念がある。アシュマンとウィンスタンリー（Ashman and Winstanley 2007）は、企業の道徳的行為者性は実践的な利点がある一方、これを認める場合、企業の対外的な統一性の裏側で非倫理的行為が覆い隠されると主張する[48]。こうした懸念は、杉本の「企業にも道徳的人格を認めるほど人格概念を劣化させてしまうことは、翻ってそこで働く経営者や従業員らの物件化になりかねない」（杉本 2019:77）のような指摘にも読み取れる。

この点に関して、本稿の議論を振り返れば、問題は企業の道徳的行為者性の正当化にではなく、その正当化を適切とみなさせる実践にあると思われる。本稿の議論が正しいとすれば、企業がその活動の道徳的責任に関する意図の主体となるかどうかは、実践の内容に依存して決まる問題である。例えば、本稿では、身体運動を含む意図のような、人間だけを主体とする意図が議論から除外されると論じることで、企業が問題となるだろう意図の主体になり得ることを主張した。これは、身体の有無のような個人と企業の違いが捨象され、両者が言語実践で機能的に等価な主体として扱われる実践で、企業の道徳的行為者性が成立することを意味する。翻っていえば、実践の内容次第で、企業の道徳的行為者性は否定されもするのである。従来の議論は、個人を基準とした企業の特徴の評価に終始してきたが、企業と個人の相互的な実践に目を向ける必要がある。

最後に、推論主義により志向的状態を企業と個人の相互的実践から説明する場合、道徳的行為者性の理解そのものが見直される必要がある。本稿で採用した道徳的行為者性の条件は、信念や意図などの志向的状態を、主体の内的な資源と考える心理主義的な前提において設定されている。これに対し推論主義は、ここまでに見たように、志向的状態の理解を大きく変える。特に注意すべきは、本稿の議論に従う場合、個人の志向的状態もまた、企業とのやり取りの影響を免れない点である。例えば、ヒースは次のように言う。「心的状態［志向的状態］がその内容を決定的に言語に依存するなら、そして言語が社会的実践から生じるなら、個人の志向的な計画能力…は非常に重要な仕方で行為者の社会的環境に依存するように見えるだろう」（Heath 2008:102）。このような志向的状態の特徴は、従来から指摘される企業内の個人の脆弱性、例えば、意志の弱さ（Garrett 1989）、マルクス主義的な疎外（Ladd 1970）、企業のプレッシャーや拘束力（Gilbert 2014; Pettit 2003; Werhane 1985）、社会心理学の集団浅慮（Phillips 1995）などと親和的に見える。主体の意思決定の資源が社会的産物だと考えられる場合の、道徳的行為者性の理解が明らかにされる必要がある[49]。

---

47 フレンチは、人間のいない完全にプログラム化された企業の例をあげているが、その場合は異なる議論になるだろう（French 1995:34-35）。プログラムを作成した人間の責任など、人工知能の責任論も参照する必要がある。

48 二人は、心を持たない企業に志向的状態を認めることはできないというヴェラスキーズの議論を支持し、企業の道徳的行為者性は理論的には誤りだと考える（Ashman and Winstanley 2007:87-88）。

49 ブランダムがこうした脆弱性をどう評価するかは検討が必要である。ブランダムの想定する行為者は、言語実践で理由を与え求めることを通じて、自らを発展させていく合理的主体である（Brandom 1994:3）。ヒースは、個人は規範に従って行為する存在だが、その内容が規範倫理学で論じられるような道徳的内容を伴うことは保証されないと主張

結　論

本稿では、企業の道徳的行為者性をめぐる企業の意図の問題について、レンネガードとヴェラスキーズの議論に反論を提起する形で推論主義を導入し、二人の主張する心理主義が妥当性を欠き、推論主義が適切であることを示した。その上で本稿では、推論主義から企業の意図、また企業の道徳的行為者性が、条件つきではあるが正当化できることを示した。ここまでの議論は、あくまで行為の意図に焦点を当てたものであり、企業の道徳的行為者性をめぐって提示される論点を尽くしたものではない。しかし、本稿で示した推論主義に基づく議論は、従来の機能主義に基づく議論よりも、企業の道徳的行為者性をめぐる議論において、機能主義が適切であることを示している。推論主義に基づくさらなる分析により、企業の道徳的行為者性に関するより適切な理解を提示することが期待される[50]。

参考文献

Arnold, Denis G. 2006. “Corporate Moral Agency.” *Midwest Studies in Philosophy* 30 (1):279-91.

Ashman, Ian, and Diana Winstanley. 2007. “For or Against Corporate Identity? Personification and the Problem of Moral Agency.” *Journal of Business Ethics* 76 (1):83-95.

Brandom, Robert. 1994. *Making It Explicit : Reasoning, Representing, and Discursive Commitment*. Harvard University Press.

Brandom, Robert. 2000. *Articulating Reasons : An Introduction to Inferentialism*. Harvard University Press.

Bratman, Michael E. (1987) 1999. *Intention, Plans, and Practical Reason*. CSLI Publications.

Bratman, Michael E. 2014. *Shared Agency : A Planning Theory of Acting Together*. Oxford University Press.

Bratman, Michael E. 2017. “The Intentions of a Group.” Pp. 36-52 in *The Moral Responsibility of Firms*, edited by E. W. Orts and N. C. Smith. Oxford University Press.

Danley, John R. 1980. “Corporate Moral Agency.” *Bowling Green Studies in Applied Philosophy* 2:140-49.

Dennett, Daniel. 1987. *The Intentional Stance*. MIT Press. (若島正 , 河田学訳『「志向姿勢」の哲学 : 人は人の行動を読めるのか ?』, 白揚社 , 1996 年).

Donaldson, Thomas. 1980. “Moral Agency and Corporations.” *Philosophy in Context* 10:54-70.

French, Peter A. 1979. “The Corporation as a Moral Person.” *American Philosophical Quarterly* 16 (3) :207-15.

French, Peter A. 1984. *Collective and Corporate Responsibility*. New York: Columbia University Press.

French, Peter A. 1992. *Responsibility Matters*. University Press of Kansas.

French, Peter A. 1995. *Corporate Ethics*. Harcourt Brace College Publishers.

Friedman, Nick. 2020. “Corporations as Moral Agents: Trade - Offs in Criminal Liability and Human

---

する (Heath 2008:258-85)。

50 本稿は、JSPS 特別研究員奨励費 JP20J11383、公益財団法人上廣倫理財団平成 30 年度研究助成金の助成を受けた研究成果の一部である。また、本稿のもとになる発表を行った北日本哲学研究会、北海道大学哲学会、応用哲学会、Society for Business Ethics の会場でコメントをいただいた方々、論文審査の過程で大変丁重なコメントをいただいた匿名査読者の方々にお礼を申し上げる。

Rights for Corporations." *The Modern Law Review* 83 (2):255-84.

Garrett, Jan Edward. 1989. "Unredistributable Corporate Moral Responsibility." *Journal of Business Ethics* 8 (7):535-45.

Gilbert, Margaret. 2014. *Joint Commitment : How We Make the Social World*. Oxford University Press.

González de Prado Salas, Javier, and Jesús Zamora-Bonilla. 2015. "Collective Actors without Collective Minds: An Inferentialist Approach." *Philosophy of the Social Sciences* 45 (1):3-25.

Heath, Joseph. 2008. *Following the Rules : Practical Reasoning and Deontic Constraint*. Oxford University Press.（瀧澤弘和訳『ルールに従う：社会科学の規範理論序説』, NTT 出版 , 2013 年）.

Hess, Kendy. 2010. "The Modern Corporation as Moral Agent: The Capacity for 'Thought' and a 'First-Person Perspective.'" *Southwest Philosophy Review* 26 (1):61-69.

Hess, Kendy M. 2013. "The Free Will of Corporations (and Other Collectives)." *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 168:241-60.

Hussain, Waheed, and Joakim Sandberg. 2017. "Pluralistic Functionalism about Corporate Agency." Pp. 66-86 in *The Moral Responsibility of Firms*, edited by E. W. Orts and N. C. Smith. Oxford University Press.

Ladd, John. 1970. "Morality and the Ideal of Rationality in Formal Organizations" edited by S. J. B. Sugden. *Monist* 54 (4):488-516.

List, Christian., and Philip Pettit. 2011. *Group Agency : The Possibility, Design, and Status of Corporate Agents*. Oxford University Press.

Moore, Geoff. 1999. "Corporate Moral Agency: Review and Implications." *Journal of Business Ethics* 21 (4):329-43.

Orts, Eric W., and N. Craig Smith. 2017. *The Moral Responsibility of Firms*. Oxford University Press.

Pettit, Philip. 2003. "Groups with Minds of Their Own." Pp. 167-93 in *Socializing Metaphisics*, edited by F. Schmitt. Rowman and Littlefield.

Pettit, Philip. 2007. "Responsibility Incorporated." *Ethics* 117:171-201.

Pettit, Philip. 2017. "The Conversable, Responsible Corporation." Pp. 15-35 in *The Moral Responsibility of Firms*, edited by E. W. Orts and N. C. Smith. Oxford University Press.

Phillips, Michael J. 1995. "Corporate Moral Responsibility: When It Might Matter." *Business Ethics Quarterly* 5 (3):555-76.

Rönnegard, David. 2013. "How Autonomy Alone Debunks Corporate Moral Agency." *Business and Professional Ethics Journal* 32 (2):77-107.

Rönnegard, David. 2015. *The Fallacy of Corporate Moral Agency*. Springer.

Rönnegard, David, and Manuel Velasquez. 2017. "On (Not) Attributing Moral Responsibility to Organizations." Pp. 123-42 in *The Moral Responsibility of Firms*, edited by E. W. Orts and N. C. Smith. Oxford University Press.

Salis, Pietro. 2017. "Conceptions of Original Intentionality (and Social Ontology)." pp. 7-15 in *Mind, Collective Agency, Norms*, edited by P. Salis and G. Seddone. Germany: Shaker: Aachen.

Sanders, John. 1993. "Assessing Responsibility: Fixing Blame versus Fixing Problems." *Business and Professional Ethics Journal* 12 (4):73-86.

Searle, John R. 1980. "Minds, Brains, and Programs." *Behavioral and Brain Sciences* 3 (3):417-24.

Searle, John R. 1983. *Intentionality : An Essay in the Philosophy of Mind*. Cambridge University Press.

（坂本百大監訳『志向性：心の哲学』, 誠信書房 , 1997 年）.

Searle, John R. 1992. *The Rediscovery of the Mind*. MIT Press.

Searle, John R. 1995. *The Construction of Social Reality*. Free Press.

Searle, John R. 1997. *The Mystery of Consciousness*. New York Review of Books.（笹倉明子他訳『意識の神秘：生物学的自然主義からの挑戦』, 新曜社 , 2015 年）.

Searle, John R. 1998. *Mind, Language and Society: Philosophy in the Real World*. Basic Books.

Searle, John R. 2001. *Rationality in Action*. MIT Press.（塩野直之訳『行為と合理性』, 勁草書房 , 2008 年）.

Searle, John R. 2010. *Making the Social World : The Structure of Human Civilization*. Oxford University Press.（三谷武司訳『社会的世界の制作：人間文明の構造』, 勁草書房 , 2018 年）.

Sepinwall, Amy J. 2016. “Corporate Moral Responsibility.” *Philosophy Compass* 11 (1):3-13.

Smythe, Thomas W. 1985. “Problems about Corporate Moral Personhood.” *The Journal of Value Inquiry* 19 (4):327-33.

Velasquez, Manuel G. 1983. “Why Corporations Are Not Morally Responsible for Anything They Do.” *Business & Professional Ethics Journal* 2 (3):1-18.

Velasquez, Manuel G. 2003. “Debunking Corporate Moral Responsibility.” *Business Ethics Quarterly* 13 (4):531-62.

Weaver, William. 1998. “Corporations as Intentional Systems.” *Journal of Business Ethics* 17 (1):87-97.

Werhane, Patricia H. 1985. *Persons, Rights, and Corporations*. Prentice-Hall.

Wittgenstein, Ludwig. 1958. *Philosophical Investigations, Translated by G.E.M. Anscombe*. Basil Blackwell（藤本隆志訳『ウィトゲンシュタイン全集 8』, 大修館書店 , 1976 年）.

杉本俊介 . 2008. “ 企業の道徳的行為者性を擁護する：デイヴィッド・ゴティエの理論を応用する試み .” 実践哲学研究 = *Studies for Practical Philosophy* (31):41-59.

杉本俊介 . 2019. “ ビジネス倫理 企業それ自体の責任を問うことの困難さ：ビジネス倫理学の新展開（特集 倫理学の論点 23）.” 現代思想 47 (12):72-78.

白川晋太郎 . 2015. “ ブランダムにおける客観性 .” アルケー：関西哲学会年報 23:117-28.

白川晋太郎 . 2017. “ なぜ推論主義をとるべきなのか .” 京都大学文学部哲学研究室紀要 *: PROSPECTUS* (19):1-17.

北海道大学応用倫理・応用哲学研究教育センター

『応用倫理 ── 理論と実践の架橋』第 13 号　論文公募のお知らせ

『応用倫理 ── 理論と実践の架橋』編集委員会では、応用倫理学に関する研究論文、研究ノート、書評を下記の要項・投稿規定において公募いたします。なお、投稿は随時受け付けておりますが、第 13 号への掲載は 2021 年 11 月 30 日までの投稿を目安とします。皆様の御投稿をお待ちしております。

1. テーマは応用倫理学に関わるものとする。
2. 論文は独創性を有する学術研究成果をまとめたものとし、研究ノートは萌芽的研究の中間報告等とする。
3. 応募論文および研究ノートは未発表のもので、本『応用倫理』以外に同時投稿していないものに限る。二重投稿の場合、審査対象としない。

※ただし、外国語で既に発表された論文を著者本人が日本語に訳したものを投稿する場合は二重投稿とは見なさない。また著者本人が外国語で発表した論文に基づいて執筆されたために、もとの外国語論文と内容的に重複が多い場合も二重投稿とは見なさない。その際には投稿する際に、当該外国語論文の翻訳であること、ないしは当該外国語論文に基づくものであることを別紙により記載して申し出ることとし、掲載が決定した際には、その点を論文の中で明記することとする。

4. 使用言語は日本語とする。英語論文については *Journal of Applied Ethics and Philosophy* にて受け付ける。
5. 論文および研究ノートの分量は 1 万～ 2 万字を目安とする。書評は 2000 ～ 4000 字程度とする。
6. 論文または研究ノート投稿者は『応用倫理』編集事務局に、①論文または研究ノートの原稿、②論文または研究ノートの和文要旨（500 字程度）および英文要旨（250 語程度）、③著者略歴（100 字程度）の電子媒体テキスト（MS ワードのファイルを記録した CD-R を添付）およびハードコピー 3 部を送付する。また電子媒体のものは本センター事務局宛にメールでも送付すること。
7. 書評投稿者は、『応用倫理』編集事務局に書評原稿を電子テキスト（MS ワードによる添付ファイル）にて送付する。
8. 投稿された論文及び研究ノートは、編集委員会が定める査読者 2 名により審査され、編集委員会において選考される。
9. 編集委員会は査読者の審査の結果を踏まえ、投稿者に対して修正・書き直しを求めることができる。修正・書き直し後に再投稿されたものについては、必要に応じて再査読を行う。
10. 掲載可となった論文及び書評は、ウェブページ及び冊子体により公開する。
11. 掲載の可否については編集委員会が最終決定を行う。

※ 本誌の査読はダブル・ブラインドで行っているので、論文本体には著者氏名は書かず、「拙論」等の表現も使わないこと。

○ 過去の本誌の内容は、北海道大学応用倫理・応用哲学研究教育センターのウェブサイト上及び北海道大学学術成果レポジトリ「HUSCAP」でご覧いただくことができます。　http://caep-hu.sakura.ne.jp/

http://eprints.lib.hokudai.ac.jp/dspace/bulletin.jsp

論文送付先／問い合わせ先

〒 060-0810　札幌市北区北 10 条西 7 丁目　北海道大学大学院文学研究院　応用倫理・応用哲学研究教育センター

（電子媒体テキスト送付アドレス）　E-mail：caep@let.hokudai.ac.jp

応用倫理 ── 理論と実践の架橋　vol. 12
2021 年 3 月 25 日発行

ISSN 1883-0110

〒 060-0810
札幌市北区北 10 条西 7 丁目
北海道大学大学院文学研究院
応用倫理・応用哲学研究教育センター

Tel：011-706-4088
E-mail：caep@let.hokudai.ac.jp
URL：http://caep-hu.sakura.ne.jp/