

The "Responsibility" and "Accountability" of Moral Agent in the Information Environment

メタデータ	言語: jpn 出版者: 公開日: 2021-01-04 キーワード (Ja): キーワード (En): 作成者: 中村, 美智太郎 メールアドレス: 所属:
URL	https://doi.org/10.14945/00027845

情報環境における道徳的行為者の「責任」と「答責性」

The "Responsibility" and "Accountability" of Moral Agent in the Information Environment

中村 美智太郎¹

Michitaro NAKAMURA

(令和 2 年 11 月 30 日受理)

ABSTRACT

Information ethics is inseparable from the consideration of ethics and morality itself. I suggest that issues of information ethics need to be discussed within the framework of what it means to be human and what ethics in action is. From this discussion, it seems that we can derive some hints for today's human development from the perspective of information ethics. Based on these issues, this paper focuses on moral agents in the information environment and considers them thematically in the following steps. In section 1, I start with a brief description of moral agents, depict their moral structure and the premise of their morality, and highlight the issue of "ownership of responsibility" as a point of contention. In the second section, I reconsider the moral agent as a responsible subject and consider the issue of "personality" which is based on free will and obeys moral laws. The process of this study confirms the distinction between human and non-human entities, and we take up the issue of "robots" as a non-human entity. In this attempt, we introduce the perspective of categorizing "ethical agents" and capture the issue of "responsibility". The third section positions moral agents in the "information sphere" and clarifies the situation of the agents. From the analysis, I argue that the problem of moral agents can include the "morality of events". Section 4 points out the distinction between moral responsibility and moral accountability, and extends the possibility of moral agents in terms of accountability. Through the above discussion, I conclude by summarizing the issue of moral agents in the context of information ethics and suggesting the possibility of education in a pluralistic world of diverse ethics.

はじめに

情報倫理は、生命倫理や医療倫理、あるいは環境倫理、ビジネス倫理といったいわゆる応用倫理の中のひとつの領域に位置づけられることが多い。この位置づけを受け入れるとすれば、その他の倫理的なアプローチと同様に情報倫理の議論においても、行動規範についての道徳的な評価を考察するだけでなく、情報倫理の枠組みで論じられる問題にも「規範それ自体を問い直す」ことが含まれていることになる。従って、この文脈において情報倫理の課題は、社会

¹ 学校教育系列

や集団においてすでにある程度共有された規範を受け入れることや、そのために効率の良い教育方法を開発することにのみあるわけではない。むしろ、そうしたすでに存在している規範そのものを問い直すことにこそ、その代替しがたい役割があるとも言える。特に、インターネット上にその都度出現する新しいデジタル・ツールがもたらす諸問題を考察することの裏側には、規範そのものの再検討が含まれることになる。

この意味では、情報倫理は倫理や道德そのものを考察することとは不可分だということになる。この問題について、西垣（2007）は、グローバルな資本主義においては「デジタルICTは単にビジネス遂行の道具として用いられているだけではない」く、「今や人間の生産消費活動そのものが、デジタルICTの空間のなかで行われつつある」ため、「むしろ倫理や道德を、『情報』という視座から、根本的に『思想』として捉え直すことが急務なのである」（p.10）と指摘している。ただし西垣は、情報概念が多義的であるため共通概念として確立しておらず、また倫理や道德がローカルな文化に依存していることを理由に、こうした議論が「たやすいものではない」ことも同時に指摘する。根村（2017）もまた、情報社会において倫理が問題となるのはプライバシーや有害コンテンツ、知的財産権等といった社会問題として扱われる場合が「中心」になっているが、これらの個別問題だけが「情報社会における〈倫理〉の考察ではない」として、「情報社会の進展によって、人間社会のあり方がどう変化しつつあり、その変化が人間の〈倫理的な営み〉にどのような影響を与えつつあるのかなどを分析することもまた、情報社会における倫理について考える際の重要なテーマになりうる」（p.96）と述べる。これらの指摘は、情報倫理の問題が、人間とは何かを問うこと、また行為における倫理とはどのようにあるかといった枠組みにおいて議論される必要性を示唆している。そして、こうした議論からは、情報倫理の観点から今日的な人間形成のあり方への示唆を導くことができるように思われる。

こうした問題意識に基づき、本論文では、情報環境における道徳的行為者に焦点を当てて、次の手順で主題的に考察を行う。第一節では道徳的行為者についての端的な規定から出発し、その行為者における道徳性の前提と構造を描き、「責任の帰属」の問題を論点として浮かび上がらせる。続いて第二節では、道徳的行為者を責任主体として捉え直し、自由意志に基づきながら道徳法則に従う「人格」の問題を検討する。この検討のプロセスから人間と人間以外の存在を区別することが課題となることを確認し、人間以外の存在として「ロボット」を取り上げ、情報倫理上の課題と接合することを試みる。この試みにおいて、「倫理的なエージェント」を分類する視点を導入し、「責任」の問題を捉える。さらに第三節では、道徳的行為者を「情報圏」の中で位置付け、その行為者が置かれた状況を明らかにする。その分析から、道徳的行為者の問題には「出来事 of 道徳性」が含まれ得ることを指摘することで、問題を別の位相に移す必要があることをみる。これを踏まえ、第四節では、道徳的な責任と道徳的な答責性とが区別されることを指摘し、答責性の観点から道徳的行為者の可能性を拡張する。あわせて、このプロセスで道徳的行為者としてのロボットの限界と可能性を析出したい。以上の考察と検討を通じて、最後に情報倫理の文脈における道徳的行為者の問題を整理し、また多様な倫理が共存する多元主義的な世界における教育の可能性についての示唆を得る。

第1節 道徳的行為者における主体と責任

道徳的な振る舞いを「人間」に教えることは少なくとも形式的には可能であるし、その人間の行為を道徳的に良いか悪いかを判定する道徳的な評価も可能であると考えられることは、決して

特殊なことではない。小学校や中学校で教科化されもした「特別な教科道徳」(道徳科)では、子どもの道徳性は数値や評語等で評価されることはないが、教科書や教材に登場する人物の行為が道徳的であるか否かについての判断は、多かれ少なかれ関わることになるはずである。そのような判断は「道徳的評価」とみなして差し支えないものであり、この点では、子どもは道徳科の授業を通じてそうした道徳的評価について理解することになり、自らが道徳的評価の対象者にならなくても、評価そのものと無縁であるというわけでは決してない。道徳科は義務教育における「特別な教科」でもあるために、子どもは全員「道徳的評価者」になるということになる。道徳的評価者であるのは、何も子どもだけではない。私たち大人もまた、日常生活においては道徳的評価者になる。とりわけ、教師は自らも道徳的な評価者でありながら、同時に子どもが「適切な」道徳的評価者になれるよう支援する責務を負った存在である。

いずれにせよ、私たちが道徳的に評価を行うとき「道徳的行為者」(moral agent)における道徳性のあり方がまさに問題となる。「道徳的行為者」とは端的には「道徳的に善い、ないし道徳的に悪い行為を行う者」のことだが、この場合の「行為」とは、一般的な意味内容よりも厳密に捉えられて「ある出来事がある主体に帰属可能な行い」のことを指す。例えば、横に並びあって立っている主体Aと主体Bがいる。ある瞬間に、主体Bが主体Aの足を踏んだとする。この場合は、主体Aの足の痛みは、主体Aの足を踏んだ主体Bの「行為」に帰属する。このため、主体Aは主体Bに対して、足の痛みの責任を問うことができる。ところが実際は、主体Bの横には主体Cが立っていて、実はその主体Cが主体Bの体を意図して横に押した結果、主体Bが主体Aの足を踏むことになってしまったという場合は、主体Aは足の痛みの責任をBとCのどちらの主体に問うことになるだろうか。この一連の出来事だけで判断する場合は、主体Cに責任があると判断されるだろう。なぜなら、主体Bはたしかに主体Aの足を踏んだ当人ではあるが、その足を踏むことの原因を作ったのは主体Cだからである。つまり、この場合、主体Aにとっては責任の主体は主体Cの「行為」にあることになる。主体Aが主体Bの責任を問うことは、主体Bが避けることができたのに意図してその選択をしなかった場合など、別の条件が明らかになって初めて可能となるだろう。

このように、「道徳的行為者」とは、行為の「行為性」が前提され、かつその上でその行為主体の道徳性が問われる、という構造を持つことになる。この前提と構造を受け入れてはじめて、私たちは「責任がその主体に帰属するかどうか」を議論することができるようになる。

例えば、いわゆる「トロリー問題」(日本では「トロッコ問題」と呼ばれることが多い)として知られるテーマがある。この問題の骨子は、線路の分岐点に近づきつつある暴走したトロリーをそのまま走らせると5名が命を落とし、別の支線に進路を変えれば1名が命を落とす状況下で、どちらを選択することがより倫理的な選択になるか、というものである。これは、フット(1967)によって提案され、その後トムソン(1976)やカム(1989)らによってより広範囲に取り扱われた、いわゆるジレンマ問題のひとつであり、類似の問題はいくつも提案されている。よく似た構造の問題としては、古くは「カルネアデスの板」がある。難破した船から投げ出されたある一人の人間が船の板につかまるところ、別の人間が同じ板にすがりつこうとするが、二人がつかまると板が沈むと考えた最初の者がそれを拒否して結果的に水死させたことをめぐり問題である。古代ギリシアのカルネアデスは、裁判においてこの者の罪を問わず免責するという結末を与えているが、他者を助けて自らと共に命を落とすリスクを負うことが倫理的な選択か、他者の命を犠牲にしても自らの命が助かる選択肢を選ぶことが倫理的かという観

点に立って議論すれば、ジレンマ問題として把握することができる。同様に、*covid-19* をめぐる状況でも議論された、治療の対象者の重症レベルに応じて治療の優先度を決定して選択を行うトリアージの問題も、同様の問題構造を持つように思われる。この意味では、普遍的な問題を提起していると考えることができる。

このトロリー問題には一義的な解答が与えられるものではないが、その主体は「道徳的行為者」が想定されているとあって差し支えない。いずれかの選択肢を選択するとしても、あるいは困難に向き合いながらも別の方法を考えるにしても、その基準となるのは「その行為が道徳的行為であるか否か」を決定し、その結果を引き受ける責任主体となるであろうからである。

第2節 道徳的行為者は道徳法則にどのように従うか

このように道徳的行為者における主体性が問題にされる場合、その行為性が帰属される責任主体が同時に問題となっている。この責任主体をめぐる問題については、とりわけ近代以降に主題的に検討されている。例えば、カントが提供する「人格」概念もその一つである。カントは「善、あるいは悪は常に、意志が理性法則によって規定されて、あるものを自己の客体にする限りにおいて、意志に対する関係を意味している」ために「善あるいは悪は本来人格の行為に関係させられるもの」(Kant, 105) だと考える。「意志」はこの場合、「理性の規則を自己の行為(…)の動因とする能力」(ibid.) であって、「自由」なものでなければならない。カントにおいては、道徳的行為者の主体は、自然の法則から自由でありながら、自由意志に基づいて道徳法則に従っているかのように行為する「自律」的な主体でもある。このことは、道徳的行為者に「人格」を設定することで、人間にのみ理性に基づく行為を可能にしたことを意味する¹。このように、カントにおいては、人間以外のいかなる存在も、自由意志と自律性を持つ存在でない限り、道徳的行為者になることは原理的にできないことになる。このような位置づけは、人間存在を神に一層接近させ、他の動物等と区別されるその特異性を際立たせるという働きがある一方、「道徳法則にどのように従うか」という観点を導入することで、例えばロボットのような人間以外の存在が道徳的行為者になることについての視座を提供することにもなるように思われる。

そこで「ロボット」と比較することで、問題をさらに明瞭にしてみよう。人間と同様にロボットもまた、倫理的な原則の範囲に置かれる、とする設定が成り立つ。実際に、例えば日常世界にあふれるコンピュータの中にも、設計者の最初の命令の完全な範疇には入らず、自動的に学習を進めて判断領域を拡張し続けるものもある。こうしたコンピュータを内蔵したマシンとしてのロボットは、必ずしも人間の身体と類似した外見(例えば本田技研工業による ASIMO 等がその典型例である)を持たないものとして、すでに日常生活に深く入り込んでいる。

ただし、いわゆる「エンハンスメント」を可能にするようなロボットスーツをめぐる議論については、ここでは区別されることに注意が必要である。例えば、サイボーグ型ロボット「HAL (Hybrid Assistive Limb)」²は、身体に直接装着して装着者の意志を反映させることができる。このロボットスーツを装着することで、筋ジストロフィーなどの筋原性疾患を持った患者が自らの意志通りに身体を動作させることや、あるいは重量物を持ち上げる動作を行う際の身体への負荷を低減することなどが可能になり得る。HAL におけるこのような身体機能の拡張・増幅

は、筋肉を動かそうとするときに皮膚表面に発生する微弱な生体電位信号を読み取ることで実現しており、装着者自身がこの装置の仕組みを理解しているか否かに関わらず、装着者の意志の範囲内に収まるとみなして差し支えないものである³。このため、ロボットスーツそのものが道徳的行為者とみなされ得るかどうかは疑わしい。他方、ロボット概念には、ソフトウェアとハードウェアの双方が含まれることも同時にここで確認しておきたい。

さて、ロボットスーツのような装着型のロボットと区別した上で、このような「ロボット」についての設定の最も基礎的なものに SF 作家アイザック・アシモフが提案した「ロボット三原則」である。すなわち、「ロボットは人間に危害を加えてはならない。また何も手を下さずに人間が危害を受けるのを黙視してはならない。」「ロボットは人間の命令に従わなくてはならない。ただし第一原則に反する命令は、この限りではない。」「ロボットは自らの存在を守らなければならない。ただしそれは第一原則及び第二原則に反しない場合に限る。」の三原則である。

これらの原則は、道徳的な人間が持つ原則と同じ構造を持つようにみえる。試みにロボット三原則の「ロボット」と表現されている箇所を「人間」に置き換えてみれば、きわめて道徳的な人間の姿が表現されていることに気づくだろう。この場合、行為の主体は「道徳的行為者」になり、このように考えられるとすれば、直感的に、人工的な道徳的マシン＝ロボットであろうと人間であろうと、「道徳的行為者」になることができるということになる。

この可能性を追求するためには、人工的な道徳的マシンの性質をさらに検討する必要がある。この検討のために、ムーア (2009) の議論を参照してみよう。ムーアは「倫理的エージェント」を4つの種類に分類し、それらの階層化を試みている。第一のエージェントは、「倫理的な影響を与えるエージェント」である。このエージェントは、それが意図しているかどうかに関わらず、その行動が倫理的な結果をもたらすものである。あらゆるロボットは、その行動が人間に利益や危害をもたらす可能性がある限りにおいて、このエージェントになる。だから、もしデジタル時計が、その所有者に約束の時間を守るように促すとすれば、やはり倫理的な影響を与えるエージェントであるとみなしても良いことになる。第二のエージェントは、「暗黙の倫理的エージェント」である。これは、設計において倫理的配慮が組み込まれているものである。例えば、他の飛行機が衝突経路上に接近してきたときにパイロットに警告するための警告装置を備えた飛行機や、あるいは多数の犠牲者にスパムメールを送信するように設定されたウイルスに感染してゾンビ化したコンピュータなどがこれに当たる。第三のエージェントは「明示的な倫理エージェント」である。これは、単に倫理に従って行動するのではなく、倫理から行動していると考えられることができるエージェントであり、様々な状況に関する倫理情報を特定して処理し、何をすべきかについて敏感な判断を下すことができるため、倫理原則が矛盾している場合にも合理的な解決策を導き出すことができる。つまり、様々な種類の状況に合わせて調整・解釈される倫理的行動の一般的な原則・ルールを持つエージェントであると言える。そして、最後に第四のエージェント「完全な倫理エージェント」は、第三の「明示的な倫理のエージェント」と同様に、多様な状況について倫理的判断を行い、しかも多くの場合その判断に何らかの正当性を与えることができるが、第四の「完全な倫理のエージェント」は、道徳的行為者としての人間に帰属する特徴である「意識」(あるいは「意識のような特徴」)を持つ。

ムーアのこうした分析においてポイントとなるのは、これら四種類の倫理のエージェントは階層的なものであると指摘する点である。この分類に従えば、前述のアシモフが想定するロボ

ットは、行動する前に倫理的なガイドライン（「ロボット三原則」）を必ず参照するため第三の「明示的な倫理的エージェント」に該当することになる。「一見すると理にかなっているようにみえるが、徹底して適用しようとする場合に予想外の結果を生むことがある」としてアシモフの三原則を批判しながら、ムーアは「人間の意思決定者の能力に関する証拠を収集するのと同じように、ロボットの意思決定を人間の意思決定と比較したり、ロボットにその意思決定の正当化を求めたりすることで、ロボットの倫理的能力に関する証拠を収集することができる」と指摘し、「ロボットが人間の優れた倫理的意思決定者と同等かそれ以上の倫理的意思決定のための説得力のある正当化を与えることができれば、そのロボットの能力は倫理的意思決定のその分野について帰納的に確立される」と主張する。

ムーアが分析したこの第三の倫理のエージェントこそが、「道徳的行為者」としての人間とロボットとを分ける分水嶺となるように思われる。つまり、単に倫理に従うのではなく倫理から行動し、倫理をめぐる情報を特定して処理し、なすべきことについての判断を適切に下し、倫理原則が矛盾しても合理的な解決策を導き出すことができるかどうかが課題となる。このようなエージェントの設定とそのため判断基準は、カントが主張するような理想的な人間イメージと現実世界に生きる実際の人間イメージとの間にある距離への歩みを進めるものであるように思える。

だが、他方で、ロボットに倫理的な責任を負わせることができないという問題は残されたままである。ムーアも言及しているように、ロボットも人間と同様に褒めたり罰を与えたりして「学習」するかもしれないが、結局のところ「完全な倫理のエージェント」ではない限り、修理や再プログラミングによって対応する以外にないとなれば、人間がロボットの行動に責任を負い、将来的により良い行動ができるようにその責任を無限に持ち続けるしかないことになる。

第3節 遷移システムとしての道徳的行為者における「責任」の問題

ここで、改めて「責任」の問題について考察を行いたい。特に道徳的行為者に「責任が帰属される」ということはどのような事態なのか、そしてまた、責任は道徳的行為者に常に帰属する必要があるのかといった観点から、検討を進める。

これまでの議論のうち特に注目すべきは、情報倫理に対する包括的なアプローチを提供しているフロリディ（2007）の議論である。フロリディは、情報倫理の観点に立った「道徳的行為者」（moral agent）についての議論を包括的に取り扱っている。ここでは、フロリディの議論を参照することで、道徳的行為者をめぐる問題を情報倫理の問題圏に改めて移すことで、「道徳的行為者」をめぐる問題を照射したい。

さてフロリディによれば、与えられた状況において、自身にとって最適と思われる一連の行為を行おうとする場合、道徳的行為者は、「入力資源としての情報」を利用して「出力生産物としての情報」を生み出し、そのプロセスで「標的としての情報」である「情報環境」に影響を与えるという図式で把握される。こうした状況をめぐるこれまでの特に初期の議論においては「情報倫理の広大な領域を無視したり、たった一つの情報作用とこれに対応するマイクロ倫理だけがあたかも正しいアプローチを与えたりするように議論」されてきた傾向があり、道徳的行為者の行為とその評価、すなわち「情報環境への影響」は、「入力情報」と「出力情報」から「独立」して位置づけられてしまっている点を指摘している。これに対してフロリディは、これらの情報作用をすべて結びつけ、またそれらのサイクルが考慮された「情報圏」（infosphere）を

提案する。この「情報圏」は、道徳的行為者をめぐる三要素すべてが帰属されるような情報環境のことである。この提案に従えば、情報環境概念はマイクロなものではなく、よりエコロジカルなものとして捉えることが可能となり、「情報のライフサイクル全体を大局的に捉えることが可能であり、ゆえに他のマイクロ倫理的アプローチの限界を超えられる」(p.58)。ここに至って情報倫理は「マイクロ倫理」の束縛から逃れ、「マクロ倫理」となるのである。

この「マクロ倫理」において把握される情報圏に置かれた「道徳的行為者」は、フロリディ／サンダースによれば、「道徳的に正しい行為をなすような、相互行為的、自律的、かつ適応的な遷移システム」と規定され得る (Floridi and Sanders, 2004)。「システム」は属性値によって特徴づけられるものだが、ここでのポイントはシステムが「変化するシステム」である場合、この属性値も同様に変化するものであり、逆に属性値が変化すればシステムもまた変化するものであるという点である。つまり、そのシステムを持つ実態 (例えばここでは「行為者」) は、それぞれの時点において異なる「状態」を持つことになるといえる。この属性値が変化すれば、この「状態」もまたそれに対応して変化するために、システムは遷移的なものとなる。

この遷移システムにおける「変化」を「進化発展」と捉えてきたのが、伝統的な立場であると言えるが、ここではこのシステムにおける「自律性」に注目したい。この概念も「人格」概念と同様に人間性を規定する概念として歴史的に古いものである。例えば、プラトンは『パイドロス』において、ソクラテスに語らせるという形式をとりながら、「死すべき者」としての自己を探求することこそが、人間を「第一義的な主体」とする「自律性」としての自己の姿であることを示していた。三浦 (2013) によれば、ソクラテス以前においても「自律性」とは「不合理な宗教的運命論の頸木から逃れて、欲求であれ情欲であれ理性であれ、あるいはそれらの複合体であれ、ともかく自らが有する行為の内的起源に従って行為選択を行う、そのあり方自体」のことを指す。このように自律性を捉える際、人間は「行為決定や人生の目的としての幸福の実現について、神ではなく人間を第一義的な主体であり帰責される行為者」となる。いわば、「自律性」もまた責任主体としての規定を支える役割を果たしている。ただし、このような自律性は、近代以降のそれとは原理的に区別される。

再びカントの議論に戻ろう。カントは、自らの道徳法則に従うという意志におけるものとして自律性を把握している⁴。カントは『人倫の形而上学の基礎づけ』(1785)の中で「義務」概念を分析しながら、人間が従わなければならない道徳性の形態を、いわゆる「定言命法」(kategorischer Imperativ)であるとみなし [GMS:A421/63]、自律性はこの定言命法に従いながら意志の自由も確保するという形態をとることを示している。このような自律性の形態についてヘッフェ (1991) は、カントが「二重の問い」に回答したと位置付けている (p.208ff.)。すなわち、「すべての道徳的行為にとっての概念と最高の基準」については「定言命法」によって応え、「その概念と基準に従って行為できるための究極的な根拠」については「自己立法」としての「意志の自律」によって応えているとみなしている⁵。ヘッフェが「定言命法は自律的意志を存立させる概念と法則とを挙げ、自律は定言命法の要求を遂行することを可能にする」と整理しているように、たしかにカントにおいて、このふたつの観点は相互に関連しあい、道徳法則に従いながらも自由であるという形態の中で分かちがたく結びついている。

もし情報圏のうちに置かれた道徳的行為者がその「自律的」な遷移システムであるなら、道徳法則に従いながらも自由であるという、カント由来の自律性の範囲のうちにもあることになる。他方、もしこの道徳的行為者が、ロボットのような人間以外の存在を含むと考える場合、

問題は少し複雑になるだろう。フロリディらは、道徳的行為者が定義された通りの「遷移システム」であるのなら、当該行為の道徳的な答責性を負う限り、デジタル・エージェント・企業・人間機械混合システム等を含む「人工的な行為者」を道徳的行為者であるとみなすことができる。この意味で、道徳的行為者の範囲は拡大されることになる。道徳性には責任が含まれるが、責任の欠如した道徳的行為者を想定することは可能であるとフロリディは主張する。すなわち、責任のとれない道徳的行為者を識別しても虚しいので責任のとれる道徳的行為者だけを問題にする方が良いと考えるのではなく、責任の所在を問う問題とは独立して道徳的行為者を識別することが重要なポイントだ、と考えるのだ。フロリディらの挙げている例を引用すれば、行方不明者を探索し、しばしば人命救助に貢献する捜索救助犬としての訓練を受けた犬の例が典型的である。この犬は、たしかに遭難した人間を救助することで人命を救うという意味では道徳的な行為者となる。このとき、この犬自身は報酬として餌や賞賛を受けるかもしれないが、道徳的な責任を持つわけではない。たしかに、この犬が人命救助に失敗した場合や、人命救助のプロセスから逃走した場合、この犬に道徳的な責任を問うことにはならないと考える方が自然である。では、いかにしてこの犬は道徳的行為者であると考えられるのか。この犬を、道徳と関係のある「ゲーム」に参加した存在とみなし、その上で「出来事の道徳性」について説明が可能な存在であるとみなす、というやり方である。このようにすれば、この犬を道徳的行為者であるとみなすことができる、とフロリディらは考える。

第4節 「答責性」を持つ道徳的受容者の可能性

フロリディらの立場において重要なポイントは、道徳的な責任 (responsibility) と道徳的な答責性 (accountability) とが区別されている点にある。ここで「道徳的な責任」とは、その行為の意図や自覚、あるいは心情といったものが必要なものだと考えられる。例えば、自己立法した道徳法則に従って自ら道徳的な行為を実行することができる根拠とされるカントにおける「人格」も、まさにこの道徳的な行為における意図・自覚・心情を重視したものだった。そしてだからこそ、その他の動物と区別される特別な存在として人間を位置づけることが可能になったとも言える。ところが、これに対して道徳的な「答責性」には、それらに加えて「人格」も必要とはされない。フロリディらは「あらゆる規範的な言説は責任分析にきせられるべきだ」という「仮定」を「法的誤謬」として退け、「道徳的行為者をはっきり識別することが必要」と主張する。そして、当該の道徳的な行為者を識別することは、道徳的な「評価」と「等価」であるとみなし、道徳的行為者についての考察の次元を、「一次オーダーの道徳的行為」の源として考察することから、「二次オーダーの道徳的行為」の「受容者」として考察することへと転換する。すなわち、そのエージェントが道徳的な行為を行うことを意図しているのかといった問題は、そのエージェントが自分の行為に道徳的な責任を持つか否かを評価する段階においてのみ関連することになる、ということになる。

この「転換」についての例をフロリディらは挙げていないのだが、前節でみた遭難救助犬の例を再び取り上げて考察してみよう。遭難救助犬が行方不明者を探し出し、その人命を救助したとする。このときのこの犬の行為は、道徳的な責任から切り離された場合（そして実際に犬が道徳的な責任を負うことにはならない）、「人命を救助する」という道徳的行為の源となるのではなく、「遭難者の人命が救助された」という出来事を受容する受容者となる、ということだ。犬自身の一連の道徳的な行為は、遭難した行方不明者を救助することと、その行方不明者が救

助されることの少なくともふたつの面において構成されている。前者には動機や心情といった要因が関与し得るが、後者にはそれらは必ずしも要求されないという点で、区別される。こうして、この「道徳的行為の受容者」という点から、犬が道徳的行為者であることを考察することが可能になる。

ここでは、道徳的な責任を問題としなくても、道徳的行為者は存在することが可能となる、とフロリディは主張していることになる。道徳的な責任と道徳的な答責性とをこのように区別することは、原理的にはたしかに可能である。そして、もしこの区別が可能であるとすれば、道徳的行為者の概念は人間のみに限定されるのではなく、新しい行為者として拡張していくことができるようになる。この場合の「新しい行為者」にはロボットが、より精確にはムーアが分析した第三のエージェントとしての「明示的な倫理エージェント」が、含まれることになるだろう⁶。

だが、フロリディのこうした議論にはさらなる検討の余地がないわけではない。自由意志の取り扱いについての課題や、「区別」と「評価」を等値することをめぐる発展的な問題については、異論の余地もある⁷。そもそも、犬に道徳的な動機がないことを証明することは難しいという問題も残される。これらに加えて、特に道徳的行為者が人間を越えて「人工知能のシステム」にまで拡張される問題について考察するウォラック／アレン（2009/2019）が、フロリディらの議論の有効性を一定程度認めつつも、工学的な実装という面では限界があるとみなしていることも無視できない。ウォラック／アレンは「人工道徳的エージェント」（artificial moral agents : AMA）をめぐる問題を、とりわけ実装の観点から論じている。ウォラック／アレンによれば、宗教や価値観、あるいは関心といった伝統的に人間特有の倫理的な原理を構成してきた要素をマシンに適用することは容易ではないが、アシモフのロボット三原則のような、規則に基づく倫理体系であれば簡単にマシンに実装することができるように、一見すると見える。しかしながら、すでにムーアを媒介して検討したように、このような単純な三原則であっても、それほど簡単に倫理的なジレンマから脱出することは許されない。ウォラック／アレンは、心を持たないシステムを道徳的行為者とみなすフロリディらの「抽象化」を実装するケースを、「閾値関数」（それを上回れば道徳的だと考えられ、それを下回れば不道徳だと考えられる閾値）によって定義しようとするが、その実現可能性は難しいと判断している。例えば、彼らが挙げている例を引用すれば、「社会の人間の福利に貢献する」といった項目に従う行動を生成する定義として利用可能な閾値となる変数を想定することは極めて困難だという。

しかし、こうした困難を踏まえてもなお、道徳的行為者をめぐる議論は、社会システムの変更に常に向き合わざるを得ない時代に生きる私たちに対して、一定の示唆を与え得るように思える。特に、「答責性」という観点は、岡田（2012）が主張するような「弱いロボット」の可能性に応答するものであるとみなすことができる。岡田は、「すごいといわれるロボット技術からかなり距離がある」ロボットとして、『役に立つロボット』であることから降りてしまったような『弱い』ロボットを構想し、実装にまで至っている。この岡田が開発した「弱いロボット」には、「む〜」や「ゴミ箱ロボット」などがある。例えば「ゴミ箱ロボット」では、ゴミ箱のような姿をしたロボットが「とぼとぼ」と歩き、ゴミを見つけて腰をかがめて拾おうとする動作を見せるが、もともとゴミを拾うための手や腕を装備していないため、見つけたそのゴミ自体を拾うことはできない。そこで、拾うことを手伝ってくれる他者を周囲から探し出し、ゴミを拾ってもらい、そのゴミを自らに入れてもらう。他者にゴミを拾ってもらうためには、ロ

ボットの性能としては不完全であることと、その不完全であることを他者に伝えることが必要だという。このことは、「このロボットとの関わりを感じる『思わず助けようとしてしまう』という場」(岡田, 2017) を創り出すことを意味しているだろう。ここでのロボットの能力は、「共同行為を生みだすためのポイントは、みずからの状況を相手からも参照可能なように表示しておくこと」である。岡田の「ゴミ箱ロボット」は、たしかに道徳的行為者と言えるが、道徳的な責任を負うことなく、むしろ道徳的な答責性を持つ受容者として、道徳的行為を成立させている。ムーアの分類における「完全な倫理エージェント」になることない。むしろ階層的には下部に位置づけられる「暗黙の倫理的エージェント」あるいは「倫理的影響を与えるエージェント」に該当するかもしれない。それでも、ロボットによってむしろ人間の道徳的行為は惹起され、道徳的行為の共同行為者となることができる⁸。このことは、道徳的行為者の議論に別の地平を提供するだけでなく、情報環境を共同体として再構築していく時代に情報圏に住まう者が道徳的行為者として生きるための示唆を与えているように思われる。

おわりに

ここまで考察してきたことを改めて整理しながら、最後に一定の示唆を得たい。本論文では、主に道徳的行為者に焦点をあてながら考察を進めてきた。特にカントに代表されるような古典的な道徳をめぐる議論を参照しながら、情報圏における道徳的行為者における道徳的行為と道徳法則、責任と答責性を検討してきた。

近年、社会一般においても教育機関においても、サイバースペースや ICT 環境の整備が急速に進められている。日本においては、内閣府が提唱する「Society 5.0」⁹の時代への対応が求められている。この Society 5.0 それ自体については、議論や批判の余地がある。例えば、イノベーションを実現する営みやそれに関わる学問以外は排除される傾向にあるように見える。ある学問やその成果がイノベティブであるかどうかは一義的に決定することは多くの場合困難であることを考えれば、こうした傾向が継続され強化されれば、短期的な利益の獲得への注力が進められ、結果的にイノベーションを継続的に実現し続けることは困難に陥るだろう。さらに、特定の主体にとって有用とされるもの以外は認められないというような社会に陥ってしまうリスクもある。このような多様性やユニバーサルな環境を否定するような世界の到来は、長期的な視野に立った戦略的なイノベーションの実現だけでなく、その萌芽の保護、あるいは健全な社会の実現を目指す上では、避けなければならないと言える。

他方で、誰にとってもより幸福で生きやすい社会を実現していくためには、デジタル・ツールの利活用だけでなく、それらがもたらす様々な「差」を架橋していく必要がある。この課題に応答するためには、「情報」を倫理的に扱うことと、情報倫理が人間形成にどのようにアプローチするののかといった問題についての理解を深めることが欠かせない。この問題は、情報環境において道徳的行為者はいかなる存在かを検討することでもある。Society 5.0 の概念提案の際にも、先行する「Society 4.0」の弱点として「知識や情報が共有されず、分野横断的な連携が不十分」、「人が行う能力に限界があるため、あふれる情報から必要な情報を見つけて分析する作業が負担」、「年齢や障害などによる労働や行動範囲に」つきまとう「制約」等に「十分に対応することが困難」であることがすでに指摘されていた。こうした困難の対応する準備として、岡田が提唱するような「弱いロボット」の発想があるとみなすことができる。「Society 5.0」と呼称するかどうかは別として、これからの新しい社会ないし共同体の構築には、そうした弱点

を乗り越える枠組みを提供する必要がある。本論文で進めてきた道徳的行為者をめぐる問題の検討は、理性や人格、あるいは責任を負うことが可能な「強い」存在ではなく、むしろそれと平行に存在することができる「弱い」存在をめぐる枠組みの検討でもある。この視点は特に、どのようにロボットを「道徳的行為者」や「道徳的受容者」としてみなすかという問いの中で、あり得る、あるいはあるべき共同体の姿を構想する場合には、避けて通ることができない問題である。特に道徳的行為者のうち、ロボットなどの人間でないものにおける道徳性や道徳的行為を検討する際、その先には人間の共同体にそうした人間ではないものをどのように組み入れるか、また組み入れたいかという問題への眼差しが含まれるはずである。すなわち、ここで問われているのは、道徳的行為者という観点から、既存の共同体の変容をどのように設計し直すかという問題である。もちろんこの問いには一義的な回答はなく、文字通り「遷移的」なものになるだろう。次のようなハーバーマスの言葉が示唆的である。

つまり、われわれが、すべての人を同じに尊重し、すべての人に対して連帯の責任を負うことを要求する共同体のメンバーというステータスを持っていたいと思うかどうかは、必ずしも自明のことではないからである。われわれが道徳的に行為しなければならないという当為 (*sollen*) は、(義務論的に理解された) 道徳そのものの意味内容に含まれているにはちがいない。しかし、バイオ技術が類的存在としてのわれわれのアイデンティティを暗黙のうちに掘り崩しつつあるとすれば、なぜわれわれは道徳的でありたいと望まねば (*wollen*) ならないというのだろうか。道徳を全体としてどう評価するかということは、それ自身は道徳的判断に属することではなく、倫理的判断、いや類倫理的判断なのである。(ハーバーマス, 2012)

ハーバーマスはここでロボットや情報倫理、情報環境の問題を考察しているわけではなく、遺伝子工学の進歩による技術革新とそれに基づく着床前診断・妊娠中絶等の問題を考察しているのだが、情報倫理の視野から道徳的行為者の問題を考察してきた本論文における検討と強く響き合う。もちろん、例えば「ヒトが生まれる前にどのような存在であるか」というような問いは、本論文の関心を越えてはいるが、多元主義的な世界において、多様な倫理が共存する状況で、結果的に人間存在と共同体のありかを問い直すことになるという点では、「類倫理的判断」が要請されるとする課題意識を共有するものである。また、2019年度に「GIGA スクール構想」に関わる経費が補正予算案に盛り込まれ、日本の学校においては「GIGA スクール構想」が「Global and Innovation Gateway for All」として、すなわち「多様な子供たちを誰一人取り残すことなく、子供たち一人一人に公正に個別最適化され、資質・能力を一層確実に育成できる教育 ICT 環境の実現」を目指すものとして、位置づけられた。諸学校における教育もまた、これからの社会や共同体に対して情報倫理の思想から把握された課題や困難を克服を意図するものであるなら、子どもにとってそのような能力や資質を育成するような「情報圏」になることが求められていると言えるだろう。

参考文献

アイザック・アシモフ (小尾英佐訳) 『われはロボット』, 早川書房, 1983年。

Foot, Philippa: "The Problem of Abortion and the Doctrine of the Double Effect" in *Oxford Review*(5),

Oxford: Blackwell, 1967.

ルチアーノ・フロリディ（西垣通訳）「情報倫理の本質と範囲」, 西垣通・竹之内禎編著『情報倫理の思想』NTT出版, 2007年, pp.47-98.

Floridi, Luciano and J.W. Sanders: “On the Morality of Artificial Agents” in: *Minds and Machines*, 14(3), Netherlands: Kluwer Academic Publishers, 2004, pp.349-379.

ユルゲン・ハーバーマス（三島憲一訳）『人間の将来とバイオエシックス』法政大学出版局, 2012年。

オットフリート・ヘッフエ（藪木栄夫訳）『イマヌエル・カント』法政大学出版局, 1991年。

Kamm, Francis . “Harming Some to Save Others” in *Philosophical Studies*, 57 (3), Netherlands: Kluwer Academic Publishers, 1989, pp.227-260.

Kant, Immanuel: *Kritik der praktischen Vernunft*, Hamburg: Felix Meiner Verlag, 1990. 訳出に際しては, 次を参照した。深作守文訳「実践理性批判」, 『カント全集』第7巻, 岩波書店, 1965年。引用に際しては, アカデミー版のページ数を記す。

Kant, Immanuel: “Metaphysik der Sitten” in: *Kants gesammelte Schriften*, Bd. VI., begonnen von der Königlich Preußischen Akademie der Wissenschaften. Herausgeber: Paul Natorp. Berlin, 1914. 同書からの引用は, 以下「MS」と略記する。訳出に際しては, 次を参照した。吉澤傳三郎・尾田幸雄訳「人倫の形而上学」, 『カント全集』第11巻所収, 理想社, 1969年。アカデミー版の頁数に, 邦訳頁数を併置して示す。また, 傍点は原文に従う。ただし, 必要に応じて適宜変更を加えた。久木田水生・神崎宣次・佐々木拓『ロボットからの倫理学入門』名古屋大学出版会, 2017年。岡田美智男『弱いロボット』医学書院, 2012年。

岡田美智男『〈弱いロボット〉の思考——わたし・身体・コミュニケーション』講談社, 2017年。

坂田武久「カントの人格概念の意義と問題」, 『関西学院哲学研究年報』第36巻, 2002年, pp.87-98.

Thomson, Myrna Judith Jarvis: “Killing, Letting Die, and the Trolley Problem”, *The Monist*(59), Oxford: Oxford University Press, 1976, pp.204-217.

三浦要「自律的自己の起源について」, 金沢大学哲学・人間学研究会編『哲学・人間学論叢』第4巻, 2013年, pp.57-76.

Moor, James: “Four Kinds of Ethical Robots” in: *Philosophy*(72), 2009, pp.12-14.

中村美智太郎・藤井基貴（2017）「道徳教育における内容項目『自由』『自律』に関する基礎的研究」, 『静岡大学教育学部研究報告（教科教育学篇）』第48号, pp.75-87.

根村直美『ポストヒューマン・エシックス序説——サイバー・カルチャーの身体を問う』青弓社, 2017年。

日本経済団体連合会「Society5.0——ともに創造する未来」
<https://www.keidanren.or.jp/policy/society5.0.html> [最終閲覧日 2020年11月28日]

西垣通「普遍倫理への模索」, 西垣通・竹之内禎編著『情報倫理の思想』NTT出版, 2007年, pp.8-45.

ピーター・シンガー（戸田清訳）『動物の解放』人文書院, 2011年。

ウェンデル・ウォラック／コリン・アレン（岡本慎平・久木田水生訳）『ロボットに倫理を教える』名古屋大学出版会, 2019年。Wallach, Wendell and Colin Allen: *Moral Machines: Teaching Robots Right from Wrong*, Oxford: Oxford University Press, 2009.

註

¹ カントは実際、現象界に属す有限な存在でありながら、あらゆる自然の法則から自由である存在として、人間をより高次の存在として位置づける際に、人間以外の存在を「物件 (Sache)」として区別している。この問題については、例えば坂田武久 (2002) を参照のこと。

² HAL については、下記を参照のこと。 <https://robocare.jp/what-is-hal/> [最終閲覧日 2020 年 11 月 28 日]

³ ただし、だからといってもちろん重要性が低いというわけではない。特に後述する共同体の形成という視点に立って考察する場合や、同じ文脈でスポーツにおける身体性の問題を考察する場合に、一定の役割を果たすことが見込まれる。この問題については、稿を改めて論じたい。

⁴ なおこの自律性の問題については、筆者はすでに道徳教育との関連で論じている。下記を参照のこと。中村美智太郎・藤井基貴 (2017) 「道徳教育における内容項目「自由」「自律」に関する基礎的研究」、『静岡大学教育学部研究報告 (教科教育学篇)』第 48 号, pp.75-87.

⁵ なお、ヘッフェは、「自己立法」概念の起源をルソーの『社会契約論』第 1 章第 8 節にみている。ただし、「倫理学全体の基本原理を発見し、その基礎づけをしたのはカントがはじめてであった」としている (p.208ff.)

⁶ 他方、このことは実は、カントの「人格」概念と構造的にそれほど矛盾しないようにも見える。カントの「人格」もまた、現実に存在する個別の固有な名詞がつけられる人間そのものから切り離されて、いわば仮想的に立てられるものだからである。「仮想的」であることは、人間存在を原理的に二分している状況であるということを目指す。

⁷ 特に、人間以外に「道徳的受容者」概念を拡張していくという議論は、他方で動物倫理や環境倫理の問題と接点を持つことになることが予想される。人間以外の動物のうち、苦痛を感じる一部の動物のみが道徳的受容者であると考え、後に論争となっている。シンガーについては次を参照のこと。ピーター・シンガー (戸田清訳) 『動物の解放』人文書院, 2011 年。また、ロボットと動物の問題をめぐる整理については次を参照。久木田水生・神崎宣次・佐々木拓『ロボットからの倫理学入門』名古屋大学出版会, 2017 年。

⁸ 「弱いロボット」によって全員が道徳的行為を引き起こされるわけではもちろんない。岡田 (2012) によれば、場合によっては、特に子どもの攻撃の対象になることがあると報告されている。この場合、ロボットにおいても子どもにおいても道徳的行為は成立しないが、ロボットがお辞儀をする動作をみせると行動変容を起こすこともあることが指摘されている。

⁹ Society5.0 は、総合科学技術・イノベーション会議が策定した「第 5 期科学技術基本計画」で提案されたものである。日本経済団体連合会 (経団連) はこの Society5.0 を、経済界がデジタル革新によってもたらされる「革命的な変化」の主体となって実現していくとしている。日本経済団体連合会「Society 5.0——ともに創造する未来」

<https://www.keidanren.or.jp/policy/society5.0.html> [最終閲覧日 2020 年 11 月 28 日]