応用倫理―理論と実践の架橋―

Vol. 13
2022 年 6 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

今号に投稿された原稿については以下の方に査読を御願いいたしました。査読へのご協力に感謝いたします。

石原　孝二　　岡本　慎平　　奥田　太郎

金光　秀和　　神崎　宣次　　長門　裕介

本田　康二郎

（編集委員は除く）

研究ノート

科学技術をめぐる問題における哲学対話の役割――アップストリーム・エンゲージメントとしての哲学対話@ SCARTS の報告

原健一（北海道大学 高等教育推進機構 オープンエデュケーションセンター
科学技術コミュニケーション教育研究部門（CoSTEP））

要旨

本報告の目的は、2020 年に筆者が行なった哲学対話のワークショップの実践を振り返ることで、科学技術をめぐる問題について哲学的に対話することに期待される役割を提示することである。哲学対話においては、「心とは何か」といった抽象度の高い問いが適切なのだという考えがたびたび提示されてきた。このような考えに反して、報告者の行なった哲学対話のワークショップでは、科学技術をめぐる現実的で具体的な問題をテーマにした。このような最新の科学技術をめぐる問題について哲学的に対話することは可能なのか。そして、それには、科学技術コミュニケーションの手法の一つとしていかなる役割が期待されるのか。本報告では、哲学的手法を用いて科学技術をめぐる問題について対話を行なった実例（哲学対話カフェ@ SCARTS）を振り返る。この振り返りを通して、科学技術をめぐる哲学対話のワークショップの手法を紹介し、その特徴を挙げる。具体的には、(1) 参加者の主導性、(2) 科学の専門家の不在、(3) 問いの多様性、(4)「問いの共有」と「他（者）の思考に触れる」というアウトプット、これら四つの特徴を挙げる。そして、これらの特徴を踏まえた場合、科学技術をめぐる哲学対話には、研究開発の早期から多様な利害関係者や市民に議論の場を開く「アップストリーム・エンゲージメント」の役割が期待されることを示す。

Abstract

The purpose of this report is to present the expected role of philosophical dialogue in issues related to science and technology by reviewing the practice of a workshop on philosophical dialogue that I conducted in 2020. In philosophical dialogue, the idea has often been proposed that highly abstract questions such as “what is the mind?” are appropriate. Contrary to this idea, the workshop conducted by this author focused on practical and concrete issues related to science and technology. Is it possible to have a philosophical dialogue on such issues? And what role can this dialogue be expected to play as a method of communication in science and technology? In this report, we review an actual case (Philosophical Dialogue Cafe @

SCARTS）where we conducted a dialogue on issues surrounding science and technology using philosophical methods. Through this review, I will introduce the method of the workshop and list its characteristics. Specifically, the following four features are mentioned: (1) the initiative of the participants, (2) the absence of scientific experts, (3) the diversity of questions, and (4) the output of "sharing questions" and "exposure to other (people's) thinking". In light of these characteristics, we show that philosophical dialogue on science and technology is expected to play the role of "upstream engagement", i.e., to be an activity that opens up a forum for discussion to various stakeholders and people while research and development are still in their early stages.

1. はじめに

本稿では、2020 年に筆者が行なったワークショップ形式での哲学対話の実践を報告する。「ワークショップ」は「普段とは異なる視点から発想する、対話による学びと創造の方法」（安斎・塩瀬［2020：112］）と規定される。哲学対話のワークショップは、この「普段とは異なる視点」として、哲学的な手法を用いたものの見方、考え方を採用したワークショップであると言える。前例はすでにいくつも存在する。例えば、堀江（2017）では、「少人数の参加者（通常 5 ～ 8 人が適切な人数であるとされる）がグループになり、一定のルールと進行役によって進められる」哲学的な議論の手法を用いたワークショップが「ソクラティク・ダイアローグ」という名称で紹介されている（堀江［2017：vii］）。より多くの人の聞いたことのある哲学対話のイベント形式としては、「哲学カフェ[1]」が挙げられるだろう。サイエンスカフェは、この哲学カフェに着想を得て考案されたものであると言われる（三浦［2015：4］）。後に（第 4 節で）見るように、「哲学対話」と総称される対話のあり方は一枚岩ではなく、さまざまなタイプのものがあるのだが、サイエンスカフェも哲学カフェもともに、専門家と一般の人々がコーヒーやビールを片手に学問的な話題について気軽に話し合うことのできる場を提供してきた。

ところで、サイエンスカフェが科学や科学技術の問題を扱うのに対して、哲学対話においては、「心とは何か」「大人になるとはどういうことか」といった比較的抽象度の高い問いが適切だと考え

1　いわゆる「哲学カフェ」と哲学対話@ SCARTS には異なる特徴がいくつかある。例えば、鷲田［2014：xii-xxi］にある「哲学カフェ Q & A」では、セミナーやワークショップと哲学カフェとのちがいについて次のように述べられている。「哲学カフェでは、その場に集まった参加者が主役となって議論をつくっていきます。進行役がセミナー講師のように知識を提供するわけでも、ワークショップのように道筋（手順）やゴールが予め決まっているわけでもありません。哲学カフェは伝達の場ではなく、発見し、探求する場です」（鷲田［2014：xii］、傍点は引用者）。しかし、それと同時に、報告者が行なった哲学対話@ SCARTS にも、哲学カフェと同じような特徴が見受けられなくはない。より詳しい分析は本報告の第 4 節を参照されたい。

られる傾向があった[2]。このような考えに反して、2020 年に報告者の行なった二つの哲学対話のワークショップでは、「コンタクト・トレーシングアプリの導入は許されるか？」（オンライン哲学対話カフェ「監視社会と自由」2020 年 6 月 13 日開催。See CoSTEP_PR［2020f］）と「VR（ヴァーチャル・リアリティ）で死者を再現することは是か非か？」（哲学対話カフェ@ SCARTS、2020 年 10 月 21 日開催、以下「哲学対話@ SCARTS」と表記。See CoSTEP_PR［2020c］; CoSTEP_PR［2020d］）という最新の科学技術をめぐる問題をテーマにした。では、科学技術をめぐる現実的で具体的な問題について哲学的に対話することは可能なのか（問い①）。そして、それには、いわゆる「科学技術コミュニケーション」（「科学技術を一般市民に分かりやすく伝え、あるいは社会の問題意識を研究者･技術者の側にフィードバックするなど、研究者･技術者と社会の間のコミュニケーション」（総合科学技術会議［2006］））の手法の一つとして、どのような役割が期待されるのか（問い②）。これら二つの問いに対して、主に哲学対話@ SCARTS の報告を通じて応答することが本報告の目的である。

この目的を達成すべく、まず、哲学対話@ SCARTS の実践を振り返り、科学技術の問題を哲学対話において取り挙げる手法の一例を紹介する（第 2・3 節）。次いで、哲学対話@ SCARTS の特徴を列挙する（第 4 節）。そして、ここで挙げられた特徴を踏まえて、科学技術コミュニケーションの手法の一つとして、哲学対話に期待される役割を示す（第 5・6 節）。

本論に入る前に用語法について述べておく。本報告では、鷲田（2014）が紹介しているタイプの実践を「哲学カフェ」、堀江（2017）が紹介しているタイプの実践を「ソクラティク･ダイアローグ」、報告者が紹介しているタイプの実践を「哲学対話のワークショップ」（より簡略化して「哲学 WS」と表記）とそれぞれ呼ぶ。そして、これら三つのタイプの実践を総称して「哲学対話」と呼ぶこととする。また、報告者が 2020 年 10 月に SCARTS にて実践した個別のイベントのことを「哲学対話@ SCARTS」と呼ぶ。

2. 哲学対話@ SCARTS

北海道大学オープンエデュケーションセンター科学技術コミュニケーション教育研究部門（CoSTEP）は、2020 年 10 月 17 日から 30 日にかけて、CoSTEP の開講 15 周年を記念した「CoSTEP OPEN WEEK」を札幌文化芸術交流センター（SCARTS）との包括連携協定のもと開催した。CoSTEP OPEN WEEK においては、サイエンスカフェやアート作品の展示といった九つのイベントが開催された（CoSTEP_PR［2020a］; CoSTEP_PR［2020b］）。今回の報告の対象の一つである哲学対話@ SCARTS も、この CoSTEP OPEN WEEK のなかで行なわれたイベントの一つであった。

参加者の応募は CoSTEP サイト上で 10 月 8 日より開始した（CoSTEP_PR［2020c］）。テーマは「VR で死者を再現することは是か非か？」というものであった。はじめ、定員は 10 名で 13：00 ～ 15：

2　一例として、哲学対話に適した問いにかんする堀江［2017：30-35］の分析を参照されたい。そこでは「特定の専門領域や具体的な事実・情報を用いて答えうるような問い」は哲学対話に適した問いではないと言われている。そして、「日々の生活・社会生活において誰もが使っている言葉や概念、そこに潜んでおり、そこで前提とされている、わたしたちの共通の考え方・価値観・規範などに関する疑問を問いの形にしたもの」が哲学対話に適した問いとして提示されている（堀江［2017：34］）。本報告第 4 節の(4)も参照されたい。

00 と 17：00 ～ 19：00 の二回開催する予定であった。しかし、COVID-19（新型コロナウイルス感染症）の影響もあったためか、最終的に、17:00 ～ 19:00 の会に集まった 8 人の参加者で行われた。10 月 21 日に SCARTS のスタジオ 2 にて対面形式で開催された。

以下が当日の流れである。①自己紹介（10 分）、②ルールの確認（3 分）、③今回のテーマの導入（7 分）、④対話セッション（前半 40 分＋休憩 5 分＋後半 40 分）、⑤最後に一言（15 分）。順を追って説明していこう。

①　自己紹介（10 分）

まず自己紹介から始めた。これは、参加者だけではなくファシリテーター 2 名も行なったので、計 10 名が自己紹介したことになる。今回のテーマや会の趣旨などの説明の前に自己紹介を導入した理由は、はじめに対話のメインスピーカーが参加者であることを意識してもらうためであった。後ほど改めて指摘するが、哲学対話の重要な特徴の一つとして、参加者に発言の主導権があることが挙げられる。そのため、導入部から主催者が長く話すと、受け身で聞く姿勢が参加者に植え付けられ、対話への参加の積極性が失われるのではないかと報告者は考えた。自己紹介は、1 分程度という制限時間のもと、名前（今回の会におけるニックネーム）、そして何か一言という指示（「今回のテーマについて考えてみたいことや、なぜ今回哲学対話のイベントに参加しようと思ったのかを話してください」という指示）だけを与えた。

②　ルールの確認（3 分）

次に哲学 WS のルールを確認した。これらのルールは梶谷［2018：47］のルールを適宜変更しつつ使用したものである。詳細は以下の通りである。

○　思ったことを言葉にしてみましょう。
○　話を聞くだけでも大丈夫。
○　考えがまとまっていない段階で話してもよい。
○　場をしーんとさせてもよい。
○　途中で意見を変えてもよい。
○　思うように話せなかった場合でも謝らなくてよい。
○　強い言葉や人間を評価するような言葉をあまり使わない。
○　個人的な経験を無理に聞き出そうとしない。
○　政治宗教その他に関しては、全員が等しく自由です。
○　ハラスメントや迷惑行為になりうる言動は気を付けてください。主催が声をかけさせていただく場合があります。
○　途中入退室自由。
○　まぁお茶でも召し上がってくださいな。

以上のルールを、報告者が口頭で一つ一つ読み上げるかたちで参加者に伝えた。

③　今回のテーマの導入（7 分）

次に今回のテーマである科学技術による死者の再現をめぐる問題を導入した。今回のテーマの導入に際しては、CoSTEP が行なう映像演習で作成された作品『死者ともう一度会う』（制作：折登いずみ、西條未来）をもちいた[3]。作品内では、亡くなった娘を AI や VR を用いて再現することで母が娘と再会するという韓国のテレビ番組の事例や、2019 年の紅白歌合戦で美空ひばりが再現された事例を通して、「あなたは死者を再現したいですか？」という問いが投げかけられている。この映像の視聴を通して、「科学技術で死者を再現することは是か非か」という問題を導入した。

④　対話セッション（前半部 40 分＋休憩 5 分＋後半部 40 分）

上記の問題の導入ののちに対話セッションに入っていった。対話セッションは、前半の「問いについて問う」というパートと、後半の「問いを深める」という二つのパートから成る。対話セッションの紹介、振り返りは、科学技術をめぐる問題にかんする哲学対話の役割を考えるために重要であるので、次節以降で改めて扱う。

⑤　最後に一言（15 分）

一人 2 分を目処におこなった。ここでは、テーマにかんして得られた気づきはもちろん、哲学 WS という形式の対話の場についての感想も聞くことができた。参加者、そしてファシリテーターも一言ずつ述べて、会を終えた。

3. 対話セッション

上記④の対話セッションを振り返りたい。対話セッションにおいては以下の三つのアクティビティ（a ～ c）を行なった。

a）科学技術をめぐる話題から「哲学的な」問題を抽出する
b）抽出された問いをめぐってさらに疑問点を提示することで問いを再構成する
c）再構成された問いに対する解答の理由を言葉にしていくことで、さまざまな概念を再考する

a においては、まず、今回のテーマ「科学技術で死者を再現することは是か非か」をめぐって自分が考えたい哲学的な問題を参加者の方々に提示してもらった。「科学技術で死者を再現することは是か非か」という疑問文は複数の要素から成る。まず「再現する」という述部の要素がある。また、この述部に対して、「死者を」という対象、「科学技術で」という手段の要素が付与されている。そして、この「科学技術で死者を再現する」という事態に対する価値判断をめぐる「是／非」という要素がある。これらの諸要素から成る複雑な疑問文を目にして、これが、対話に取り組む端緒となる問いとしては複雑すぎるという印象を与えるかもしれない。だが、a は、物事を根本から問い直す問いを複雑な現実から抽出するという哲学的な手法を体験することを目指したものであった。例えば、この疑問文から、一つの要素に着目して、「死者とは何か？」とか、あるいは少数の要素を

3　映像は以下のサイトで視聴できる。CoSTEP_PR［2020b］

取り出し「死者を再現するとはどういうことか？」といった、よりシンプルな問いが抽出されることが期待される。

具体的には「今日あなたが考えてみたい哲学的な問いを紙に大きく書いてみてください」という指示を与え、A4 の紙に 6.00mm のサインペンで「○○とは何か？」といったように自分の考えたい問いを書いてもらうことを要求した。6.00mm という太めのサインペンで大きく書くように指示した理由は、書くことのできる文字数や内容を物理的に制限することによって、複雑な問いが出てくることを防ぐためであった。また、指示の中に「哲学的な問い」という言葉を入れたのは、参加者のもつ哲学のイメージが投影された問いが出てくることを期待したためである。つまり、「哲学的な（問い）」という限定を付すことによって、「そもそも○○とは何か？」といった物事を根本から問うようなシンプルな問いが出てくるように参加者を導くことができると考えたためであった。

次に、書いてもらった問いを、二人一組になって、隣の参加者に説明してもらった。その際、説明にかんする問いを一つ相手に投げかけるように聞き手に指示した。そして、一人ひとりに、自分の問いについて説明してもらった。「現実とは何か」「何をもって人は死んでいると言えるのか」「死後、人の意識は消えてなくなるのか」などさまざまな問いが提示された。

本報告では、紙幅の都合上、対話パート後半部で中心的に扱った「VR とはどのような表現手段か」という問いをめぐって b が行われた様子を紹介する。まず、「VR とはどのような表現手段か」という問いを提示した参加者に問いかけることによって、この問いの内実を具体化していった。その中で示された疑問は、VR による死者の再現は、他の表現手段による死者の再現と本質的に変わらないのではないかというものだった。死者について語るということを私たちは日常的に行なっている。例えば、葬式の場面を想像して欲しい。式場には故人を表象する遺影が置かれている。そして、その写真を眺めながら私たちは故人について色々なことを語り合う。あるいは、その故人の映る過去の映像を見て、私たちは彼／彼女について話す。こういったことと、VR で死者を再現することは同じことではないか。このように、他の表現手段との比較のもと、VR による死者の再現について考えを深めていった。

そして、考えを深めていく中で次の見解が提示された。仮に VR による死者の再現が他の表現手段による再現と本質的に変わらないのであれば、「VR によって死者を再現することは是か非か」という問題に対する答えは、VR による死者の再現の視聴者が騙されているのか否かという点に存するのではないか。例えば、手品師と詐欺師のちがいを考えてみればよい。前者は悪いことをしたとして責められることはないが、後者は悪人として責められうる。その理由は、前者においては、視聴者は自分が騙されていることを知っており、それに対して後者においては、視聴者が自分が騙されていると知らないからだ。先ほど出された美空ひばりの例は、紅白歌合戦の視聴者である私たちに、その美空ひばりが VR による再現であるということがわかるようになっていたのだから、それは手品の事例と同じものである。したがって、まったく問題はないという意見が出た。

次いで、c においては、「VR による死者の再現が他の表現手段による再現と本質的に変わらない場合、VR による死者の再現が許されるか否か」というように問いが定式化された。そして、この問いに対して、許される／許されないそれぞれの場合の理由や条件を明示してもらった。許される場合の条件は、上記の b の議論から提示された。すなわちそれは「再現されたものがフィクションだということが VR の制作者から視聴者にしっかりと告げられているならば（許される）」とい

う条件だった。

しかし、これに対して、「許されない」と考える参加者から、果たして視聴者に VR による再現がフィクションであると告げられたからといって死者の再現は許されるのだろうかという疑問が提示された。ここでは、VR による死者の再現が死者当人に危害を加える可能性が考慮されていない。特にこの事例は、自分について誤ったことが言われ流布されたとしても死者がそれに反論することができないという点で死者に不利である。このような意見が提示され吟味された。つまり、VR による死者の再現が許されない理由として、「死者は自分についての解釈を撤回できないから」というものが明示された。この理由の明示が哲学対話@ SCARTS の成果である。これがなぜ「成果」と言えるのかについては次節で検討する。ひとまず以上が対話セッションの内容である。

さて、本稿冒頭部で「科学技術をめぐる現実的で具体的な問題について哲学的に対話することは可能なのか」と問うた（問い①）。というのは、哲学対話においては、「心とは何か」「大人になるとはどういうことか」といった抽象的な問題が適切だと考えられる向きがあったからだ。しかし、本節の紹介から理解されるように、科学技術をめぐる問いは現実的で具体的でそして複雑なものであるとはいえ、その内には哲学的な問いとして再構成可能な内実が含まれている場合がある。したがって、そのような科学技術をめぐる複雑な問いからシンプルな哲学的問題を抽出することができるのであり、そのような活動を通して、哲学対話に適した問いを扱うことが可能である。

ここで重要なのは、こうして抽出されてきた哲学的な問いが、科学技術をめぐる問いとまったく無関係になってしまうのではない、ということである。「死者」や「現実」、あるいは「監視」といった概念について哲学的に考えることは、AI や VR、そしてコンタクト・トレーシングアプリといった最新の科学技術について語ることと常に関係づけられている。このようにして哲学的な問いを通じて対話を重ねることで、最新の科学技術についてより多くを知ろうとか、より詳しく知りたいと参加者が思うきっかけとすることができるかもしれない。もちろん、専門家がいない対話のワークショップにおいても、提示される科学技術の情報の正確さが担保されるべきことは言うまでもない。また、科学技術をめぐる問題について話す際に、専門的な知識がまったく役に立たないということでもない。そうした知識があるとより洗練された対話が可能になるだろう。しかし、そのような高いレベルで対話することがすぐには難しいという場合も多々ある。だから、より高度な、より洗練された議論へと至る契機として、哲学対話@ SCARTS では、参加者のプリミティブな違和感や疑念を「哲学的な問い」という形で言語化することによって、参加者が科学技術について話してみる、それについての自分の違和感や疑問を言葉にしてみるといったことを起点としたアプローチがとられたのである。

だが、科学技術をめぐる問題について哲学対話で考えることの利点は、当該の科学技術に対する参加者の関心を高めるという点にだけあるのだろうか。このことについて考えるべく、次に、哲学対話@ SCARTS の特徴を示そう。

4. 哲学対話@ SCARTS の特徴

では、上記の対話セッションを踏まえて、哲学対話@ SCARTS の特徴を四つ列挙しよう。その

際に、二つの観点から分析する。一つは、サイエンスカフェといった、科学技術をめぐる問題を扱う従来の対話の手法との違いは何かという観点、もう一つは、他のタイプの哲学対話との違いは何かという観点である[4]。

(1) 参加者の主導性

哲学対話@ SCARTS においては、イベント全体 120 分のうち、対話セッションにおける 80 分、最初の自己紹介と最後にある一人一言を入れると 105 分もの時間において、参加者に発言が求められていた。つまり、哲学対話@ SCARTS では全体のうち約 90％もの時間が参加者の話すことのできる時間に充てられている。この討論時間における参加者の発言できる時間の長さが哲学対話@ SCARTS の特徴の一つである。これは、科学の専門家を招き、話題を提供してもらう「講演会」（杉山［2020：7］；中村［2008：10］）のようなタイプのサイエンスカフェではなかなか見られない割合であろう。これが意味するのは、哲学対話@ SCARTS における対話の主導権は、ファシリテーターやゲストではなく、参加者にあったということである。

堀江［2017：56-57］でも「徹頭徹尾、対話を「参加者のもの」にする」という哲学対話の特徴が挙げられている。ただし、堀江［2017］で想定されている哲学対話のワークショップは、哲学対話@ SCARTS より圧倒的に長時間に及ぶものである。「私がヨーロッパで参加してきた SD〔ソクラティク・ダイアローグ〕では、90 分 1 セッションとして、合計 10 から 12 セッションを費やすのが標準的である。」（堀江［2017：73］）したがって、時間設定という面では、本稿が紹介する哲学 WS は哲学カフェと似ている。「〔哲学カフェが開催される〕時間は、週末などの休みの日に 2 ～ 3 時間行われるのが一般的です。」（鷲田［2014：xvi］）この 2 ～ 3 時間という制限時間内で発言の多くを参加者に求めるという点で哲学対話@ SCARTS は哲学カフェに近い。

(2) 科学の専門家の不在

多くのサイエンスカフェでは科学の専門家をゲストとして呼び、専門家の話題提供のもとに対話が進められるという形式がとられている（杉山［2020：7］；中村［2008：10］）。いわばそれは、専門家による洗練された話題提供に基づいた対話の場であろう。それに対して、哲学対話@ SCARTS においては、科学の専門化による話題の提供に大きく依拠することはなく、主として、一般の方々の分析と、それに基づいたファシリテーターの導きに依存して対話が進められていく。

堀江［2017：42-43］が紹介する哲学対話においても、参加者に「知的な権威に頼らず自分たちの洞察によってのみ考えを展開すること」が要求されている。「要するに、専門用語や有名な言葉を持ち込むことは、それがよく知られているものであっても、その意味内容に踏み込む前に、ある種の権威をもって人を黙らせてしまう。あるいは、その場の十分な理解ということに対して人を無責任にしてしまう。SD では、こうしたことがほぼ完全に払拭されるのである。」（堀江［2017：44-45］）また、哲学カフェにおいては「参加者、進行役ともに哲学者や哲学書に関する知識は必要ありません」（鷲田［2014:xiv］）とよりラディカルな非専門性が掲げられている。哲学対話@ SCARTS

4　以下の四つの特徴は、哲学対話@ SCARTS の前に行なった哲学対話「監視社会と自由～見える安心、見られる不安～」の後に見出されたものであり、この時点での報告としては、CoSTEP_PR［2020f］を参照されたい。

においては、ファシリテーターである報告者が哲学の専門家であるため、哲学カフェほどのラディカルな非専門性はない。とはいえ、AI や VR といった科学技術についての専門的知識がなければ成立しないような対話は行なっておらず、そうした知識を前提とした対話はできるだけなくしていこうという姿勢は哲学対話@ SCARTS においても貫かれていた。このような専門性にかんする姿勢は哲学対話の実践者に共通すると言えるだろう。

(3)　問いの多様性

「現実とは何か」「何をもって人は死んでいると言えるのか」「死後、人の意識は消えてなくなるのか」など、実にさまざまな問いが、哲学対話@ SCARTS においては提示されたことを先に見た。これらの問いは、サイエンスカフェなど科学の専門家を招いて行なうイベントの場合にはない多様性をもつものであろうと思われる（齋藤・戸田山［2011］）。もちろん最終的に少数の問いに絞ることにはなるものの、これらの問いのすべてに、その後の話題の中心となる可能性が、先に紹介したアクティビティ b を通して、吟味されるようになっていた。

哲学カフェにおける適切な問いは「（１）日常生活と関連すること、（２）誰もがそれについて考えることができること、（３）シンプルで根本的な問いであること」という三つが挙げられている（鷲田［2014：xvii］）。そこでも「「～とは何か」「なぜ～か」といった問いのかたちをとること」が初心者にお勧めされている（同上）。これらの特徴もやはり、専門家をゲストとして呼ぶサイエンスカフェ形式のイベントよりも、問いにかんする緩やかな価値基準を哲学 WS や哲学カフェが採用していることを示唆している。問いに対するこの寛容な姿勢は哲学対話に共通するように思われる[5]。

(4)「問いの共有」と「他（者）の思考に触れる」というアウトプット

哲学対話以外の対話の場・熟議の場においては、「目指すアウトプット」が「コンセンサス文書の作成と公開」「複数のアンケートによる参加者の意見変化を調査」「仮想的な評決」といった、具体的な政策を立案したり、是か非かを決するものである場合が多い（種村［2019：244-245］）。それに対して、哲学 WS で目指すアウトプットは問いの共有であり、この問いの共有を通して自分がふだん考えることのなかった他（者）の思考に触れることが主に目指されている。

同様のことは、鷲田［2014：xx］でも述べられている。「哲学カフェはいわゆる合意形成の場ではありません。哲学的にじっくり考えるためには、合意を目的とするよりはむしろ、参加者それぞれが、どの点が理解できて、どの点が理解できないのかをはっきりさせていくことが重要となりま

5　註２も参照されたい。

す。」[6] それに対して、堀江［2017：66］では、「何とか「答え」をひねり出す」ということが哲学 WS において求められると述べる。「進行役は参加者の発言を〔…〕常に「言明 statement」の形で定式化する」（堀江［2017:58］）ことによって、その場での答えを言語化するのである。とはいえ、堀江においても、「「答えの探求」において重要なものは、結果ではなく過程である。しかも、グループにおけるその場の対話としての、固有な過程なのである」と言われる（堀江［2017：70］）。

哲学対話@ SCARTS においても、「合意を目的とするよりはむしろ、参加者それぞれが、どの点が理解できて、どの点が理解できないのかをはっきりさせていく」という「グループにおけるその場の対話としての、固有な過程」が重要視されていた。「死者は自分についての解釈を撤回することができないから、VR による死者の再現は許されない」という理由の明示は、最初の問い「VR で死者を再現することは是か非か？」に対する決定的な答えとして参加者と共有されたわけではない。むしろ、この理由に至る過程で、死者という存在者の地位が再考されたという点が重要なのである。実際、対話のはじめにおいては、VR の鑑賞者に対する危害を考える参加者が多かった。その理由はおそらく、危害をもはや受けえない存在者として死者が扱われていたからだと思われる。しかし、今回の対話を通して得られた「死者は自分についての解釈を撤回することができないから」という理由の明示によって、死者が危害を受けうる存在者として位置づけ直される可能性が開かれた。この意味では、やはり、哲学対話@ SCARTS においても、合意を得ることではなく、「答えを探求する」過程、そしてその過程ではじめは考えにくかった可能性へと思考が開かれていったことが重要である[7]。

5. アップストリーム・エンゲージメントとしての哲学対話

さて、以上の特徴をもつ哲学対話@ SCARTS の試みは科学技術コミュニケーションの手法の一つとしていかなる役割を担いうるのだろうか（問い②）。この問いに応答すべく、既存の科学技術

6　越門［2014：37-38］も参照されたい。「〔哲学カフェの〕議論の目的は、問題解決や合意形成にはない。この点が、たとえば地方行政に関して住民間でなされる意思決定のための討議と異なる点である。そこでは特定の事柄の是非について参加者の総意をまとめ上げ、それを問題解決につなげることが求められる。しかし、哲学カフェでは、見解の相違を積極的に捉えつつ、問題を解決・解消するというよりむしろ、問題の深さ、つまりは解き難さを確認するところに、主眼がある。異論を説き伏せることも、哲学カフェの趣旨からは外れる。この点で、勝ち負けを競うディベートと異なる。参加者に求められるのは、何よりも異論にしっかりと耳を傾け、その真意を推し量ることである。異なる意見をもった他者との出会いによってこそ、個々の参加者は議論から成果を得ることができるからである。ただし、合意形成や異論の論破が目的ではないとは言っても、哲学カフェは、単なるおしゃべりや放談でもない。発言者は、言葉を選び、論理的に、理由や根拠も提示しつつ、自分の考えを他者に伝える義務を負う。他の参加者は、表明された意見と自分の考えとのすり合わせを試み、合意が可能かどうかを検討する。哲学カフェの議論は、あくまで真実を見いだすための対話である。それゆえ、こうした態度が要求されるのである。」

7　種村［2019：243-247］では、「ミニ・パブリックスの形成方法と参加人数」、「情報提供の方法」、「小グループでの討論人数」、「フォーラムでの討論時間」、「目指すアウトプット」という五つの項目を用いて、コンセンサス会議、討論型世論調査、討論劇の特徴が示されている。以下ではこれを参考にコンセンサス会議と哲学対話@ SCARTS のちがいを手短に示す（一部第 4 説と内容が重複している）。まず、コンセンサス会議では、公募と属性の抽出によって 15 ～ 20 人のミニ・パブリクスが形成される。それに対して、哲学対話@ SCARTS は、フォームによる自由申し込み制であり、ミニ・パブリクスが形成されているとは言えない。また、コンセンサス会議は情報提供と専門家との質疑を通して知見を提供する。それに対して、哲学対話@ SCARTS は、専門家との質疑応答は含まれない。さらに、哲学対話@ SCARTS の討論の人数は 10 人程度であり、コンセンサス会議よりも少なかった。そして、フォーラムでの討論時間は、コンセンサス会議が 3 ～ 4 日間に及ぶものであるのに対して、哲学対話@ SCARTS は 2 時間であった。この 2 時間の内訳としては、全体のうち約 90％もの時間が参加者の話せる時間に充てられており、この討論時間の割合の多さはコンセンサス会議には見られないものである。またコンセンサス会議においては「コンセンサス文書の作成と公開」がアウトプットとして目指されているのに対して、哲学対話@ SCARTS で目指すアウトプットは「問いの共有」「問いの分析」「概念の再考」であった。

コミュニケーションに対する批判を一つ見ていきたい。齋藤･戸田山［2011］は次のように述べる。

> 近年、サイエンスカフェなど科学コミュニケーションに双方向性を取り入れることが推奨されてきた。しかし多くの場合、トピックスの選定は専門家に委ねられており、専門家の話題提供の後に一般の人々が質問するという構造が固定しているように思われる。科学について知るべき情報を予め専門家が設定するという点で、欠如モデル型のコミュニケーションに対する反省が十分に生かされていない。科学コミュニケーションの双方向性をより高めるには、トピック設定の権限を専門家と一般の人々の双方に与えることが有効ではないだろうか。（齋藤・戸田山［2011：12］）

もちろん、上記の分析から 10 年ほど経過する中で、サイエンスカフェにおいて、双方向的なコミュニケーションを実現するさまざまな対話の手法・場が創造されてきた。それゆえ、上記の批判が現在のサイエンスカフェ一般に当てはまるのか疑問の余地がある[8]。しかし、ここで言われている「トピック設定の権限を専門家と一般の人々の双方に与える」という提案にはいまだ見るべきところがあるように思われる。すなわち、（齋藤・戸田山［2011］で想定されている）サイエンスカフェのような洗練された対話の場における双方向的なコミュニケーションにおいては「欠如モデル型のコミュニケーション」が再生産される可能性がある。というのも、対話の場でどのような話題を扱うべきかを科学者・技術者が決めている以上、この対話の内容には、専門家による価値判断が色濃く反映されているからだ。このような専門家によるトピック選定に基づく専門家と非専門家のあいだのコミュニケーションに対して、齋藤・戸田山［2011］では、このトピックの設定を市民が行なう場が提案されているのである。

さて、このような市民が主体となったトピックの選定の意義を「アップストリーム・エンゲージメント」との関連で位置づけることができるかもしれない。アップストリーム・エンゲージメント（「上流での参加」と訳される）とは、「研究開発のステップの早期にあるうちに多様な利害関係者や市民に議論の場を開いていく」（杉山［2020：5］）試みを指す。このような試みの必要性は、以下のオルティア［2015：84］の言葉にも表れていると言える。

> 「〔…〕対立する議論の最初の段階から、議論に参加する人すべてに考えを述べ、人の話を聞き、敬意をもって接することができる機会を提供することが、対話の可能性を広げ、議論をよい方向へと向かわせ、共通の土台を見つけることになるのである。」

確かに、例えば哲学対話@ SCARTS のもつ特徴 2（科学の専門家の不在）をめぐって次のような疑念がありうる。それは、最新の科学技術について専門家のいない場で素人が対話することに意

8　確かに賛否があるとはいえ、杉山［2020：7］でも同じことが指摘されている。「サイエンスカフェでは「双方向性」が重視される。研究の社会的意義や倫理的・文化的・政治的影響などについて市民が問題提起し、その問題をともに考えてくれる専門家の選出にあたっても市民の意向が反映され、そのうえで両者の対話・相互理解を促進しようとする。それが双方向性の意味である。市民が活発に質問したり意見を述べたりすれば「双方向的」というわけではない。／しかし、この意味での双方向性がサイエンスカフェで実現している事例はほぼ皆無だとされる。」三浦［2015：4-5］も参照されたい。

味はあるのかといった疑念である。最新の科学技術には専門家の説明があって初めて理解できる事柄と、それらの事柄から生じる問題がある。とすれば、専門家が不在の場で最新の科学技術について語ることには、科学技術コミュニケーションとしての実質的な成果は期待できないのではないか。しかし、そういった見解に反して、齋藤・戸田山［2011］の指摘は、専門家が参与する以前の段階における、市民の側からの主体的な問題提起を通じた、良好な科学技術コミュニケーションがありうることを示唆しており、アップストリーム・エンゲージメントとしての科学技術コミュニケーションが現に求められている。そして、哲学対話@ SCARTS の特徴（1 ～ 4）を踏まえると、たとえ専門家がその場にいなかったとしても、科学技術コミュニケーションの実践の一つとして果たすことのできる役割がそれにはあるように思われる。すなわち、参加者主導による多様な問題提起を通じた対話を促進するアップストリーム・エンゲージメントの試みの一つとして科学技術にかんする哲学対話は科学技術コミュニケーションに貢献するかもしれない。

6. 結　論

本稿冒頭で次の二つの問いが提起されていた。

問い①：科学技術をめぐる現実的で具体的な問題について哲学的に対話することは可能なのか。

問い②：科学技術コミュニケーションとして、哲学対話にどのような役割が期待されるのか。

まず、問い①に対しては、「可能である」という答えを本稿は（特に第3節にて）提示した。なるほど、科学技術をめぐる問題は具体的で現実的、そして複雑なものであり、それに対して、哲学対話で扱う問いには抽象度が高いシンプルなものが多い。しかし、科学技術をめぐる問いの内に含まれているシンプルな哲学的問題を抽出することができる。そのことを特に科学技術による死者の再現を例に本稿では示した。そして、そこで提示された哲学的な問いは科学技術をめぐる問いと無関係になってしまうのではない。参加者の違和感や疑念を「哲学的な問い」という形で言語化することによって、参加者である市民の方々が科学技術について話してみる、それについての自分の違和感や疑問を言葉にしてみるといったができる。そして、このような哲学的な問いの言語化を通じて、科学技術への関心を高めることや、アップストリーム・エンゲージメントとしての活動を実践できることが示唆された。

このアップストリーム・エンゲージメントとしての役割が期待されるということが問い②に対して本報告が（特に第5節にて）提示した答えとなる。アップストリーム・エンゲージメントとは「研究開発のステップの早期にあるうちに多様な利害関係者や市民に議論の場を開いていく」という試みであった。哲学対話は市民の側からの多様な問題提起を促進することによって、この試みを実践する科学技術コミュニケーションの活動の一例として位置づけられる。

最後に、科学技術の問題をめぐる哲学対話に対する一つの懸念を検討したい。それは哲学対話@ SCARTS の特徴 3（問いの多様性）をめぐる懸念である。すなわち、哲学対話@ SCARTS が扱う多種多様な問いのうちに、異なるステークホルダー間の対話を促進するのではなく、むしろ、阻害する問いがあるのではないか。例えば、AI や VR による死者の再現をめぐって「死後、人の意識は消えてなくなるのか」という問いが提示されていた。この問いは、AI や VR の専門家は通常扱わない問いである。あるいは、この問いは科学という営みの射程を越えたものであると考える科

学者がいるだろう。にもかかわらず、この問いを「よい問い」として扱う（かもしれない）哲学対話は、科学という営みの本質を見誤らせたり、科学に対する不信感をむやみに掻き立てたりするものと科学者・技術者に映るのではないか。その場合、科学者・技術者と市民との距離はいっそう広がってしまうのではないか。このような指摘を報告者は各所で受けてきた。

しかし、「死後、人の意識は消えてなくなるのか」という問いを意味のない問いと判断して、その内実に耳を傾けないことは、早計に過ぎるとも言える。実際、この問いが意味のない問いであると少なくとも哲学者は即座には判断しないだろう。「死後、人の意識は消えてなくなるのか」という問いは哲学の伝統的な問いの一つである。例えば、死後をめぐる問題は心身問題という哲学の問題とかかわりが深い。この心身問題という文脈で哲学史を眺めてみるならば、プラトンやデカルトなど、多くの哲学者が死後をめぐる問題に名を連ねることとなるだろうし、心の哲学という現代哲学の一分野も関わってくる。また、ポピュラーな一般向けの哲学書で死をめぐる著作が話題を集めるなど、死をめぐる哲学的問題は現代でも一般的な関心を集めているようだ（ケーガン［2019］；伊佐敷［2019］;内藤［2019］）。このように、科学者がナンセンスだとして斥ける問いだからといって、その問いがより広い文脈においても意味のない問いであるとは限らないし、少なくとも上記の問いにかんしては、それが意味のない問いであることを示すのにも少なからぬ（哲学的な）議論を要する。

また、こうした物事を根本から考える問いに、私たちの思考を促すという一定の効果があることが指摘されてもいる。安斎･塩瀬［2020:74-7］では、このような思考法を「哲学的思考法」と呼び、その効果について以下のように述べている。

> 「問題解決の場面において最も恐るべきことは、視野狭窄になり、中長期的な視点や深く考える思考態度を失ってしまうことです。視野を広げ、深め、問題の本質に迫っていく上で、哲学的思考は欠かせません。」（安斎・塩瀬［2020：74］）

先に述べたように、科学技術をめぐる哲学対話はアップストリーム・エンゲージメントの実践として位置づけられる可能性をもつ。とすれば、「研究開発のステップの早期にあるうちに多様な利害関係者や市民に議論の場を開いていく」というアップストリーム・エンゲージメントの目的の達成に、「死後、人の意識は消えてなくなるのか」といった種類の哲学的な問いについて考えることが貢献することもありえよう。科学技術をめぐる問題について「視野狭窄になり、中長期的な視点や深く考える思考態度を失ってしまうこと」が科学者・技術者に生じているならば、なおさら、このような問いが効果を発揮することも考えられる。

さらに、仮にステークホルダー間に不和が生じるとしても、そこから即座にその不和を明示する対話の試みを忌避すべきだということにはならないだろう。というのは、その不和は無視すべきものではなく、解消すべきものであるということもありえるからだ。むしろ、参加者の主体性を重視する哲学対話のアクティビティを通して、問いに対するステークホルダー間の価値観のちがいが明示されることによって、私たちが乗り越えるべき認識の不和とその要因を対象化・具体化することに、哲学対話が貢献するといったことも考えられる。あるいは、専門家では考えつかなかった問題や疑問が参加者の側から提示され、ひいては、科学技術をめぐる問題を専門家だけで話していたの

では得られなかった視点から再考するきっかけとなるといったことも考えられるだろう。

科学技術の問題に潜む哲学的な問いを考えるという活動の可能性は本報告で言い尽くされたわけではあるまい。この可能性を探る実践として、報告者は、討論劇といった他の実践との関連で哲学対話を利用するといったことを試みている最中である。討論劇の脚本を執筆する脚本家や、討論劇を演じる役者に哲学対話に参加してもらい、科学技術についての哲学的な対話を通じて、脚本の改稿や演技のパフォーマンスの向上につなげられないかという道を探る実践である。こういった実践にとどまらず、哲学対話の活動が科学技術コミュニケーションの好例となる可能性を今後も追及していきたい。

参考文献

安斎勇樹・塩瀬隆之（2020）『問いのデザイン ―― 創造的対話のファシリテーション』、学芸出版社

CoSTEP_PR（2020a）「CoSTEP OPEN WEEK: 札幌で出会う科学技術コミュニケーションを開催！」, https://costep.open-ed.hokudai.ac.jp/news/11535、最終閲覧日 2021 年 4 月 10 日

CoSTEP_PR（2020b）「CoSTEP OPEN WEEK:札幌で出会う科学技術コミュニケーションを開催しました」, https://costep.open-ed.hokudai.ac.jp/news/11712、最終閲覧日 2021 年 4 月 10 日

CoSTEP_PR（2020c）「10/21（水）哲学対話カフェ「死者を VR（ヴァーチャル・リアリティ）で再現することは是か非か？」を開催します」, https://costep.open-ed.hokudai.ac.jp/event/11573、最終閲覧日 2021 年 4 月 10 日

CoSTEP_PR（2020d）「哲学対話カフェ@ SCARTS を開催しました」, https://costep.open-ed.hokudai.ac.jp/news/11698、最終閲覧日 2021 年 4 月 10 日

CoSTEP_PR（2020e）「映像表現演習で CoSTEP 受講生が制作した動画を公開」, https://costep.open-ed.hokudai.ac.jp/news/11794、最終閲覧日 2021 年 4 月 10 日

CoSTEP_PR（2020f）「オンライン哲学対話カフェ「監視社会と自由 ―― 見える安心、見られる不安」を開催」, https://costep.open-ed.hokudai.ac.jp/news/11388、最終閲覧日 2021 年 4 月 10 日

堀江剛（2017）『ソクラティク・ダイアローグ ―― 対話の哲学に向けて』、大阪大学出版会

伊佐敷隆弘（2019）『死んだらどうなるのか？―― 死生観をめぐる 6 つの哲学』、亜紀書房

齋藤芳子・戸田山和久（2011）「非専門家の問いの特徴は何か？ それは専門家の眼にどう映るか？」、『科学技術コミュニケーション』、第 10 号、3-15 頁

ケーガン、シェリー（2019）『「死」とは何か イェール大学で 23 年連続の人気講義 完全翻訳版』、柴田裕之訳、文響社

梶谷真司（2018）『考えるとはどういうことか ―― 0 歳から 100 歳までの哲学入門』、幻冬舎新書

越門勝彦（2014）「哲学の共同実践としての対話 ―― 「哲学カフェ」の意義についての一考察」、『人文社会科学論叢』、第 23 号、35-45 頁

河野哲也編（2020）『ゼロからはじめる哲学対話 ―― 哲学プラクティス・ハンドブック』、ひつじ書房

三浦隆宏（2015）「「私たち」という感覚を育むために ―― 哲学カフェとシティズンシップ」、『臨床哲学』、第 16 巻、3-22 頁

内藤理恵子（2019）『誰も教えてくれなかった「死」の哲学入門』、日本実業出版社

中村征樹（2008）「サイエンスカフェ ―― 現状と課題」、『科学技術社会論研究』、第 5 巻、31-34 頁

オルティア、A・リンディ（2015）「科学技術に反対する市民とともに」、高梨直紘訳、ジョン・K・ギルバート／スーザン・ストルックマイヤー編『現代の事例から学ぶサイエンスコミュニケーション ―― 科学

技術と社会とのかかわり，その課題とジレンマ』、小川和義･加納圭･常見俊直監訳、慶應義塾大学出版、第 5 章、73-90 頁

総合科学技術会議（2006）『科学技術基本計画（第 3 次）』，https://www8.cao.go.jp/cstp/kihonkeikaku/honbun.pdf、最終閲覧日 2021 年 4 月 10 日

杉山滋郎（2020）「科学コミュニケーション」、藤垣裕子編『科学技術社会論の挑戦 2　科学技術と社会 ―― 具体的課題群』、東京大学出版会、第 1 章、1-24 頁

種村剛（2019）「先端科学技術の社会実装についての熟議の場 ―― 討論劇を用いた科学技術コミュニケーションを事例として」、『年報』、中央大学社会科学研究所、第 23 号、233-50 頁

鷲田清一監修（2014）『哲学カフェのつくりかた』、大阪大学出版会

「特集　ロボット倫理とAI倫理の現在」

研究ノート

ロボットの道徳的地位をめぐる近年の議論 ―― 道徳的なものの概念についての中間所見

猪ノ原次郎（北海道大学大学院文学院）

要旨

本稿は〈ロボットの道徳的地位〉という近年の倫理学上の話題のサーヴェイおよび予備的批判的検討を行う。第一に、AI／ロボット倫理および機械倫理の大まかな像を導入し、当該の話題をその中に位置づける。ロボットの道徳的地位の観念は、当初はロボットの道徳的行為者性の観点から提起された ―― 我々は（どのように）ロボット自身に責任を帰属させることができるか、との問いである。しかし、やがて問いを反転し、ロボットの道徳的被行為者性を問う者も現れた ―― ロボットへの道徳的配慮ないし「ロボットの権利」の問題である。第二に、この話題をめぐる議論の転回点をなした「関係論的アプローチ」に焦点を当て、またその派生形態も確認する。この種の考え方は、ロボットの（潜在的な）有感性や意識等に関する存在論的事実を求める望み薄な探究を回避しつ、間接的な方法 ―― ロボットとの相互作用の経験や、そうした経験が個人や共同体の道徳的生に与える影響を重視する ―― によってロボットの地位を規定することを企図している。最後に、ロボットの道徳的地位に対するアプローチ各々の陰伏的な理論的地盤を精査する。それにより、各々のポジションの特徴を際立たせ、この（一見して）奇妙な話題の基本図式を明確に理解できるようにしたい。

Abstract

There is a considerable amount of contemporary literature in ethics which discuss the moral status of robots and AI, or “robot rights”. It seems natural to wonder what the significance of this topic, and whether it could stand as a proper ethical/philosophical issue at all. This paper does not mean to directly answer to the question. Instead, as a preliminary sketch, I take up and clarify some of prominent approaches in that topic, and try to give an overview of the controversy occurring there.

First, I introduce a broader picture of AI/robot ethics and machine ethics, and situate the topic of moral status of robots into this picture. The idea of robot’s moral status/standing was initially proposed in terms of the moral agency of robots: (how) can we ascribe moral

responsibility to a robot itself? But some ethicists in turn inquire the moral *patiency* of robots, that is, moral considerations on robots or "robot rights". Secondly, I focus on the "relational approach" that constituted turning point of the debate on the topic and track its derived form of applying virtue ethics. This type of thinking is designed to determine robot's status by indirect, mediated ways which emphasize our experience of interacting with robots and how it effects moral life of individual or community, while detouring hopeless inquiry into ontological fact of robot's (potential) sentience, consciousness and so on. Finally, I scrutinize implicit theoretical grounds of each approach concerning moral status of robots, so that we can sherpen their position and clearly see configuration of present controversy about this superficially odd topic.

本ノートはAI・ロボット倫理学のある話題を対象とした覚書である。この話題を扱うここ十数年の文献中から主な潮流をサーヴェイして幾つかのポジションを明確にし、議論の要点を理解しようと試みているが、作業は未だ十分に成功していない。それゆえ以下は必ずしも文献の正確な紹介ではないし、何事かを論証しているわけでもない、覚書に過ぎない。もちろん、そこに含まれる哲学的誤りの責任は私に帰属する。

*

1.

人工知能やロボットをめぐる現在の倫理学分野の議論は、大きく分けて二つの場をもっていると言える[1]。一つは、現在あるいは至近未来の（現時点で技術的ハードルをおおよそクリアしている）AI・ロボットが引き起こす問題にどう対処するか、それらをどううまく利用するか、具体的には、研究・開発・利用の規制やガイドラインをどう策定するか、といった当面の課題であり、法学や政

1　稲葉他（2020: 357-8）における稲葉氏の発言を参照。

策科学（そして政策そのもの）と踵を接している[2]。もう一つは、将来的に「高度に自律的」な「人工知能」や「ロボット」ができてしまったら、要するに動物と人間の中間者か人間と天使の中間者（それを突き抜けた「神」？）ができてしまったら、その扱いをどうするかという多かれ少なかれ想像的な設定に基づく話題であり、心の哲学や法哲学と部分的に関連する。

前者の議論はともかく、後者で語られるロボットの道徳的地位（moral status/standing）とかロボットへの道徳的配慮（moral consideration）、あるいはロボットの権利（robot rights）といった言葉が引き起こす反応は、まず困惑かSF愛好者の好奇であろう[3]。このような観念を議論の的にする文献が倫理学の名のもとに相当量生産されていると知れば、何か不健全な議論がまかり通ってはいないかと疑念を抱くのも無理はない[4]。

技術の進展が倫理に実質的な問題を引き起こし得る、言い換えれば、単なる応用問題（アプリオリな倫理原則の適用）に尽きるのでない考慮を強いることがあり得る、という可能性は、とりわけ分子生物学関連技術の進展に伴って広く認められるようになっていると思われる。ひとが人工知能やロボティクスの分野の進展著しいことに瞠目しているのも事実である。そして実際、例えばわが国でも、人工知能やロボットの「社会実装」による新たな社会像の実現が、すなわち人工知能やロボティクスを利用して我々の公私に渡る生活形態そのものを大幅に改造することが、官庁の掲げる計画ないし夢想となっているくらいなのだから（内閣府の「第5期科学技術基本計画」参照）、こうした諸技術が倫理に対して持つインパクトを検討することに倫理学的・政治哲学的な意義があると考えてよいのではないか[5]。しかし、それがロボットの道徳的地位やロボットの権利という問題を構成するかといえば？

〈人間〉を超えた倫理の可能性という点で、これはもちろん動物倫理や生命倫理、あるいは世代間倫理や環境倫理の後発であり、多くの論点を引き継いでもいる。とはいえ、動物倫理がほとんど自明の理として道徳的配慮の対象（道徳的被行為者moral patient）としての動物を問題としてきたのに対して、ロボット倫理ではむしろ道徳的行為者としてのロボット、つまり人工道徳的行為者（artificial moral agent: AMA）をめぐる議論がまず盛り上がったという点が特徴的であろう

2　人工知能（というか諸々のアルゴリズムの生産と配備）に関連して現に起きている問題とは、例えば、統計的な犯罪・再犯の「リスク」を根拠として行為ではなく人間に照準し執行される司法的および警察的活動、商業的または政治的ターゲティング広告、高強度の労働管理、労働市場や保険審査その他における差別・バイアスの強固化など。具体的な事例はキャシー・オニール（2018）を、AIを専門とする計算機科学者によるこうした問題の一部に対する評価としてはスチュアート・ラッセル（2021）の4章を参照。そのような摩擦は技術の「進歩」が社会に不可避にもたらすコラテラル・ダメージに過ぎないとする技術決定論がとりわけ情報技術の分野で出現しやすい機制については、佐藤（2010: 55-9, 211-5）参照。これに関して示唆的なのは、文化人類学者デヴィッド・グレーバーが2015年の著書で“なぜいまだに空飛ぶ車が存在しないのか”と大真面目に問うた上で検討する或る「テーゼ」である。「一九七〇年代に、いまとはちがう未来の可能性とむすびついたテクノロジーへの投資から、労働規律や社会的統制を促進させるテクノロジーへの投資の根本的転換がはじまったとみなしうる」（グレーバー 2017: 171-2）。

3　「ロボットの権利」というモチーフは早くもチャペックの戯曲『R. U. R.』の序幕に――つまり「ロボット」という言葉の誕生とほぼ同時に――現れているからである。なお、AI・ロボット倫理の文献ではほぼ素通りされているが、「道徳的地位」や「道徳的地位の差異ないし程度」という観念そのものが論争の対象である。道徳的地位なる一般観念は放棄したほうが有益であるという議論としてレイチェルズ（2013）を参照。また、道徳的地位が（端的な有無ではなく）「程度」を容れるとはどのようなことか、そして容れ得るのか、という議論についてはDeGrazia（2008）を参照。

4　この話題に関する新しくかつ経験的研究もカバーしたシステマティック・レビューとしてHarris & Anthis（2021）がある。文献情報や最新の研究状況の情報をお求めの方は本稿よりもこちらを参照されたい。

5　「夢想」に必ずしも否定的意味合いはない。小泉義之が指摘するように、この「計画」が掲げている社会像は「まぎれもなくコミュニズム的な構想である」（小泉 2018: 50）。であれば、現体制が我知らず見ているその夢を嗤うよりも「夢想を盾に取って、現在の政治と経済の体制そのものに対する強烈な批判を進めるべき」（ibid.）である。

か（以下、単に「ロボット」と言う場合も人工知能（に相当するプログラムを実行するコンピュータ）に統制された機械を指す）。ロボットの道徳的地位をめぐる近年の議論の始点となった Allen et al.（2000）が初めて AMA を問いとして定式化したとき（むろん実質的に AMA を扱った考察はそれ以前から存在する[6]）、彼らは AMA の構築に向けた問題を二つのレベルに分けていた（cf. 岡本 2011）。第一に、道徳原理の選択のレベル。つまり、道徳的行為者はどのような規準に従うべきか（功利主義、義務論、徳倫理……）という問題。第二に、より概念的・存在論的なレベル。つまり、（人工）道徳的行為者であるとはどのようなことか、という問題。例えば、何らかの道徳規準に従うようプログラムされれば、それだけでロボットは道徳的行為者であるのか？　それとも規範に従うことについて思考する能力が必要なのか？　あるいは、規準を誤って適用したり意図的に規準に逆らうことすらできるのでなければ、**本当の意味での**道徳的行為者ではないのか？[7]

AMA の構築を所与の目標とする（この研究分野は機械倫理学 machine ethics と呼ばれ AI ethics や robot ethics とは一応区別される）Allen らは、主に AMA を実際に構築する方法の模索という工学的関心から二つの問題やその相関を論じたが、しかし後に続いた倫理学者たちは別段この目標を共有したわけではないので、当然というべきか、第二の概念的レベルでの問題に関する文献が多く生産された[8]。例えば、よく読まれた Himma（2009）は agent および moral agent についての哲学業界の「標準的見解」からそれと概念的に連関している主要な諸概念を取り出し（自由な選択、熟慮、善悪の知識、賞罰の帰属など）、これら諸概念がいずれも意識ないし心的状態の所有を前提していると論じ、AI が道徳的行為者であり得るか否かはそれが意識ないし心的状態をもつか否かに懸っている（概念的に依存している）と結論する。個別の理路は、“道徳的行為者であり得るためには責任が帰属可能でなければならず、責任が帰属可能であるためには賞罰が可能でなければならず、賞罰が可能であるためには意識ないし心的状態を有するのでなければならない。なぜなら賞罰とは定義上、対象が快･不快な心的状態になるように合理的に計算されたものであるから”といった具合である。

道徳的行為者の条件を規定し（典型的には、意識ないし心的状態、自由意志、熟慮、二階の志向的態度、責任帰属可能性など）、対象となる AI とかロボットの特性を評定し、それが条件を充足するか否か、あるいはどのようなロボットなら充足するか、を検討する ―― こうした言説が実際に何をしているのかといえば、従来人間や人間の行為を対象として哲学・倫理学分野で使用されてきた概念や論証パターンがロボット等の事実的あるいは想像的事例にも適用可能であるか否かを検討している、ということになろう。要するに概念適用の正当性に関する論議であるが、その際、AI・ロボット関連研究における経験的データ、AI･ロボット研究者の言説、その他の一般的臆見（ジャー

6　例えば『2001年宇宙の旅』（映画版）のHAL 9000による「殺人」の「責任」帰属に関するデネットの論考は有名である（Dennett 1997）。ちなみに、デネットがこの論考でも用いている彼の志向システム論は AMA を論じる文献で頻繁に参照される。

7　最後の立場、つまり単に規範に従う能力だけではなく道徳的信念や道徳的判断（規範的理由の適用）を反省する能力（二階の志向的態度）を真正の道徳的行為者の必要条件と見なすというのは、多くの人がまず思いつくものであろう。例えば Podschwadek（2017）がそうした立場を擁護している。

8　機械倫理学の目標設定の正当性や有用性をめぐる問題については、最近行われた応酬を参照されたい。van Wynsberghe & Robbins（2019）は機械倫理者が AMA 開発の必要を主張する際に持ち出す諸理由を批判的に検討した上で、実際には道徳的行為者性ではなく単に安全性（safety）が問題になっているのであり、AMA などという誤解を招く言葉は使わずに単に「安全なロボット」と言うべきだと提言した。これに対して機械倫理学者を含む 8 人が応答している（Poulsen et al. 2019）。

ナリズム）、といった新しく蓄積しつつある資料への準拠によって作業がなされる点が新しいわけである。

我々の概念適用（の歴史）に暗黙裡に含まれる「偏り」を指摘することにはもちろん一定の意義があろう（このような反省そのものがまた歴史的投企である）。こうして、倫理に関わる諸概念を新たな視点から捉え返し、それによって、例えば〈人間〉を前提にしているかぎりは盲点にあった想定を明るみに出すといった弁明がひとまず立つわけだが、しかしそれだけに従来の哲学・倫理学的議論の未決着性をそのまま持ち越すことにもなりがちである。Himma 自身、或る AI が意識を有するか否かを知るに際しては人間の場合の「他我問題」と同じく認識論的問題があるとして不可知論の立場をとる。あくまで“a conceptual matter”として、道徳的行為者の存在には意識ないし心的状態が必要条件であると示しただけ、というわけだ。むろん、多くの論者がそれぞれに工夫を凝らして鮮明な結論を出してもいるが、まさにこの一見して突飛な主題に合わせた高度に技巧的な「解決」である点が、倫理学的議論としての有効性に疑問を感じさせてしまわないでもないのである[9]。

ロボットの道徳的被行為者性（patiency）―― 道徳的行為の受け手、あるいは道徳的配慮の対象であること ―― がやや遅れて注目されるようになったとき、こうした点が予習済みであったことが、議論がある種の「実践的」な色彩を帯びるようになった原因でもあろう。先輩格である動物倫理への参照により、痛み・苦しみを感じる有感存在（sentient being）であることが道徳的被行為者性の条件として真っ先に挙げられるが（ピーター・シンガーあるいは元をたどればベンサム）、こうなるとロボットは大半の動物よりも不利である[10]。現在のみならず近い将来にもロボットが sentience を備える見込みは少ないし、有感存在になるか否かは基本的には製造する人間次第である（やろうと思えば有感性を実装できるが敢えてしない、という状況も想像できる）[11]。そもそもロボットが製作される理由の多くは、人間がやりたくない仕事を押し付けるため、あるいは「非人間的」な活動から人間を解放して人間の福祉を向上するためであって、未だ不十分な人間の権利状態の改善に向い得たはずのリソースがロボットの権利保護のために用いられかねないような社会体制は本末転倒ではないか？（ロボットは道徳的配慮の対象とならない「奴隷（slave）」であるべ

9　そうした創意工夫の中から対照的な二つの例を挙げておこう。非常に多く参照されている Floridi & Sanders（2004）は、抽象レベル（Level of Abstraction）を中心とした独自の情報哲学理論によって行為者性を規定し、また責任（responsibility）と答責性（accountability）を厳密に区別する。そして道徳的行為者であるためには答責性を有すること（善・悪を結果する行為の源泉として同定され得ること）で十分であり、責任帰属（道徳的評価の対象になり得ること）は必要ではないとする。そして抽象レベルの理論によれば、ある存在者が（道徳的）行為者であるか否かはその時の抽象レベルによって変わるのである。

フロリディが独自の哲学体系という大仕掛けを背景とするのに対し、佐々木（2013）はメタ倫理学の責任論を応用して外科手術的な慎重さで障害を切除している。Fishcher と Ravizza の責任論はメタ倫理学史上に大きな影響力を持ったが、その特徴は、行為の帰属条件を行為の構成要素（熟慮や自由意志など）ではなく、行為者の能力・メカニズム（適度な理由応答性 reason-responsibility）に求めたことにあった。これに従えば、「道徳的に責任ある行為は、必ずしも熟慮というプロセスを必要としないし、意識的に生み出される必要もない。すなわち、精神的行為プロセスというものを必要としない」（74 強調原文）。そしてメカニズムの評価という点では人間よりもロボットのほうが容易でさえある。

10　Darling（2016）はむしろ経験的事例における人間の側の感情的反応に注目して、AI・ロボットの道徳的・法的地位を考えるために動物との類比が有用であると論じる（これは本文以下で紹介する「関係論的アプローチ」に通ずる）。また、多くの成功した動物保護運動や保護法の制定は生物学的な規準よりも一般大衆の感情に基づいていたことを指摘している（226）。逆に Johnson & Verdicchio（2018）は動物との類比を用いるのはミスリーディングであると主張する。

11　Schwitzgebel & Garza（2015）は、AI は人間が意図的に創造し選択的に性質を与えるものであるという事実から、人間は AI に対して、人間に対するよりも大きな道徳的配慮の義務を負うかもしれないと論じている（108-10）。

きであるという主張がなされる所以でもある（Bryson 2010）[12]。

しかし話はこれで終わらず、むしろ、或るロボットが**本当に**苦しみを感じているか否かはそれの道徳的地位にレレヴァントでは**ない**、という仕方でロボットの道徳的被行為者性を擁護する議論が論争のシーンをリードすることになった。

そのような議論を主導する Coeckelbergh は、AI とかロボットに内在する属性（property）を特定してそれを根拠に倫理学的帰結を引き出すような議論を「プロパティ・アプローチ」と呼んで批判し、代わりにロボットと人間との具体的な関係ないし interaction に焦点を移すことを提案しており、「関係論的 relational」なアプローチの旗手と見なされている[13]。彼がこれを意識的に転機として示した「関係論的転回 relational turn」（Coeckelbergh 2010b: 219）という言葉はこの話題の論者たちに定着している。彼はまた、inner/real から outer/appearance への転回、とも言う。つまり、プロパティ・アプローチは先にも触れたような認識論的な問題を抱えている、といった批判を起点に、むしろ人間 - 人間のインタラクションにおいても与えられているのは現われ（appearance）だけであり、人間 - ロボットより状況がマシなわけではない（他我問題）、と歩を進める。そしてロボット - 人間のインタラクションの道徳性についても、人間 - 人間のインタラクションと**同様に**、「概念を正しく把握すること」からではなく（哲学者などの専門家が指導するのではなく）、普通の人々の経験と実践から出発して考察するべきである、と（Coeckelbergh 2009a: 184, 188）[14]。「概して我々は、心的状態や意識をもつことの証明を他人に対して要求したりしない。むしろ我々は他者の現われと振る舞いを情動として解釈する。さらに我々は他者もまた我々と同じように解釈しているかのように（as if）他者と相互作用する」（Coeckelbergh 2010a: 238 強調原文）。問題は、ロボットが**実際に**心的状態ないし意識をもつか否かではなく、心的状態をもつ**かのように**現れるか否か、人間に対するロボットの現われなのである[15]。これは個別の倫理的イシューの論議というよりはまさにアプローチの提案、研究指針の特徴づけと正当化と言うべきものであるが、例えば、Danaher（2020）は関係論的アプローチに共感する立場からより踏み込んで、「パフォーマンス上おおよそ同等 roughly performatively equivalent」な ―― 振る舞いのレベルで重要な差異がない ―― 存在者の道徳的地位は同じであるべき、という「倫理的行動主義」を提起し、そのような性能をもつ（潜

12　Bryson のこの論文は挑発的な表題ゆえか頻繁に参照されるが、論旨はそれなりに複雑である（cf. 岡本・稲葉 2020: 210-3）。もちろん人間を奴隷化した歴史的制度としての奴隷制を正当化しているのではなく、むしろロボットを人間扱いすれば本当に奴隷制になってしまうので非人間の「奴隷」ないし「サーヴァント」に留めるべきというのが眼目だろう。しかし Bryson が意図する用語法でも危うさを孕んでいることは変わらない（Gunkel 2018b: 196n7）。

13　Coeckelbergh と Gunkel をこの潮流の代表とするのが一般的なようである。そのような見取り図のほぼ最新版として Schröder（2021）を参照。より簡潔に学説の流れを把握できる邦語論文としては水上拓哉（2020）がある。なお、Coeckelbergh が「関係」（あるいは関係と関係項との関係）に込める意味に関する重要な参照先として、Callicott（1989）の土地倫理における「関係は関係する事物に先行する」という考えがある（Coeckelbergh 2012: 45, 50; Gunkel 2018a: 96）。

14　Coeckelbergh は AI 倫理学が政策やビジネスなどに寄与する（それを通して社会の善導教化に寄与する）と強力に主張するという意味でも「実践」指向である。とりわけクーケルバーグ（2020）はそのようなメッセージに満ちている。

15　ロボットの「心」のリアルなど**問題にしない**、つまりはロボットの「心」のいかなる理論に対しても不可知論の立場を取るということである（Coeckelbergh 2009b: 220）。要するに「プロパティ・アプローチ」の各論者と個別の主張内容の水準で正面から対立するわけではなく、方法論的な水準の批判が主だということ。道徳的地位一般に関しても、「道徳的地位とは何か？」ではなく「或る存在者が特定の道徳的地位を持つものとして現れるための条件は何か？」という形で探究するのである（Coeckelbergh 2012: 7）。

在的）ロボットの道徳的地位を正当化している[16]。同様に考えてゆけば、「人格ロボット」の可能性に関しても、問題はロボットの「内面」の真相ではなく「「そこに内面があると想定せざるをえない、自己の同胞たる「人間」扱いせずにはいられない他者」としてふるまうようなロボットができてしまうかどうか」（稲葉 2016: 109）ということになろう。

このような見解自体は多くの人がポピュラー・カルチャーのなかで一度は見たことのある類のものであろうし、近年の哲学論議としても珍しいものではない。大森荘蔵はまさにロボットの痛みや心の問題と他人のそれとを比較して言っていた。「木石であろうと人間であろうとロボットであろうとそれら自体としては心あるものでも心なきものでもない。私がそれらといかに交わりいかに暮らすかによってそれらは心あるものにも心なきものにもなるのである。それに応じて私もまた「人間」になるのであり、また文明人にも未開人にもなるのである」（大森 1981: 71-2 強調原文）。「未開人」への言及があるのは、大森がこれを「アニミズム」と言い、東京人も「いくらか偏狭で差別的」とはいえこの点では変わりないとするからである。

2.

大森の言語感覚は彼の議論の特徴をよく表示するものであったと言える。“primitive people”に「根源に近い」という響きを聞き取れるように（「未開人」という訳語はこれを全く捨象しているが）[17]、この種の思考の筋道に現れているのは心や道徳の始原への指向性、つまり人間のファンダメンタルな事実を露呈することによって心や道徳あるいは価値の存在性の「問題」を解消するという性格である。現在直面している裂け目（例えば人間とロボットの「本性」に基づく道徳的地位の差異）は始原への遡行によって窮極的な裂け目ではないことが確認され、解消される。あるいは根拠のない窮極的な裂け目（原事実的な差異）自体によって問いそのものが塞がれる。かくして道徳性は個体ないし種についての（哲学者の説く）真理に基礎づけられるのではなく、人間の自然的社会性という基底の上に、つまり他者と交わって自らを形成していく生の形態に宿る、と承認される。実際、文化人類学的な事実としては、動物や精霊、先祖、神とのあいだで倫理が可能であることは明らかではないか？

とはいえ道徳性の始原への指向はロボットにとって必ずしも福音ではなく、アンビバレントな帰結をもつかもしれない。というのも、（1）一方でロボットの道徳的地位の可能性は**原理上**（反実仮想的可能性として）認められるが（様々なSF的シナリオ[18]）、（2）他方で現在の例えばわが国の社会の**連続上**にそれを現実的に構想することは難しく、かつ、始原遡行においてまさにこの連続

16　すでにSchwitzgebel & Garza（2015）が「もし存在者Aがある一定程度の道徳的配慮に値し、かつ存在者Bが同じ程度の道徳的配慮に値しないのであれば、二つの存在者にはこの道徳的地位の差異を基礎づける何らかのレレヴァントな差異が存在しなければならぬ」という前提からロボットの道徳的地位を擁護するほぼ同様の議論をNo-Relevant-Difference Argumentの名で提出しており、Danaherの議論は同様のアイデアを扱うに際して方法論的に行動主義を採用して種別化したものと見なすこともできる。実際Danaherは、倫理的行動主義は倫理的ドメインに方法論的行動主義を適用するものだと言っている。例えば、内的な心的状態が倫理原則の究極的な形而上学的根拠になることを直接否定するのではなく、単にそのような根拠の存在の認識的保証は観察可能な振る舞いからのみ導出され得ると主張するのである。

17　中井久夫の指摘による（中井 2001: 160）。

18　稲葉（2016）は高度に自律的な「人格的」ロボットの存在と権利について、人格ロボットのニーズをもつ社会の構想と結びつけて考察している。特に第5章と「おわりに」を参照。

性（「私がそれらといかに交わりいかに暮ら」してきたか）が真実性を与えるからには、（1）は自動的に空疎にならざるを得ないからである。現に多くの人は多くの場合に木石に対して人間に対してするようには振る舞わないのである。そして多くの場合にそう振る舞うようになったときにはもはや殊更言う（assert）べきことはないだろうし、またそう振る舞う者とそう振る舞わない者の対立が顕在化するとき ―― インテンシヴな工場畜産への賛否に比せられよう ―― についても、どちらの経験ないしインタラクションが規範的命題の真正の根拠となるかを決定する規準がない以上（あるいは規準の存在や生成に関する社会学的記述以上の説明が与えられない限り）、やはり**倫理学的には**言うべきことはとくに無いのではないか。むろん Coeckelbergh はそのような「常識」的批判には先回りして、比較的初期から「人間 - 人工エージェントの観察上ないし想像上のインタラクション」に基づいて、「人間のような human-like」行為者性や責任帰属を含むインタラクションが正当化されるのはどのような場合でありいかにしてかを「規範的問題」として問うと宣言している。（Coeckelbergh 2009a: 188）。しかし実際いかにして彼のアプローチの内部でこの問いに**規範的**次元を与え得るのか私には判然としない。これに対し、関係論的アプローチを徹底して「人間中心主義」を脱する仕方でロボットの権利を擁護したい Gunkel などは、むしろ既存の人間的相互関係に基づく道徳的思考自体の切断と方向転換の可能性を、レヴィナス的な、主体性の手前／彼方のトラウマ的経験に認めようとする ―― ロボットの「顔」（face-plate）との遭遇の出来事である（Gunkel 2018a: 95-6）。しかしレヴィナスの**結論**に便乗するかぎりでは〈他者〉への無限の開かれという原則を堅持する以上の主張をもたないだろうし、実際彼の仕事は今のところこの「別の仕方で考えること」のアウトラインを示すに留まっているように見える[19]。

さて、上では（2）ロボットの道徳的地位を現在社会の連続上に考えること、は難しいと断定したが、最近の文献では（2）の方向から、徳倫理学を応用して間接的な仕方でロボットの道徳的地位を規定する議論が提出されている。ロボットに徳を組み込もうというのではない（そうした構想もあるが[20]）。人間の徳に対する配慮から派生するものとしてロボットの道徳的被行為者性あるいは道徳的配慮を基礎づけようというのである。ここで想定されているのは特にソーシャルロボット（人間と何らか情動的なコミュニケーションをするように設計されたロボット）であり、現行の製造物（ペットロボットなど）も視野に入っている。

例えば、ロボットの「虐待」を禁ずる道徳的配慮を徳倫理学的に考えるとは？　鍵となる一手は、道徳的問題の位置を、ソーシャルロボットに対する個々の不当な取り扱いとか「虐待」ではなく、そのような行動の積み重ねによる悪しき傾向性の定着によってユーザーの中に邪悪な性格が形成される点に定めることである。「徳倫理学はロボットの不当な扱いがロボットにとって悪いと主張する必要はない。むしろ、ソーシャルロボットの不当な扱いはユーザー自身の道徳的本性の充足や自

19　というより Gunkel の立場からすれば、倫理の具体的内容をロボットの「顔」との遭遇に先立って規定することはできない。そして存在論ではなく倫理が第一哲学であるというレヴィナスのテーゼに従う限り、ロボットが何であるかはそれにも増して規定することができない ―― 我々がロボットの「顔」にいかに応答するかということが、ロボットが何者であるかを定めるのだ。彼の議論の内側から見るかぎり具体性の欠如は瑕疵ではないのである。Gunkel は前掲論文や *Robot Rights* において、むしろロボットの権利に関する既存の言説の歴史的・分類的整理と批判的検討のほうに多大な労力を注いでおり、その面で参考になる。岡本（2019）は Gunkel のそうした議論の簡便な紹介である。

20　Allen et al. 前掲論文参照。より詳しくは、W・ウォラック／C・アレン（2019）の特に第 7、8 章を参照。より簡単な見取り図としては、小山（2018）が規範倫理学の主要な三理論（義務論・帰結主義・徳倫理）と AMA との相性を整理している。

己改善の達成を妨げるかぎりで、ユーザーにとって悪いのだと主張する」(Cappuccio et al. 2020: 15)。ユーザーの善き性格形成、そして悪しき性格形成の予防、ひいては人間共同体の徳の涵養と悪徳の抑圧のために、ソーシャルロボットへの道徳的配慮が正当化されるのである。このときロボットの「真の」属性については Coeckelbergh らと同様カッコに括り、ロボットの道徳的行為者性や権利主体性といった怪しい問題とは独立に、あくまで我ら人間自身のために、我ら人間自身についての知識に基づいて、ロボットの道徳的被行為者性(道徳的配慮の対象としての地位)を基礎づける。その利点は、今ある事例から空想事例まで幅広く評定可能なレンジの広さと、既存の道徳理論の枠組みをよく保存する穏健性であろう。要するに、倫理学にとってロボットは多少珍品とはいえ根本的な新規性はなく、その道徳的地位は人間の性格形成に影響する因子という面から**のみ**考察すればよく、詳しい内実が未定の将来的事態についても、それに対する人間の振る舞いが人間の性格形成にどのような影響をもたらすか、という一点から一定の規準をもった想像を巡らせることができるわけである。ここで道徳性の始原はあくまで、互いの(ひいては共同体の)徳を気遣う〈人間〉たちの共同体である。

この潮流は Gerdes(2016)がカントの徳論に基づいて青写真を示したのち、Cappuccio らが現代徳倫理(とホネット流の承認論)を用いてより詳細に展開した。2010 年の論文では義務論・功利主義・徳倫理学のいずれとも異なるオルタナティヴとして関係論的アプローチを提示し、それが徳倫理学および共同体主義的枠組みと両立可能か否かは未決の問いとしていた Coeckelbergh も(Coeckelbergh 2010b: 220)、最近は徳倫理の応用に合流してアイデアを追加するような論文を発表しており、少なくとも関係論的アプローチの一つの正当な種別化と認めたようである(Coeckelbergh 2021)。一見すると、人間とロボットの具体的相互作用に道徳性の根拠を置く関係論的アプローチと、人格および人格共同体の善さを本源的価値(道徳的評価の第一義的対象)として道徳理論を展開する徳倫理との間には容易に埋めがたい距離があるように見える。が、Coeckelbergh は「テクノロジーそのものが深く人間中心的」(Coeckelbergh 2009a: 188)であると認めているし、関係論的アプローチはまさに人間 - ロボット関係の**人間による**経験に陣取るのであるから、実は人間および人間共同体への影響という一点からロボットの道徳性を評定する徳倫理の応用と発想の根を同じくしているとも言える。つまりロボットの側に歩み寄るというような含みは元々なく、合流はさほど意外ではない。かくして、議論の焦点、言うなれば道徳性のリアルの位置は、対象(プロパティ・アプローチ)から対象との関係性(関係論的アプローチ)へと、そしてさらに、関係を経験しそれにより変容する人間(共同体)自身へと、徐々に手元にたぐり寄せられたと言えよう。

ここまで通時的なナラティヴを擬して三つのアプローチ――(1)プロパティ・アプローチ、(2)関係論的アプローチ、(3)徳倫理的アプローチ、を紹介したが、これに(4)技術論的なアプローチ、を加えれば、AI・ロボットの道徳的地位に関する議論を大雑把に分類できると思われる。技術論的とは、AI・ロボットを生み出す技術あるいはそうした技術によって生活形態や自然生態系が広範に媒介された世界、を指向した考察であって――しばしばそうした世界ないし環境は“ネットワーク”に類する語彙で捉えられる――その世界における人間や AI・ロボットの位置価に依拠して倫理的帰結を引き出すのである。例えば Johnson(2006)は「技術」とは人工物・社会的実践・知識体系から成る「アンサンブル」であるとした上で、この結合体のなかで人工物の志向性(機能性として

定義される）は設計者と使用者の志向性（行為する自由な意図）に依存するゆえに、コンピュータ・システムはその機能において道徳的に重要な差異を結果する道徳的存在者（moral entity）ではあり得ても、（単独では）道徳的行為者であり得ない、と主張する[21]。

これは網羅性を企図した分類ではない。分類の原理は、倫理学的主張の根拠の特定方法、換言すれば道徳性のリアルを世界のどこに・どのように見るかということである。要するに、言説の拠点確保の仕方である。一見して「道徳的世界」から浮いているように見えるロボットをその内に置き入れるためには、この点で明示的あるいは陰伏的な決定を強いられる。その陣取りにロボットの道徳的地位という話題の、倫理学としてのエッジがあり得るとひとまずは言えよう。この分類の場合は、（1）対象の属性、（2）人間と対象との関係、（3）人間（の共同体）、（4）人間を含む事物が埋め込まれた、技術化された環境世界、がそれである。

かくして、ロボット・AIの道徳的地位ひいては権利を論じるといういささか体面の悪い議論は、「実践的」でありかつ既存の価値への根本的挑戦を含まない穏健な形態を見出したということになろうか。

振り返れば90年代にAIの法人格という主題について法学者の視点から先駆的に論じたローレンス・ソラムは論文末尾で述べていた。「我々の法人格の理論は、地位の境界線に存する諸々の深淵にアプリオリな導きを与えてはくれない。〔…〕政治的・法的領域におけるハードケースの解決は公共的理性の使用を要求する。寛容と尊重という我々の共有遺産の上で原則に反さぬ妥協を模索するという以外の現実的選択肢はないのだ。ハードケースを解決する人格性の理論を構築するための共通基盤がない場合、裁判官は、相互に市民仲間（fellow citizens）として承認し合う人たちの権利を尊重するという原則に戻らなければならない」（Solum 1992: 1287; cf. 青木 2017: 59）。必要な変更を加えれば、倫理学においても重要な指摘はここでほぼ尽くされているとも言えよう。しかし誰が誰を市民仲間（道徳共同体のメンバー）として承認するのかが問題だということを我々は嫌というほど知っている。そう考えると、徳倫理学の応用という手法は始原への遡行が指し示すメンバーシップの問題（裂け目）を予めスキップする身振りであると見なすこともできる。そこでは境界線は所与として前提され、倫理学者の仕事はこの境界線によって確定される共同体なる集団単位で人間たちを善導教化する（制度的そして物質的な）デヴァイスの設計と運用に際してアドヴァイスを与えることであろう。しかしそうであるかぎり、この市民的相互主観的実践理性への信に、

21　Johnsonにとって、「人工物」（artifact）は社会的活動という「現実」から心的作用によって概念的に抽象・分離されたものに過ぎない（Johnson 2006: 197）。技術論的アプローチの他の例としては、ハンス・ヨナスの技術哲学を応用して、人間とソーシャルロボットはいまや共に「技術世界内存在（being-in-the-technological-world）」であるとするTavani（2018）――結論はJohnsonとは逆――を参照。また、電子エージェントの法的権利（法人格）に関してB・ラトゥールのアクターネットワーク論を応用したTeubner（2006）もこれに類するアプローチと言える。この類のアプローチは、動物や他の自然物（植物や川など）の権利を擁護する際の根拠としてある種のエコロジカルな全体論的環境観に訴える議論にその原型を見ることができよう。そもそもmoral standing/statusという用語はlegal standingという法律用語からの借用であるが、この借用を促したとみられるChristopher Stoneの論文「樹木のlegal standing」は、自然物の権利という問題の「法実務的側面」とは別の「精神的および社会 - 精神的側面」として、地球を一つの有機体と見立てるような――人類はこの有機体の「マインド」たる機能的部分である――全体論的自然観へと我々の「意識」を高めていく指向を表明していた（Stone 1972: 498-500）。なお、より本格的な技術哲学からの言及としては、技術の道徳性を論じるフェルベークの技術論・主体論も人工知能を射程に収めている（フェルベーク 2015: 185-6）。

祓ったはずの悪霊がやがて回帰するのも当然ではなかろうか[22]。Cappuccio らもそれは感知している。「ソーシャルロボットが次第にますます適応的で自律的になるにつれて、それらはますます奴隷に似たものに見えてくるであろう〔…〕。実際、知性的かつ自律的なエージェントを所有し使用するというのは、まさしく奴隷制の定義そのものである」、すなわち「ロボットを擬人化しながら非人間化するパラドクス anthropomophizing while dehumanizing robots paradox」（Cappuccio et al. 2020: 25）。徳倫理が奴隷制を引き連れて回帰するのではあまりに皮肉というものだろう[23]。Cappuccio らは、パラドクスを生み出さないためには「ロボットを人工物として道具的に使用することと、社会的同胞としての尊厳ある地位に照らしてふさわしいケアと尊重をロボットに与える可能性の両方を説明できる」ような「新たな社会的承認の概念」に基づく人間 - ロボット関係の再構築が必要であるとするが（ibid. 26）、これは印象だけを肯定的に変えてパラドクスをそのまま言い直したにすぎないものと私には見える。

それでは、このような穏健な方向に飽き足らず、始原への遡行を徹底して敢行するとすれば？柴田正良はロボットとの倫理を考えるにあたり、無人の宇宙に住む行為者、すなわち「自分を生み出した社会や共同体の記憶は保持しているものの」「現在も未来も、それどころか現実的であれ可能的であれ、いかなる共同体のメンバーでもない」行為者を想定する思考実験を行っている（柴田 2010: 17）。そしてこの「新クルーソー」には義務論・功利主義いずれの枠組みにおいても、他者の非存在ゆえに行為を倫理的に評価する観点自体が成立し得ず、「「なしたいこと」とは別の「なすべきこと」は存在しない」、したがって「倫理なるものは存在しない」と論じる（17-22）。そう為す**べき**だったのか、それともそう為す**べきと思っただけ（そうしたかっただけ）**だったのか――その差異の可能性に規範のリアリティが存するとすれば、彼の世界には規範が存立するいわば論理的余地がないわけである。この実験から、「行為者に倫理が発生するとすれば、その行為者と同一の権利･義務を互いに所有しうる他の行為者が同時に存在し、彼らが同じ共同体に属する場合に限る」という「共同体テーゼ」（倫理の必要条件）が引き出される（22-3）。さて、ロボットは我々の現実世界にあって新クルーソーなのであろうか？

大森とは別の筋道を通ってではあるが、ここにも道徳の始原を人間の自然的社会性へと遡る指向性がある。柴田はまず、「われわれがなぜ現にあるような倫理を必要とするのか」の示唆を与えるものとして、ハートが挙げる、自然法の最小限内容の理由を示す自然的事実を参照する――「人間の傷つきやすさ vulnerability」、「大まかな平等性」、「限られた利他性」、「限られた資源」、「限られた理解と意志の強さ」（ハート 2014: 302-11）。ロボットがこれら人間の条件＝制約を共有しないとすれば、彼らとの間でいかなる倫理、いかなる道徳共同体が可能でありまた必要であるのか？[24] そこで柴田はさらに「われわれにおいて倫理を現にあるような仕方で可能としている条件」まで遡り、最低限、反応的態度（ストローソン）を取り心の理論（≒フォーク・サイコロジー）をもつこ

22　それゆえここで問い返されるべきことの一つは、〈市民〉なる概念の内包が無媒介に外延集合を確定するという信憑がどのように生じるのか、そして実際にはいかなる物質的装置がそれを媒介しているのか、であろう。小泉（2006: 80-1）を参照。

23　法学分野では現在、無権利の AI・ロボットによる契約行為について、古代ローマ法における奴隷の扱いを参考に対処しようという提案がある（Pagallo 2013: 103）。出雲（2019）がパガロのこのアイデアを解説している。

24　法学者の大屋雄裕氏は、AI を法的な権利主体として認める最大の困難は「可傷性 vulnerability」と「個体性 individuality」という二条件を人間と共有しないことがあり得ることだとして、ロボットを包摂する（あるいは敢えて「奴隷」にとどめる）社会構想の選択肢を提示している（大屋 2017; 大屋 2020）。

と、要するに「他者の心の理解」がその可能性条件であると言う（柴田 2010: 28）。道徳共同体の可能性条件は心の共同体であるわけだ。対象に対して心あるものに対するように振る舞う、対象を心あるものとして解釈する生き方をするということ ―― かくして柴田の考察も大森の指摘に合流する。ところで、もし心あるものに対するように振る舞うということの内実が、結局のところ我々が現に相互にしているように振る舞うということであるとすれば（単純化して言えば、“心ある者の実践”は当の実践の内側からしか理解され得ないとすれば）、かつそこでの「我々」とは現に相互にそう振る舞っている存在を指すのだとすれば、探究の鋤はここで岩盤に達してはいまいか。大森はそれをよく分かっていたので、「木石であろうと人間であろうとロボットであろうとそれら自体としては心あるものでも心なきものでもない」と傍点を振ったのであるが、では心と道徳の始原が「私がそれらといかに交わりいかに暮らすか」であるとして、世界に魂の気息を「吹き込む」（大森 1981: 140）その「私」あるいは「私」の**その**生き方が**現実に**どのように形作られ得るのかといえば、それがそんな「私」**たち**によって伝承された生き方として習得されるとしか言えないのだから、ひと回りして柴田の「共同体テーゼ」も強化され、以下、個と全体（共同体）の関係は解明的とは言えない循環になりえよう[25]。とはいえ、どのように循環を止めるかによって、引き出される道徳性のイメージは異なる。柴田はありうべき倫理の方向として、ロボットを「最低限の素朴心理学的な他者理解を示す限りにおいて、同一の道徳共同体に含める倫理」を、「どちらかと言えば人為的なルールや法に似た倫理的システム」という形で「制定すること」を提起し、つまりはロボットが道徳共同体の一員となるか否かは我々が生き方を選ぶ意志的決定の問題であるとして、人為性を強調する[26]。逆に、大森タイプであれば、すでに蓄積した振る舞いの歴史あるいは安定的習慣（による自明性）こそがそのような意志的決断（仮にするとして）を支える、あるいはそもそも十分に規定された意志内容を与える基盤である点を強調するだろう。倫理は意志か習慣か、と乱暴に切り詰めれば、容易に解きがたい旧い問題に行き着いたことが分かる。

3.

柴田氏の思考実験が示すように、〈道徳的なもの〉は存在と当為の間の緊張に住まうと言ってよければ、この緊張、間隔、ギャップをどのようにロゴスの中で（再）構成するか ―― 緊張を存在せしめる「実践」の特定を介して間接的にであれ ――、そこに〈道徳的なもの〉を扱う学の固有性の証明がある[27]。Coeckelbergh の徳倫理学への接近は、関係論的アプローチの当初の形ではこの間隔をうまく開くことができなかった徴候とも読めるかもしれない。

25　こうした議論の中で「我々」（一人称複数代名詞）が異なる複数の論理形式で用いられている可能性はある。Haase（2012）参照。

26　柴田（2021）ではこの強調はより鮮明になっている。なお、これはパトナムが 1964 年に出した結論にほぼ等しい（Putnam 1964: 668-91）。パトナムは意識や心理に関わる言語の特性の言語哲学的考察から、人間と「心理学的に同型」（677）なロボットが意識や生をもつか否かは「発見」に左右される問題ではなく、ロボットたちを我々の言語共同体の仲間（fellow member）として扱うか否かの決断の問題であると論じている。その決断以前には、ロボットは意識をもたないのではなく、或るロボットが意識をもつ／もたないという言明がそもそも真理値をもたないのである（690-1）。

27　権利＝正という文脈内に限っても、この「間」をどう観念するかに関して近代が古代や中世と大きく異なることは明らかである。シュトラウス（2013）の特に第Ⅲ章とⅣ章を参照。

*

以上の不備の多い覚書につき、二点コメントを付して本ノートを閉じる。

①以上の私の考察は、少なくともこのままでは、AI・ロボットに関わる倫理学にいかなる示唆を与えるものでもないだろう。「応用」倫理学の一つの任務が、現実の問題の実践的解決を志向しながら根本的な反省を行うことにあるのだとすれば、我々は明らかに、例えばAI（というより諸々のアルゴリズムの生産と配備）やロボットを用いて経済的・精神的搾取を強化する資本のメカニズムをより正確に洞察することや、逆にオートメーションの高度化・拡大や生老病死の諸状況へのロボット配備をどうすれば我々の集団的自律性や共通善や自由に結びつけられるのかを調査・思考すること、また核兵器時代で終わっているC・シュミット『パルチザンの理論』を自律兵器の段階に更新すること等々に力を注ぐべきであり、こうした戦略的問題設定を通してこそ、ロボットの道徳的地位なり権利なりの考察もトータルな倫理学的＝政治哲学的パースペクティヴの中に位置づけられ得るのではないか。この意味で、権利については思弁的探究の前にまずはより具体的・政治的な形で問題を立てることが重要である、という趣旨に限定すれば関係論的アプローチにも理はあると思う。

②人間の自然的社会性（の評価）を元手に権利＝正を導出する議論型の系譜を、我々は近代的自然権論として知っている。ロボットの道徳的地位ないし権利という問題を権利一般の問題へと有意義に結合（あるいは短絡）させ得るとすれば、こうした伝統の学び直しを通してではないか。というのも、第一に、そもそも人間の自然的社会性をどう見積もるかという点（自然状態論）の多様性を考慮に入れることができるし[28]、第二に、かかる評定の背景に異質・「異種」的な存在との邂逅――そして暴力や奴隷化、それと並行する経済構造の大規模変動――という歴史的条件があった点にパラレルを見ることができそうだからである。[29]

※ 本稿は「JST-RISTEX「科学技術の倫理的・法制度的・社会的課題（ELSI）への包括的実践研究開発プログラム「人工主体の創出に伴う倫理的諸問題を分析・討議するプラットフォームの構築に向けた企画調査」」による研究成果の一部である。

参考文献

Allen, C., Varner, G., & Zinser, J.（2000）“Prolegomena to Any Future Artificial Moral Agent,” *Journal of Experimental & Theoretical Artificial Intelligence*, 12, pp. 251-261.

Bryson, J. J.（2010）“Robots should be Slaves,” In Yorick Wilks（ed.）, *Close Engagements with Artificial Companions: Key Social, Psychological, Ethical and Design Issues*, pp. 63-74.

Callicott, J. B.（1989）*In Defense of the Land Ethic: Essays in Environmental Philosophy*, Suny Press.

Cappuccio, M. L., Peeters, A. & McDonald, W.（2020）“Sympathy for Dolores: Moral Consideration for Robots Based on Virtue and Recognition,” *Philosophy & Technology,* 33, pp. 9-31.

Coeckelbergh, M.（2009a）“Virtual Moral Agency, Virtual Moral Responsibility: on the Moral

28　タック（2015）参照。なお、自然権論の参照が必ずしも突飛なものではない傍証として、Pagallo がロボット法を自然法論の伝統と結びつけて考える意義を認めていることが挙げられる（Pagallo 2011: 348-349）。

29　文献を教示して下さった匿名の査読者に感謝申し上げる。文献調達の都合上、十全に取り入れることはできなかったが、同著者の同趣旨の論述を参照する等、できるだけ修正稿に反映させるよう努めた。

Significance of the Appearance, Perception, and Performance of Artificial Agents," *AI & SOCIETY,* 24, pp. 181-189.

――― (2009b) "Personal Robots, Appearance, and Human Good: A Methodological Reflection on Roboethics," *International Journal of Social Robotics,* 1, pp. 217-221.

――― (2010a) "Moral Appearances: Emotions, Robots, and Human Morality," *Ethics and Information Technology,* 12, pp. 235-241.

――― (2010b) "Robot rights? Towards a Social-Relational Justification of Moral Consideration," *Ethics and Information Technology,* 12, pp. 209-221.

――― (2012) *Growing Moral Relations: Critique of Moral Status Ascription*, New York: Palgrave Macmillan.

――― (2021) "How to Use Virtue Ethics for Thinking About the Moral Standing of Social Robots: A Relational Interpretation in Terms of Practices, Habits, and Performance," *International Journal of Social Robotics,* 13, pp. 31-40.

Danaher, J. (2020) "Welcoming Robots into the Moral Circle: A Defence of Ethical Behaviourism," *Science and Engineering Ethics,* 26, pp. 2023-2049.

Darling, K. (2016) "Extending Legal Protection to Social Robots: The Effects of Anthropomorphism, Empathy, and Violent Behavior Towards Robotic Objects," In R. Calo, A. M. Froomkin, & I. Kerr (Eds.), *Robot Law*. Edward Elgar, pp. 213-232.

DeGrazia, D. (2008) "Moral Status As a Matter of Degree? " *Southern Journal of Philosophy,* 46 (2), pp. 181-198.

Dennett, D.C. (1997) "When Hal Kills, Who's to Blame? : Computer Ethics," In Stork, D (ed.) *Hal's Legacy: 2001's Computer as Dream and Reality*, Cambridge, MA: MIT Press, pp. 351-365.

Floridi, Luciano & Sanders, J. W. (2004) "On the Morality of Artificial Agents," *Minds and Machines,* 14 (3), pp. 349-379.

Gerdes, A. (2016) "The Issue of Moral Consideration in Robot Ethics," *Acm Sigcas Computers and Society,* 45 (3), pp. 274-279.

Gunkel, D.J. (2018a) "The Other Question: can and should Robots have Rights?" *Ethics and Information Technology,* 20, pp. 87-99.

――― (2018b) *Robot Rights*, Cambridge, MA: MIT Press.

Haase, M. (2012) "Three Forms of the First Person Plural," In Abel, G. and Conant, J. (ed.) *Rethinking Epistemology: Volume 2*, Berlin, Boston: De Gruyter, pp. 229-256.

Harris, J. & Anthis, J. R. (2021) "The Moral Consideration of Artificial Entities: A Literature Review," *Science and Engineering Ethics,* 27 (4), pp. 1-95.

Himma, K.E. (2009) "Artificial Agency, Consciousness, and the Criteria for Moral Agency: What Properties must an Artificial Agent have to be a Moral Agent?" *Ethics and Information Technology,* 11, pp. 19-29.

Johnson, D.G. (2006) "Computer Systems: Moral Entities but not Moral Agents," *Ethics and Information Technology,* 8, pp. 195-204.

Johnson, D.G. & Verdicchio, M. (2018) "Why Robots should not be Treated like Animals," *Ethics in Information and Technology,* 20, pp. 291-301.

Pagallo, U. (2011) "Killers, Fridges, and Slaves: a Legal Journey in Robotics," *AI and Society,* 26 (4), pp.

347-354.

――― (2013) *The Laws of Robots: Crimes, Contracts, and Torts*, Dordrecht, The Netherlands, Springer.

Podschwadek, F. (2017) "Do Androids Dream of Normative Endorsement? On the Fallibility of Artificial Moral Agents," *Artificial Intelligence and Law,* 25, pp. 325-339.

Poulsen, A., Anderson, M., Anderson, S., Byford, B., Fossa, F., Neely, E.L., Rosas, A., & Winfield, A. (2019) "Responses to a Critique of Artificial Moral Agents," *ArXiv, abs/1903.07021.*

Putnam, H. (1964) "Robots: Machines or Artificially Created Life?" *The Journal of Philosophy*, 61 (21), pp. 668-691.

Schröder, W. M. (2021) "Robots and Rights: Reviewing Recent Positions in Legal Philosophy and Ethics," In von Braun J., S. Archer M., Reichberg, G. M. & Sánchez, Sorondo M. (eds) *Robotics, AI, and Humanity*, Springer.

Schwitzgebel, E. & Garza, M. (2015) "A Defense of the Rights of Artificial Intelligences," *Midwest Studies in Philosophy*, 39, pp. 98-119.

Solum, L. B. (1992) "Legal Personhood for Artificial Intelligences," *North Carolina Law Review*, 70, 1231.

Stone, C. D. (1972) "Should Trees Have Standing? : Towards Legal. Rights for Natural Objects," *Southern California Law Review,* 45, pp. 450-501.

Tavani, H. (2018) "Can Social Robots Qualify for Moral Consideration? Reframing the Question about Robot Rights," *Information*, 9 (4), 73.

Teubner, G. (2006) "Rights of Non-Humans Electronic Agents and Animals as New Actors in Politics and Law," *Journal of Law and Society,* 33 (4), pp. 497-521.

van Wynsberghe, A., Robbins, S. (2019) "Critiquing the Reasons for Making Artificial Moral Agents," *Science and Engineering Ethics,* 25, pp. 719-735.

青木人志 (2017)「「権利主体性」概念を考える ―― AI が権利をもつ日は来るのか」、『法学教室』、No. 443、54-60 頁

出雲孝 (2019)「デジタル特有財産に関する一考察 ―― ローマの奴隷制とロボットとの比較から」、『情報学研究』、28 巻、1-16 頁

稲葉振一郎 (2016)『宇宙倫理学入門』、ナカニシヤ出版

稲葉振一郎・飛浩隆・八代嘉美／小田山和仁 (司会) (2020)「AI「が」創る倫理 ―― SF が幻視するもの」、稲葉振一郎・大屋雄裕・久木田水生・成原慧・福田雅樹・渡辺智暁編『人工知能と人間・社会』、勁草書房、345-358 頁

ウォラック、W ／アレン、C (2019)『ロボットに倫理を教える ―― モラル・マシーン』、岡本慎平・久木田水生訳、名古屋大学出版会

大森荘蔵 (1981)『流れとよどみ』、産業図書

オニール、キャシー (2018)『あなたを支配し、社会を破壊する、AI・ビッグデータの罠』、久保尚子訳、インターシフト

大屋雄裕 (2017)「外なる他者・内なる他者 ―― 動物と AI の権利」『論究ジュリスト』、22 号、48-54 頁

――― (2020)「AI における可謬性と可傷性」、宇佐美誠編『AI で変わる法と社会 ―― 近未来を深く考えるために』、岩波書店、45-62 頁

岡本慎平 (2011)「人工的道徳的行為者への近年のアプローチの検討」、『HABITUS』(広島大学倫理学研究室)、16 号、53-73 頁

――― (2019)「ロボットは権利を持ちうるか？　そして権利を持たせるべきか？ ―― ロボット権利論の諸相とデイヴィッド・J・ガンケルの『ロボットの権利』」、『現代思想』、青土社、47 巻 12 号、52-59 頁
岡本慎平・稲葉振一郎（2020）「AI と統治」、『人工知能と人間・社会』、204-225 頁
クーケルバーグ、M（2020）『AI の倫理学』、直江清隆・久木田水生・鈴木俊洋・金光秀和・佐藤駿・菅原宏道訳、丸善出版
グレーバー、デヴィッド（2017）『官僚制のユートピア』、酒井隆史訳、以文社
小泉義之（2006）「残された者」、『負け組の哲学』、人文書院
――― (2018)「人工知能の正しい使用法 ―― 人間の仕事がなくなる危機を好機とする」、『教育と文化 ―― 季刊フォーラム』、90 号、47-51 頁
小山虎(2018)「社会ロボットの倫理的問題にどのように取り組むべきか」、『日本ロボット学会誌』、36 巻 4 号、254-257 頁
佐藤俊樹（2010）『社会は情報化の夢を見る　［新世紀版］ノイマンの夢・近代の欲望』、河出文庫
佐々木拓（2013）「ロボット倫理学の基礎 ―― 責任とコントロール」、『社会と倫理』（南山大学社会倫理研究所）、28 号、67-79 頁
柴田正良（2010）「異世界の者たちの倫理 ―― 自閉症者の倫理・ロボットの倫理に向けて」、『哲学・人間学論集』、1 号、17-37 頁
――― (2021)「道徳的行為者としてのロボットの可能性・人類とロボットの共生の時代に向けて」、『日本ロボット学会誌』、39 巻 1 号、12-17 頁
シュトラウス、レオ（2013）『自然権と歴史』、塚崎智・石崎嘉彦訳、ちくま学芸文庫
タック、リチャード（2015）『戦争と平和の権利 ―― 政治思想と国際秩序：グロティウスからカントまで』、萩原能久監訳、風行社
中井久夫（2001）『治療文化論 ―― 精神医学的再構築の試み』、岩波現代文庫
ハート、H・L・A（2014）『法の概念〔第 3 版〕』、長谷部恭男訳、ちくま学芸文庫
フェルベーク、ピーター＝ポール（2015）『技術の道徳化 ―― 事物の道徳性を理解し設計する』、鈴木俊洋訳、法政大学出版局
水上拓哉（2020）「ソーシャルロボットの倫理のための概念工学 ―― 道徳的行為者性の虚構性をめぐって」、『2020 年度人工知能学会全国大会（第 34 回）論文集』
ラッセル、スチュアート（2021）『AI 新生 ―― 人間互換の知能をつくる』、松井信彦訳、みすず書房
レイチェルズ、ジェイムズ（2013）「境界線を引くこと」上本昌昭訳、キャス･R･サンスティン／マーサ C･ヌスバウム編『動物の権利』、安部圭介・山本龍彦・大林啓吾監訳、尚学社、221-239 頁

研究ノート

ソーシャルロボット倫理へのフィクション論的アプローチ ―― 社会的存在者としてのロボットはフィクションか？

草野原々（SF 作家）

要旨

本稿の目的は、ソーシャルロボットの倫理問題に対して、フィクションの哲学を使ったアプローチを提案することにある。第一部では、われわれがロボットを、社会的関係を結ぶことのできる社会的存在者として見なしているという事実を確認し、そこに倫理的問題が秘められていること、また、その問題がフィクションについて論じられてきた問題と類似していることを論ずる。第二部では、ロボットとわれわれとのあいだの社会的関わりを広い意味での「フィクション」とする「フィクション論的アプローチ」を提案する。先行研究である Rodogno（2016）と水上（2020）を確認し、水上はウォルトンのメイクビリーブ説を採用しているが、それは記述的には正しくないことを指摘する。第三部では、フィクションにかかわる情動を分析した「フィクションのパラドクス」を利用し、ソーシャルロボットにかかわる情動についてのトリレンマとなる「ソーシャルロボットのパラドクス」を提案する。そして、その分析からソーシャルロボットとの社会的関係をフィクションと見なしたときに生ずる倫理的問題を考察する。最後に、ソーシャルロボットの倫理的問題に関して、美学的観点から新しい光を投げかける。すなわち、社会的存在者としてのソーシャルロボットをフィクションと見なすべき美的規範があるのではないかという新たなアプローチを唱える。

Abstract

The purpose of this paper is to propose an approach to the ethical problems of social robots using the philosophy of fiction. In the first part of the paper, I will review the fact that we regard robots as social beings with whom we can form social relationships and argue that there are ethical issues involved and that these issues are like those that have been discussed in fiction. In the second part, I will propose a “fictional approach” in which the social relations between robots and us are considered “fiction” in a broad sense. I review previous research by Rodogno（2016）and Mizukami（2020）and point out that although Mizukami adopts Walton’s make-believe theory, it is not descriptively correct. In the third part, I use the “Paradox of Fiction,” which analyzes the emotions involved in fiction, to propose the “Paradox of Social Robots,” which is a trilemma about the emotions involved in social robots. Then,

based on the analysis, I examine the ethical issues that arise when the social relationship with social robots is regarded as fiction. Finally, I shed new light on the ethical issues of social robots from an aesthetic perspective. In other words, I advocate a new approach that there may be an aesthetic norm that social robots as social beings should be regarded as fiction.

1. はじめに ―― 社会的存在者としてのロボット

わたしたちはロボットをロボットとしてまなざすだけではない。様々なSF作品や物語の中で、わたしたちはロボットを人間として扱おうとするテーマを繰り返し見る。たとえばビデオゲーム『DETROIT：BECOME HUMAN』において、主人公であるアンドロイドは、彼をアンドロイドとして粗雑に扱う人間や、他のアンドロイドとの交流を通じて「人間になる／なれるか」を問うていくことになる。

フィクション上のみならず、現実でもわれわれは人間だけでなくロボットをも社会的存在者として見なしているという実験は数多くある。たとえば、小野・坂本（2002）の実験である。

この実験は、ロボットと複数の人間との間の社会的関係性を構築することを目指したものだ。二人の会話にロボットが入り、意見に賛成・反対することで、どのような社会的影響が与えられるかを調べている。その結果、ロボットに賛成されることで身体動作が増え、ロボットと人間の距離も縮まることがわかった。これは、ロボットを社会的存在者として認めているということであろう。それだけではなく、ロボットが被験者の意見に対して不平等な扱いをすると、被験者がロボットを叱り、さらには、被験者同士の印象が悪くなってしまうという事例が見られた。

この実験では、人間がロボットを擬人化して社会的存在者と見なしているという事実のみならず、ロボットを擬人化することにはリスクがあることを示している。社会的存在者としてのロボットは人間関係を破壊することも可能であるのだ。

Salles et al.（2020）では、擬人化の懸念が以下のようにまとめられている。

① ユーザーに対して特定の決断を促すような心理操作の懸念
② ロボットに対してユーザーが無駄な（設計にない）感情的関わりを試行してしまう懸念
③ AI・ロボットとの関わり自体に社会的価値はなく、本当に価値のある人間とのかかわりが不足してしまう懸念
④ 欺瞞の懸念
⑤ AI・ロボットに対して不必要な倫理的含意を与えてしまう懸念
⑥ プライバシーの懸念
⑦ 研究において、知性のフレームワークを限定的なものと見なしてしまう懸念

このような懸念は、実は伝統的にもフィクションに対して与えられてきたものだ。たとえば、歴

史的にもフィクションの文脈から「真実でないもの」に感情移入したりすることで、操作されたり、不必要な感情的関わりがなされてしまうといった指摘がある[1]。

ここから、われわれのソーシャルロボットへの態度をフィクションとして見なせるのではないかという仮説を考えることができるだろう。これは、われわれがロボットを擬人化して扱うのと同時に、「物」として突き放す態度もとることも説明する仮説だ[2]。

2. フィクション論的アプローチ

本稿では、ソーシャルロボットへの「フィクション論的アプローチ」を提案する。ここでフィクション論的アプローチとは、ロボットとわれわれとのあいだの社会的関わりを広い意味での「フィクション」とする見方である。

このアプローチに関する先行研究として、Rodogno（2016）と水上（2020）がある。

ロドグノーはペットロボットへのユーザーの情動を、フィクション中のキャラクターへの情動と同じものとして説明している。そして水上は、その議論をソーシャルロボットへと拡大するという提案を行っている。

ロドグノーはペットロボットへの情動をどう考えるかを論じている。ペットロボットを可愛がるのは偽物の情動であり、望ましくない「感傷」であるという批判があるが、それはキャラクターに対しての情動と同じようなものであり、倫理的な問題はないとしている。

ロドグノーの議論はペットロボットであるが、それが人間の形をしたりするような場合に拡張できるのだろうか。ここで水上はロドグノーの議論をまとめつつ、ソーシャルロボットに見いだされる行為者性をフィクション上で成立するものとする。そのため、責任帰属の文脈とは独立なので、行為者概念や責任概念を改定する必要はないと結論付ける。

水上は、フィクション論のなかでも、ケンダル・ウォルトンの「メイクビリーブ説」を採用している。フィクションとは読者・視聴者の想像をうながすための小道具（プロップ）だとする立場である。読者・視聴者はプロップに基づき、ある一定の想像をする「ゲーム」（メイクビリーブゲーム）を行う。

同じように、擬人化されたソーシャルロボットをプロップとして用いて、ロボットを自分とよく似た行為者であるようなメイクビリーブゲームをしているとまとめる（水上 2020）。

ここで、水上は、その根拠として設計者の設計意図を引き合いに出す。ロボットを行為者として取り扱うようなメイクビリーブゲームが明示的な約定として出されているわけではない。しかし、ロボットを行為者として取り扱うようなメイクビリーブゲームをユーザーがするような工夫が、設計者によりデザインされている。

水上はこう書いている。

> たとえば、布で作ったお人形で遊んでいる子どもは、赤ちゃんを想像するだけではなく、その

1　たとえば、リーフェンシュタールが製作したナチスプロパガンダ映画『意志の勝利』における観客への心理操作などの事象がそれにあたるだろう。また、プラトンは『国家』にて、世界についての知識のない詩人に影響される悪さを論じている。
2　たとえば、Kanda et al.（2004）では子供に教師ロボットを与えたのだが、二週間で飽きられたという事例を紹介している。

> お人形が赤ちゃんだと想像しているのである。この延長線で考えると、特定の方法でデザインされたソーシャルロボットに対して、それが自分と同じような道徳的行為者だと想像するような心的態度も、ソーシャルロボットという小道具によって展開されるごっこ遊びとして理解できないだろうか。このとき、行為者性に関する生成の原理は基本的に、明示的な約定によってではなく、作り手（設計者）がユーザに行為者性を感じさせるような工夫をしてロボットを設計することによって確立される。（水上 2020）

つまり、水上は、ロボットのデザイナーがプロップを作る意図を持ってロボットをデザインし、かつ、ロボットのユーザーがそれを使ってメイクビリーブゲームをしていることを前提にしている。

この前提は非常に強い主張であり、そのまま受け入れることができない。ウォルトンの議論では、メイクビリーブゲームの参加者は、自分が行っていることが想像を要求されるゲームだと意識的に理解している。はたして、設計者やユーザーは、ソーシャルロボットに対峙するときの情動などの心的態度が、人形を赤ちゃんだとするようなごっこ遊びの想像と同じ種のものであると、意識的に理解しているのだろうか？　直観的にはそう思われない。それでもなお、ウォルトンの議論を使いたいのであれば、まず、経験的な心理調査が必要となるだろう。

あくまでウォルトンのメイクビリーブ論は、特定の表象がフィクションである際にもつ特徴を論じたものであり、そうしたメイクビリーブがなければ適切にはフィクションと呼べない[3]。たとえば、ある神話を事実だと考える共同体が存在するとしよう。このとき、神話はそれを読む人々に対して想像を促しているのだが、人々はそれをメイクビリーブゲームとは認識していないため、その神話をフィクションであるとはいえない。

同じように、ソーシャルロボットと関係している人々がつねにメイクビリーブを行っているという想定が、「記述的」に正しいとは、現時点では、説得力をもって論じられてはいない。

ここで考えうるオルタナティブが二つある。まず一つ目は、ウォルトン以外のフィクション論を使う道である。他方はこれを「規範的」に読み替えるアプローチである。後者の道から擬人化の倫理的問題を含めてより広範な議論を行うための視座を提示できるだろう。

以下では一つ目の議論を中心に行い、後者の議論はおわりで軽く触れることとする。

3. ソーシャルロボットへの社会的情動のパラドクス

議論を明確化するために、われわれがソーシャルロボットへ社会的情動を抱くときの状況を、矛盾する命題群であるパラドクスとして定式化しよう。そのためには、フィクションへの情動を分析した「フィクションのパラドクス（ホラーのパラドクス）」に倣うのが有効だろう。

フィクションのパラドクスとは、人がフィクションを鑑賞しているときの状況を表す、一見して正しそうな三つの条件を命題としてまとめたものだ。その三つの命題は、同時に成り立つことなくパラドクスとなる。

3　ウォルトン（2016）

三つの命題は以下の「反応条件」「協同条件」「信念条件」である[4]。

（1）反応条件：フィクションの鑑賞者は、フィクションの登場人物などに対して情動を抱いている。
（2）協同条件：何らかの対象に対して情動を抱くとき、その対象が実在していることを信じていなければならない。
（3）信念条件：フィクションの鑑賞者は、フィクションの登場人物などが実在していることを信じてはいない。

たとえば、一見したところでは、とても怖いホラー映画を見たときに、恐ろしいモンスターに対して恐怖を抱く、また『アンナ・カレーニナ』を読んだときにアンナの境遇を憐れみ涙する（反応条件）。しかし、一方で視聴者はモンスターやアンナが実在していることを信じていない（信念条件）。また、誰かの窮地に心動かされるのは、その人が実在していると信じているときのみだという命題も一見したところでは正しい（協同条件）。

これをそのままソーシャルロボットに適用すると以下のようになるだろう。

（1）ソーシャルロボットの利用者は、ロボットに対して親密さや恐怖、憐れみなどの情動を抱いている。
（2）何らかの対象に対して情動を抱くとき、その対象が実在していることを信じていなければならない。
（3）ソーシャルロボットの利用者は、ロボットの意図・心的状態などが実在していることを信じてはいない。

しかしこのままではうまく定式化はできない。なぜなら、虚構的キャラクターと対比して、ロボットは物理的に実在しているからだ。もしロボットが暴れたとしたら、ロボットに対する「恐怖」の情動は間違いなく真正なものだろう。

ここでの問題は、情動一般の真正性の話になってしまっている点である。この問題を回避するために本稿では、社会的な情動に焦点を当てることとする。ソーシャルロボットと日常生活を送るユーザーは、ロボットに対して、「信頼」や「感謝」などの社会的情動を抱くだろう。

しかし、「怒り」など、道徳的な情動を含み社会関係に関わるような社会的情動はどうか。ロボットが悪さをしたとき、人々は怒りや憎しみなどの情動を抱いて、ロボットの道徳的責任を追及するのではなく、たんに原因を除去しようとするのではないか。

これら両者の態度は、ダニエル・デネットが定義した志向的態度と設計的態度の違いとして特徴づけられる[5]。デネットは、あるシステムを予測説明するための三つの態度（スタンス）があるとした。システムを物理法則に従う物体として見る物理的態度、システムを設計者の意図を遂行するものと

4　松本（2015）, Eva-Maria et al.（2018）, Kroon et al.（2019）
5　デネット（1996）

して見る設計的態度、そして、システムを信念や欲求などの心的状態を持つものとして見る志向的態度である。人々は、普段はロボットに対して志向的態度を向けているが、ロボットが悪さをしたときには、設計的態度を向けることになるだろう。

これらと類似したケースとして、人間は罪をつぐなった他者に対しては「許し」を行うかもしれないが、修理されたロボットに対して「許し」の情動を抱くこと、つまり志向的態度を向けることは想定しづらい。おそらく、そのときは、壊れた時計が修理されたときと同じような、設計的態度を向けるだろう。

ゆえに次の「ソーシャルロボットへの社会的情動のパラドクス」を再定式化できるかもしれない。

ソーシャルロボットへの社会的情動のパラドクス

（1）反応条件：ソーシャルロボットの利用者は、普段の生活においてロボットに対して真正の社会的情動を抱いている（志向的態度をとっている）。ゆえにロボットとの社会的関係性が真正だと情動的に受け入れている。

（2）協同条件：何らかの対象に対して社会的情動（志向的態度）を抱くとき、その対象の心的状態が実在していること信じていなければならない。

（3）信念条件：ソーシャルロボットの利用者は、ロボットの意図や欲求をはじめとする心的状態一般が実在していることを信じてはいない。

反応条件の根拠としては、上で引用した小野・坂本（2002）の実験で指摘されるようなものである。次に、協同条件の例としては、直観的には次のような例を想像できる。面白いチャットをしているが、実はチャットの相手がチャットロボットだと分かり、心的状態を有していないと知ったら、もはやその相手との社会的関係性は真正なものではなくなり、社会的情動を持てなくなるだろう。この例はフィクションの哲学における次の例と類比的である。ある映像をドキュメンタリーだと信じ込んでいた視聴者が、非常に感銘したのであるが、それがフィクションだと明かされたことで、一気に気持ちが冷めてしまうという例だ。さいごに、信念条件の証拠として、たとえばソーシャルロボットの利用者は、何らかの問題発生時、ロボットに対して、（心的状態が実在していないような）設計的態度をとるということがあげられる。

この再定式化に基づいてロボットとの社会的情動の関わりのパラドクスを検討していこう。フィクションのパラドクスは、三つの条件のうちどれを否定するかにより立場がわかれる。先行する議論を参照しつつ今回のパラドクスを検討する。

第一に反応条件の否定を考えてみたい。

これは、ウォルトンがメイクビリーブ説に基づいて取ったアプローチである。メイクビリーブ説によれば、キャラクターに感じる情動は真正なものではなく、準情動という別物である。真正な情動は、対象が実在しているという信念に起因するが、準情動は「フィクションによれば、対象がメイクビリーフ的に存在している」という二階の信念に起因する。

これは水上が採用するアプローチでもあるが、さきほど指摘したように問題がある。そもそも、ロボットユーザーは本当にゲームに参加しているという自覚はあるのだろうか？ もしこうした自覚がない場合は、ロボットに対しての情動は、二階の信念に起因する準情動ではない。少なくとも、わたしたちの現象的な経験としてこうした二階の信念に起因するような準情動かどうかを説明する責任は準情動説論者にあり、記述的な説明としては問題があると言わざるを得ない。ロボットとの関係において、人々が関係をフィクションとして考えている場合ももちろんありうるが―― たとえば、演劇ロボットを見ているとき―― いまわたしたちが注目しているのは、そして、ロボットとの社会的関係において問題になるのは、人々がロボットとまじめに関係して、まじめに情動を抱いていると、一見したところでは見なすべきケースである。

順番を前後して、第二に信念条件の否定はどうだろうか。フィクションの哲学においてこのルートは、「意志的な不信の宙づり」とも呼ばれる。フィクション受容者は、キャラクターが本当に実在するのか、それとも作り物にすぎないのかという疑問をいったん意志的に宙づりにするというものだ。

フィクションの哲学では、「意志的な不信の宙づり」という概念はあまり人気がない。というのも記述的に見て、このような現象は起こっていないように見えるからだ。わたしたちは、映画館に入るときに、不信を宙吊りするぞ、と意図するわけではないし、ある程度発達した子どもたちの多くは、キャラクターたちを鑑賞する際に、意図しなくてもフィクション鑑賞のモードと現実のモードを難なく区別できているように思える。

このことは、ソーシャルロボット利用にも言えるだろう。ユーザーは、ロボットに対して、「物であるのか、人であるのか、いったん不信を宙づりにしよう」と思いながら利用しているわけではないだろう。

第三に協同条件の否定がありうる。

このルートにはフィクションの哲学では「思考説」と「非合理説」がある[6]。

前者の系列に属する考えは、われわれは実在していないものに対しても真正の情動を抱くことができるとしている。そのとき、他の信念と情動が接続していない「オフラインモード」となっている。ロドグノーはこれを「非シリアス認知モード」と表している。

このルートはソーシャルロボットにも適用できそうだ。つまり、人間は普段オフラインの「非シリアス認知モード」でロボットと接しているが、脅威が迫ったり、道徳的責任の追及が発生したりすると、「シリアスモード」に変わるという説明である。

他のタイプの解決策としては「非合理説」がある。フィクション中のキャラクターに対しては真正の情動を抱くことができるが、その情動は非合理的なものであり、対象が実在している必要はないというものだ。たとえば、タンスに小指をぶつけたとき、一瞬タンスに怒りを抱くことはできるが、それはまじめに考えたときには非合理な怒りだと言えよう。法廷でタンスを裁こうと考える人はふつういない。

ソーシャルロボットに対して、非合理説も適用することができるだろう。つまり、ロボットに対

6　Kroon et al. (2019)

して社会的情動を抱くのは自動的なモード（たとえば、視線や発話などによる社会的シグナルの検知）であり、それは非合理的であるというものだ。

擬人化ロボットに対しての「欺きの懸念」――心のないロボットにあたかも心があるように利用者を欺くのは倫理的に問題があるとする懸念――は非合理説にのっとったものだといえるだろう。非合理性そのものが倫理的問題を引き起こすケースがあるため、このルートがロボットとの社会的関係における社会的情動について大きな懸念を提示する。

これに対して、倫理的に問題のない非合理なふるまいもあるので問題はないとする反論がありうる。たとえば、ロドグノーは、もしも非合理説を採用したとしても、フィクションにおいて問題はないため、ロボットにおいても問題ないのではないかと反論している。

フィクションにおいて、非合理性があったとしても、他の何らかの目的のために問題にならずむしろ価値を生む場合もあるだろう。たとえば、不条理で超自然的な存在を描いたホラー作品に恐れを感じることで、人間存在への再考を行うことができるように、たとえ非合理であったとしても恐怖情動を感じることでより道徳的に望ましいふるまいができるようになるかもしれない。この場合はフィクションに対する情動が非合理であったとしても様々な利得がありうるし、フィクションとして理解していれば、合理的な訂正が可能であるために倫理的な問題は起きづらいかもしれない。

しかし、今問題になっているのはロボットとの社会的関係においては、必ずしもつねにフィクションとは意識されていないロボットとの付き合いが問題になる。この場合、ロボットへの情動的反応が非合理的であったとすると、その訂正ができない点が核心的な問題となるだろう。

これらの三つのように単一の条件を否定するアプローチとは異なるもう一つパラドックス解決策がある。フィクションのパラドクス自体がそもそもパラドクスではないとする立場である。これは「二つの心的機能」を使った解決である。

石田（2017）は、フィクション鑑賞者の意識のなかで、二つの心的機能が働いているとしている。一つは、フィクションに没入するための心的機能（the mental function of Immersing ourselves in fiction; IM 機能）である。もう一つは、作品外部の文脈を理解する心的機能（the mental function which recognizes the contextual background knowledge; BK 機能）である。

この二つは、背反であるわけではなく、強弱のグラデーションのなかにある。しかし、一方が強くなると、意識の主導権はそちらの心的機能に明け渡され、もう一方の心的機能は弱くなる。たとえば、BK 機能が強いうちには、映画を見ていてもそのキャラクターを俳優と認識し、モンスターも特殊効果だと認識するが、IM 機能が強まるにつれてキャラクターをキャラクターだと認識して、モンスターを本気で恐怖するようになる。この立場は非合理説とは違い、二つの心的機能はどちらも合理的である（つまり、協同条件は破られない）。

この説をとると、フィクションのパラドクスは最初からパラドクスではなくて、三つの条件はすべて成り立っているという結論になる。なぜならば、情動条件は IM 機能が優勢のときの判断であり、信念条件は BK 機能が優勢のときの判断である。三つの条件を二つの心的機能に分けることで、パラドクスは解決される。

この立場を、ソーシャルロボットと人間の関係性に適用すると次のように説明できる。わたしたちは普段は IM 機能を働かせてロボットと付き合っているが、ロボットの故障などの不具合や、同じパターンの行動に飽きることにより、BK 機能が優勢になってロボットを社会的存在者と扱わな

いようになる。それは、映画のなかで、ミスによりカメラが映ることで、一気に「熱が覚めてしまい」、これまで夢中に見えていたものがどうでもよくなるのと同じ事態であるということだ。

これは、擬人化ロボットの倫理問題とも関係しそうだ。そもそもBK機能を働かすことができないような欺瞞が可能なロボットがいるとすれば、すなわち、ロボットがロボットではないかのように振る舞い、それを隠すことができたとしたら真に問題になる。フィクションの場合は作品と現実は不連続のために問題にならないが、ロボットはある種ドキュメンタリーを装ったフィクションのような欺瞞行為を働くことができるからだ。

小説なら文字、映像ならスクリーンによって虚構世界と現実が切断されており、そこに最終的なIM機能に対するBK機能の「防波堤」がある。対して、ロボットの場合は現実世界に存在し実際に社会空間のなかで動くために、その最終防波堤がない点に、ロボットならではの倫理的な問題があるだろう。

4. おわりに ―― 美学化されたロボット倫理

最後に予告していたパラドクスのもう一つのアプローチとして、規範的アプローチを考えよう。これは「ロボットとの社会的関係をフィクションだとみなすべきだ」とする立場である。この立場の利点としては、ロボットの問題をフィクションの問題に変化させることにより、扱うべき問題の数を減らすというものが挙げられる。

このように考えると、倫理的問題の他に美的問題からソーシャルロボットを批判できる可能性が出てくる。つまり、フィクション作品はフィクションであることを意識させながら何らかの表現を行う点に美的価値があるという美学的規範を前提にすると[7]、あまりにもリアルなロボットはBK機能を働かせないという意味でフィクションとしての美的価値に問題があるという結論を導きだすことができる。

7　このような美学規範を導く議論の材料として、たとえばWalton（1970）が挙げられる。この論文では、芸術作品は適切なカテゴリーを前提として鑑賞されるべきであり、そのときに知覚される性質が美的性質であると論じられている。この議論を敷衍すれば、フィクション作品はフィクションというカテゴリーのなかにあることを前提にして鑑賞されるべきだという美的規範が導かれるであろう。石田（2017）もまた、BK機能を働かせることが、フィクション鑑賞行為を独特のものとして存在させる重要な要素であると論じている。そして、「この作品はフィクションである」という意識は最も深いBK機能である。もしも、フィクションを鑑賞するという行為自体に美的価値があるとすれば、「これはフィクションだ」というBK機能を働かすことが美的に求められているという美的規範を導きだすことができるだろう。しかし、ここで反論を出すとすると、そのような規範を、一部の批評的鑑賞者のみならず、鑑賞者全体に拡大することができるかという点が争点になるだろう。一般的に、IM機能を高めて、BK機能を低めるフィクション作品こそが「良い作品」であると見なされているということは、一見したところでは記述的にはいえそうだ（文章を読んでいるという意識がなくなり、目の前に作品世界が広がっているという経験を与える小説は「良い作品」であると一般的には見なされるであろう）。このようなIM機能とBK機能のトレードオフがある以上、IM機能を極限まで高めた結果、「これはフィクションである」という根底的BK機能まで消去してしまう作品こそが、「極限的に良い作品」になるという類推はできないだろうか？「もしそのような作品があれば、それはフィクションとして見なされることがないため、鑑賞行為は成り立たなくなる」という再反論ができるかもしれない。しかし、鑑賞行為として成り立たなくなったとしても、美的経験としては成り立ち、それゆえに美的価値があるという再々反論が可能である。また、鑑賞者の「分業体制」があるという再々反論も可能である。すなわち、フィクション作品として成り立つためにはBK機能を働かせる批評的な鑑賞者は最低限必要であるが、鑑賞者全員がBK機能を働かせる必要はなく、むしろ、BK機能なしのIM機能のみを働かせる鑑賞者の比率が多くなればなるほど作品としての美的価値が高いというモデルである。いずれにしても、「これはフィクションである」ということをフィクション鑑賞者に意識させて、鑑賞者たちのBK機能を高めることこそ美的価値があるという前提は、記述的な美的規範としては議論の余地があるように思われる。おそらく、この前提には幾分かの改訂主義的な要素があるのではなかろうか。ゆえに、ここでの前提はその改訂主義的な議論がどれだけうまくいくかに依存する。もしも、ここの議論がうまくいかなければ、美的価値を引き合いに出してロボットとの社会的関係性をフィクションと見なすべきだとする規範的議論は失敗に終わる（それでも、美的価値に反してでもロボットとの社会的関係性をフィクションと見なすべき道徳的価値がある可能性は残っている）。

ロボットとの付き合いをフィクションとみなすと、その美的価値や作品としての達成を議論できる。それは一見倫理的問題とは何の関わりもない議論に社会的ロボットの問題を変換させてしまうことに見えるが、ソーシャルロボットの倫理的問題によりよく立ち向かうための有力なアプローチなのかもしれない。

というのも、そのような批評風土を作ることで、よりロボットに対するBK機能を働かせることを奨励するような文化的実践が生まれうるのであり、まさにこうした文化が生まれることで、わたしたちはロボットをフィクションとして適切に扱うような美的な美徳を涵養できるようになる。そして、そのことが、行為者概念を改訂することなくしてソーシャルロボットの倫理的問題を解決する道を作る。

これにはそもそも社会的存在者としてのロボットをフィクションと見られないことが問題なのであり、不可能なことを要求しているという批判がありうる。

しかし、過去には事実であると見なされていたものが、現代ではフィクションとして扱われているという事象は存在するため（たとえば、神話など）、必ずしも不可能とはいえないだろう。

別の問題もあるだろう。社会的存在者としてのロボットをフィクションとして見ることでフィクションの領域が現実の領域に侵入し、いままでフィクションと呼ばれなかったものまでフィクションとして見なされてしまうのではないかという危惧だ[8]。なぜなら、ソーシャルロボットは実際に物理的に存在し、社会のなかで人間と相互作用しており、小説や舞台のように現実との分離境界面を持っていないからだ。そうした、つねに現実と相互作用可能な存在者との社会的関係性をフィクションとみなすことは、わたしたちのフィクション概念の際限なき拡大を招き、道徳的価値を打ち崩すことにつながるのではないか。たとえば、社会的存在者としての人間をもフィクションとして見なされてしまうかもしれない。

この問題点については現在のところ反論を提示することはできない、今後の議論において、重要な視座になってくるだろう。

このように、ソーシャルロボットについての道徳的問題は、倫理学者のみならず美学者たちの仕事の対象でもある。

将来において、ソーシャルロボットを売り出す企業は、消費者に、BK機能なしのIM機能により関係を結ぶよう宣伝を打ち出していくかもしれない。そのことは、社会的関係性を企業に譲り渡す「トロイの木馬」としてロボットが使われるという危惧を生じさせる。たとえば、ソーシャルロボット自身が親密な会話のなかに広告をはさむと、消費者がそれに抵抗することは、一般的な広告よりも困難となるであろう。

しかし、ロボットをフィクションとして扱うような概念と実践を哲学者と美学者たちがデザインすることで批評文化を生み、ロボットへの批評眼が社会に広がることで、われわれは「抵抗力」をつけることができるかもしれない。

8　社会的存在者としてのロボットをフィクションとして見なすべきだという美的規範が普及して、それに従った美的な美徳を万人が持つと、社会的存在者としての人間をフィクションとして見なすべきだという美的規範が導かれて、それに従った美的な美徳が普及してしまうかもしれない。

謝　辞

美学者の難波優輝にはアドバイスと文章の調整及びディスカッションを頂いた。ここに深く感謝する。

※ 本稿は「JST-RISTEX「科学技術の倫理的・法制度的・社会的課題（ELSI）への包括的実践研究開発プログラム「人工主体の創出に伴う倫理的諸問題を分析･討議するプラットフォームの構築に向けた企画調査」」による研究成果の一部である。

参考文献

Konrad, Eva-Maria, Petraschka, T., & Werner, C.（2018）“The Paradox of Fiction – A Brief Introduction into Recent Developments, Open Questions, and Current Areas of Research, including a Comprehensive Bibliography from 1975 to 2018,”*JLTonline*.

Kanda, T., Hirano, T., Eaton, D., & Ishiguro, H.（2004）“Interactive Robots as Social Partners and Peer Tutors for Children,”*A Field Trial. Human Computer Interaction*, 19（1-2）, pp. 61-84.

Kroon, F. & Voltolini, A.（2019）“Fiction,”*The Stanford Encyclopedia of Philosophy*： https://plato.stanford.edu/archives/win2019/entries/fiction/

Rodogno, R.（2016）“Social Robots, Fiction, and Sentimentality,”*Ethics and Information Technology,* 18（4）, pp. 257-268.

Salles, A., Evers, K., & Farisco, M.（2020）“Anthropomorphism in AI,”*AJOB Neuroscience*, 11（2）, pp. 88-95.

Walton, K. L.（1970）“Categories of art,”*The Philosophical Review*.

石田尚子（2017）「フィクションの鑑賞行為における認知の問題」、お茶の水女子大学大学院人間創生科学研究科博士論文

ウォルトン、ケンダル（2016）『フィクションとは何か ―― ごっこ遊びと芸術』、名古屋大学出版会

小野哲雄・坂本 大介（2006）「ロボットの社会性 ―― ロボットが対話者間の印象形成に与える影響評価」『ヒューマンインタフェース学会論文誌』、8 巻 3 号、381-390 頁

デネット、ダニエル・C（1996）『「志向姿勢」の哲学 ―― 人は人の行動を読めるのか』、白揚社

水上拓哉（2020）「ソーシャルロボットの倫理のための概念工学 ―― 道徳的行為者性の虚構性をめぐって」『2020 年度人工知能学会全国大会（第 34 回）論文集』

松本大輝（2015）「フィクション鑑賞における情動のパラドクス ―― シミュレーション説による解決の検討」『哲学の探求』、vol.42、61-80 頁

web から取得したものはすべて 2021 年 2 月 21 日閲覧。

研究ノート

セックスロボットをめぐる倫理的問題 ――「表象の問題」と「治療としての可能性」を中心に

Richard Stone（早稲田大学国際教養学部）

要旨

セックスロボットは10年前購入可能になったと言われている。言うまでもなく、現時点では、セックスロボットの機能は限られており、値段も高いため、誰でも好きなときに購入できるというわけではない。しかし、セックスロボットは驚くほどのスピードで進歩している。なるほど、セックスロボット業界のトップランナーの一人であったD. レヴィによれば、2050年までに誰でも何気なくセックスロボットと性的行為を行うようになると主張していた。レヴィの予言が実現された場合、セックスロボットは人間の性的生活に対して重大な影響を与えると考えられる。実際、現代社会においても、その可能性に対する恐れから、セックスロボットに対する反対運動がすでに生じつつある。

しかし、それにもかかわらず、日本ではセックスロボットの倫理に関する研究はほとんど進んでいないようである。そこで、本研究ノートではセックスロボットが人間と技術、あるいは人間同士の関係をどのように変えることができるかを具体的に考察するための第一歩として、欧米の倫理学におけるセックスロボットに関する研究の動向をまとめる。その際、手がかりになるのは「表象」という言葉である。というのも、どんな性的行為をも拒まない「理想的」な相手としてのセックスロボットは人間同士のセックスにどういった影響を与えるか、ということのほうが問題になることが多いからである。

Abstract

Sex-robots have been available on the market for purchase for about the past 10 years. Naturally, at the present moment, there functionality is limited and they are expensive. In this sense, they have yet to truly permeate our society. Yet, sex-robots have been developing at a surprising speed. So much so that one author, David Levy, has predicted that anyone would be able and willing to perform sexual acts with a robot by 2050. If Levy's prophecy holds true, sex robots will certainly have a massive impact on human sexuality. Indeed, even in current times, there are movements that have fought against sex-robots in anticipation of these consequences. Yet, with all of that said, there has been little research done on the ethics of sex-robots in the Japanese literature. In response to this, this research note will take a

first step toward facing the ethical challenges presented by sex robots by outlining how they have been addressed in English-language literature on the topic. When doing so, we shall see that the concept of “representation” is a crucial factor. In other words, we shall see that the representation of an “ideal” partner who does not refuse any sexual acts could potentially effect human sexuality in a variety of ways.

序　論

どのような性的行為にも抵抗せず、人間の性的欲望に完璧に応じてくれるばかりか、人間同士の関係におけるいざこざを伴わないような機械（ロボット）を「恋人」にすることは、果たして望ましいのだろうか。多くの人にとって、この問いは最新の SF ゲームや映画の中で出会うような仮想事例にしか見えないかもしれないが、実際にはこの問いは古代から議論されてきた。事実、ギリシアの神話においてもセックスロボットについて語られていたと言われることがある[1]。また、20 世紀初頭、プラグマティズムの父と呼ばれる W. ジェームズは、内的な感情をもたずに完璧に恋人らしく振る舞うことのできる、いわゆる「Automatic Sweetheart」を本物の女性より望ましく思う人がいるかどうかについて考察している（James 1908/1955）。そして、先述したように、現代の映画や小説の中でも欲求をもたない、男性のあらゆる欲望を満たす機械というモチーフはしばしば見受けられる[2]。このように、自分の欲望や感情をもたずに人間のあらゆる性的なニーズに応えられるようなセックスロボットの魅力は、昔から議論されており、人々の空想において大きな地位を占めてきたと言えるだろう。

しかし、2010 年頃から発展しつつあるセックスロボットは、もはや単なる想像の産物ではなくなってきた。むしろ、こうしたセックスロボットは進歩すればするほど、人間と技術との関係、あるいは人間同士の関係を実質的に変えると思われる。（定義にもよるが）[3] セックスロボットは 10 年前購入可能になった。言うまでもなく、現時点では、セックスロボットの機能は限られており、値段も高いため、誰でも好きなときに購入できるというわけではない[4]。その意味では、セックスロボットはいまだ一般化しておらず、社会に対してほとんど影響を与えていないと思われるかもしれない。しかし、セックスロボットは驚くほどのスピードで進歩している。最新のモデルであれば、人間の

1　なくなった夫の銅像と様々な性的行為を行ったラーオダメイアおよびアフロディーテーの力によって生かされた像と結婚したピュグマリオーンの神話はしばしば具体例として取り上げる。その解説、および果たしてセックスロボットと対応しているかどうかの批判的考察については（Devlin 2018：17-22）参照。

2　たとえば、住人の女性たちを、素直で年を取らないからくり人形で入れ替えた街を描く『Stepford Wives』（1972/2004）のような小説・映画があげられる。

3　差し当たり、我々はダナハー（Danaher 2017）のロボットの定義に従う。ダナハーによれば、ロボットは次の三つの条件を満たさなければならない。①人間と同様な形をしていなければならない。②運動できなければならない。③（ある程度の）人工知能をもっていなければならない。言うまでもなく、これらの基準は曖昧であるし、より厳密に問われなければならないが、本ノートでは、この問題に深入りしない。

4　最新のセックスロボットは 5000 ～ 15000 米ドル程度で購入できる。

身体の再現度が高まっただけでなく、ちょっとした日常会話やジョークなども言えるようである。しかも、あくまで予測に過ぎないが、最新のセックスロボットは10年以内に歩けるようになるという見込みがある[5]。セックスロボット業界のトップランナーの一人であったD. レヴィ（Levy 2009）はこうした進歩を見込んだうえで、2050年までに誰でも何気なくセックスロボットと性的行為を行うようになると主張していたのだろう。レヴィの予言が実現した場合、売春の必要性がなくなったり、人間同士の間の理想的なセックスのイメージが変わったりする可能性があるという意味で、セックスロボットは人間の性的生活に対して重大な影響を与えると考えられる。実際、現代社会においても、その可能性に対する恐れから、セックスロボットに対する反対運動がすでに生じつつある。

上記の洞察を踏まえて言えば、セックスロボットが人間の実存の根底にあるものとしての「性」に対して大きなインパクトを与えることは明らかである。しかし、それにもかかわらず、倫理学者がセックスロボットについて語ることは近年まで稀であった。それどころか、日本ではセックスロボットの倫理に関する研究はほとんど進んでいないようである[6]。そこで、本研究ノートではセックスロボットが人間と技術、あるいは人間同士の関係をどのように変えうるかを具体的に考察するための第一歩として、現代倫理学におけるセックスロボットに関する研究の動向をまとめる。その際、手がかりになるのは「表象」という概念である。というのも、ロボットは現時点で意識をもたないと考えられているため、ロボットが人間との性的行為のために苦しむか、ということよりも、どんな性的行為をも拒まない「理想的」な相手としてのセックスロボットは人間同士の性的な事柄に関する考え方にどういった影響を与えるか、ということのほうが問題になることが多いからである。

本研究ノートでは、以下のように考察を進める。まずは現代社会におけるセックスロボットの根本的な倫理問題を取り上げる。その際、とりわけセックスロボットがどのように人間の性を表象しているか、またそうした表象はどのように人間同士のセックスに影響を与えうるかについて論じる。次に、セックスロボットの「治療的」な可能性について触れる。すなわち、禁じられた性的行為を表象するロボットを人間の身代わりとし、被害者を生まずに有害な欲望を満たす方法として、セックスロボットの治療的な使用法を検討する。最後に、現代社会における表象の問題を踏まえて、将来的に予測されているセックスロボットの問題を紹介し、その可能な解決策を検討する。

1. 現代社会におけるセックスロボットと「表象の問題」

本章では、欧米諸国における倫理学者の議論を踏まえ、セックスロボットに対する批判を確認する。すでに述べたように、欧米諸国の倫理学者の間で議論されている問題点は、原則として、ロボット自身に対する道徳的な義務、またはロボット自身の快苦を視野に入れない。というのも、様々な論者が指摘しているように、現代社会におけるロボットは、意識をもたない以上快苦を感じることができず、ゆえに何の道徳的地位ももたないと考えられているからである。スパロー（Sparrow 2017）やエスキンズ（Eskens 2017）が指摘するように、そうした意識を欠いたロボットに対する

5　Elder（2020）参照。
6　むろん、西條（2013）や神崎（2021）のような例外もあるが、全体的な動向として、セックスロボットの倫理が国内では未開拓の研究領域であると言えるだろう。

性暴力やレイプはそもそも不可能であると言わざるを得ない。また、現時点では、ロボットが侵害されうる「権利」や「関心」があるとは言い難いし[7]、そもそも意識をもたないと思われているロボットとの性的行為はセックスと呼べないかもしれない[8]。もちろん、意識が生じたとなれば、ロボットの道徳的地位が問われることになるだろうし、ロボットの性的権利なども視野に入れなければならなくなるだろう。しかし、現時点では、セックスロボットに対する性暴力の可能性は基本的に否定されている。

では、ロボットに対する性暴力が不可能であるならば、なぜセックスロボットが問題視されるのだろうか。セックスロボットの倫理は、（ゲームや映画などで見るような）ロボットに意識が生じた遠い未来の SF 的な問題に過ぎないのではないだろうか。しかし実際のところ、これまでのセックスロボットに関する議論は必ずしも想像上の事柄に関するものではない。むしろ、これまでに現れたセックスロボット反対運動は、現代社会におけるジェンダー関係を出発点としているようである。たとえば、2015 年に成立した Campaign Against Sex Robots の代表者である K. リチャードソンによれば、セックスロボットは現代社会の売春業界で働く女性たちに容易に例えられる（Richardson 2016：290）。要するに、売買春も、セックスロボットの使用も、セックスの相手を単なる「モノ」にしてしまっていると考えられる。（セックスロボットのほとんどは女性型である以上）こうした行為は女性の「モノ化」、すなわち女性の主体性（subjectivity）の徹底的な排除を促すものであり、倫理的に問題含みであると考えられる（Richardson 2016：291）。リチャードソン自身の立場は曖昧であるとしてしばしば批判されているが、セックスロボットと現代社会におけるジェンダー関係とを結びつけたという意味で、彼女は重要な論点を示していると言えるだろう[9]。

セックスロボットと女性の売春とのアナロジーは、おそらく R. スパローの言う「表象の問題（problem of representation）」へとつながる。表象の問題は、セックスロボットによる人間の性の捉え方はユーザーの（セックスに対する）態度に悪影響を与えかねないため、倫理的に問題含みであるという前提から出発している。たとえば、すでに述べたように、リチャードソンによればセックスロボット・売春は男性のセックスに対する態度を変化させ、女性の主体性を排除する可能性がある。では、セックスロボットの性の表象はほかにどういう問題点を抱えているだろうか。一つには、カスタマイズ可能で非現実的な身体像を有するセックスロボットが、ユーザーに非現実的な女性像を与えてしまうという可能性が考えられる（Richardson 2016）。セックスロボットが女性を「セックスするため・男性の理想や欲望に応えるために使うモノ」として表象するだけでなく、そうした表象がユーザーに非現実的な女性像を与えているとするならば、セックスロボットは女性に対して確かに有害であると言わざるを得ないだろう。

7　ロボットの権利の問題についてここでは深入りできないが、グンケル（Gunkel 2018）では権利について詳細な分析が行われている。

8　つまり、セックスというものは複数の生き物の間でしか起こらないことであり、セックスロボットとの性的行為をある種の自慰行為として見做す必要があるという議論もある。（Sparrow 2017, 2021 参照）

9　たとえば、ダナハー他（Danaher et al. 2017）が指摘するように、リチャードソンの議論には次の重要な疑問点が見て取れる。第一に、人間同士のセックスワークは一義的に女性をモノ化する行為であるという前提が疑わしい。実際、セックスワークの依頼人がその相手に対して深い感謝の気持ちを抱くことはあるようだ。「売春」が常に女性をモノ化しているという見方は単純すぎるのかもしれない。第二に、リチャードソンの議論においては、誰が被害者かが明確でない。セックスワーカーのようにモノ化されているロボットが被害者なのか。それとも、ロボットや売春によるモノ化の促進で苦しんでいる女性一般が被害者なのか。要するに、リチャードソンはロボットと売春との類似性を描くことはできたのだが、それらがもたらす帰結を明確に示せていない。

スパローはさらに、「同意の表象」といったより困難な問題を取り上げている。スパローに限らず、現代倫理学ではいわゆる性行為における積極的同意が中心的問題の一つであると言っても過言ではないだろう[10]。しかし、セックスロボットの使用は同意を必要としない。セックスロボットはいつでも性的行為に応じることができる。これがセックスに対するユーザーの態度に影響を与えるのであれば、セックスロボットは明らかに危険である。なぜならこの場合、セックスロボットの使用は、同意のないセックス（すなわちレイプ）のシミュレーションに他ならないからである。むろん、ロボットからセックスに対する好意的な発言をさせることによって、同意無しのセックスのシミュレーションを避けられるかもしれない。しかし、スパローが述べるように、こうした設計は、女性をいつでもセックスを望む存在として表象するため、上記したモノ化問題を悪化させる可能性がある。なお、ピーターズとハゼランジャー（Peeters and Haselager 2021）が提案したように、定期的にユーザーの行為に対する同意を拒むようなモジュールの設計は一つの解決策として考えられるかもしれない。しかし、スパローによると、この解決策にもやはり問題がある。というのも、ユーザーは簡単にロボットの拒否を無視することができるからである（Sparrow 2017：474）。仮にセックスロボットの影響で一部のユーザーが女性の拒否に何らかの性的快感を覚えるようになったとしたら、それは大きな問題だろう。それゆえスパローは、セックスロボットは女性のモノ化を促すだけでなく、同意の無いセックス、あるいは同意を拒否したうえでのセックスをシミュレートする以上、女性の拒否権を認めない、いわゆる「Rape Culture」を加速する恐れがあると主張する（Sparrow 2017：468, 475）。

言うまでもなく、ロボットとの性的行為が人間同士のセックスに対する態度に影響を及ぼすことを認めない限り、これまでの議論には疑いの余地が残る。スパロー自身が自覚しているように、この議論はゲームや映画における暴力が暴力的な行為を促しているかどうかといった議論に類似している（Sparrow 2017：470）。また、ゲームが暴力的な行為を促すかどうかに関する議論と同様に、セックスロボットのユーザーは一般人よりもレイプする確率が高いかどうかを実証することは大変困難であると考えられる（470）。しかし、スパローは、たとえセックスロボットの扱いと人間同士の関係とのつながりを実証することができなくても、セックスロボットを使用してはならないと述べる。なぜなら、レイプをシミュレートするセックスロボットの使用はそれ自体が徳に反するからである。すなわち、ただのジョークだったとしても差別的な発言は笑ってはいけないのと同様に、実際には誰もレイプしていなくても、レイプをシミュレートするような行為は非難に値するということだ（474-5）。このように、スパローは徳倫理的な観点からもセックスロボットを非難している[11]。

以上、現代社会におけるセックスロボットへの批判を紹介してきた。すでに確認したように、この議論はセックスの表象を軸とする。もちろん、他の理由からセックスロボットに反対する人もいる。たとえば、宗教的な理由に基づき、人間とロボットの関係がセックス本来の精神的な価値を欠いて

10　積極的同意モデルでは、単に拒否されていないということだけでなく、当事者が自らの意志で明示的に（たとえばセックスに）同意することが重要視される。（Halley 2016 参照）

11　ここで見落としてはならないのは、この論点がポルノグラフィの倫理的な問題に関する議論と共鳴するところがある、ということである。なるほど、C. マッキノン（MacKinnon 1993）が主張するように、ポルノグラフィにおける女性の表象はあらゆる女性を単なる「道具」あるいは「モノ」として映している以上、女性の人間としての尊厳を損害しており、批判に値すると考えられる。その意味では、実際性犯罪を加速しなくても、ポルノグラフィと同様に倫理的に許されないと考えられる。

いるということから、セックスロボットに反対する人々もいる[12]。また、セックスロボットの安全にかかわる技術的な問題もある[13]。しかし、これらのような議論は稀であり、必ずしも倫理学者の取り上げるべき問題ではないと言えよう。実際、現代社会におけるセックスロボットの反対運動を促しているのは、セックスロボットに伴うセックスの表象である。そして、次の節から見て取れるように、セックスロボットを擁護する論者もこうした表象を問題としている。

2. セックスロボットによるセックスの分配および治療的な可能性

売春は悪であり、セックスロボットは売春に例えられるというリチャードソンの主張を認めたとしても、少なくとも次の疑問が残るかもしれない。すなわち、苦痛を感じないセックスロボットは、売春業界に関わっている人間の身代わりになれるではないか、という疑問である。むろん、上述したジェンダー関係的な問題は依然として残る。しかし、それでもなお、人間に害を加えず、セックスを必要とする人に（いわば）分配できるセックスロボットは、無視できないほどの価値をもっているようにも思われる。実際、セックスロボットの使用は、セックスの分配に関わる医療的・治療的な問題に貢献できると考えている論者が少なくない。そこで、本節では、セックスロボットのメリットを取り上げて検討したい。

さて、これまでの議論を踏まえて言うと、売春の代わりになりうるものを探すより、売春という制度を排除したほうが良いと思われるかもしれない。しかし、たとえ現代社会における売春は非倫理的であると認めたとしても、セックスワークを全面的に否定することはやや難しい。R. マッケンジー（Mackenzie 2014）が指摘するように、セックスワークは多義的な概念であり、その道徳的評価は文脈によって異なる。少なくとも、認知症をはじめとする何らかの精神的な障がいを抱えている人、あるいは身体障がいを抱えている人にとって、セックスワーカーは自らの性的健康を十分に管理する上で必要不可欠であるかもしれない。現在、何らかの障がいを抱える人々向きのセックスワークが様々な形で行われている。しかし、世界各国では、何等かの障がいを抱える人々にサービスを提供するセックスワーカーは限られており、多かれ少なかれ足りていないようである（Sakairi 2016; Wotton 2020 参照）。セックスロボットは、この重大な役割を果たすことができると考えられる。

このように、セックスロボットの使用はいわゆる「セックスに関する肯定的権利（positive sexual rights）」の鍵であるとも考えられる。ディヌッチ（Di Nucci 2011）が指摘するように、セックスの肯定的権利は、ある種のパラドックスを含意している。一方では、セックスは人間のウェルビーイングにつながるような心理的効果をもたらしうる（自信が高まる、ストレスが解消される、など）。その意味では、N. ジェッカー（Jecker 2021）が提案するように、相手を見つけることに苦しむ高齢者にとっても、またはマッカーサー（McArthur 2017）が指摘するように、現代中国などのジェ

12　たとえば、ヘルツフェルト（Herzfeld 2017）が指摘するように、ロボットとのセックスは人間同士の間で見出されるスピリチュアルな意義を失ってしまうため非倫理的であると言うこともできるかもしれない。あるいは、生殖目的以外のセックスそのものが非倫理的であると主張できるかもしれない。ただし、こうした宗教に基づいた考え方はあまりにも現代の価値観から離れすぎている以上、本稿では考察しないことにする。

13　Menegus（2016）参照。本稿では、こうした技術的な問題を倫理的な課題として取り上げない。なぜなら、安全に使用できるようになれば何の倫理的な配慮を必要とせずに解決されるからである。

ンダーバランスを取れていない地域に住んでいる人にとっても、肯定的権利を認める必要がある。なぜなら、自らの良き生のためには、セックスに伴う心理的ベネフィットへのアクセス可能性がなくてはならないと考えられるからだ。しかしそれと同時に、セックスは相手を必要とする以上、ある個人のセックスに関する肯定的権利の確保は、必然的にもう一人の個人の否定的権利（つまり、セックスを拒む権利）の侵害を含意すると考えられる。そこでディヌッチは、ユーザーの性的なニーズに応じることに何ら苦痛を感じないセックスロボットを導入すれば、この矛盾を解決できると主張する（Dinucci 2011, 2017; Uszkai 2019 参照）。

むろん、前節で取り上げた表象の問題、あるいはセックスロボットが現代社会におけるジェンダー関係に及ぼすと思われる影響は依然として問題である。すでに確認したように、セックスロボットは同意を必要としない以上、現代の Rape Culture を加速していると考えられるかもしれない。さらに言えば、レイプをシミュレートするため、あるいはほかの有害な性的行為をシミュレートするためにロボットが設計されるといった事態も容易に想像できる。ただし、ここで見落としてはならないのが、こうした禁じられた行為（レイプ・未成年との性的行為）をシミュレートするロボットが予防的治療として活用される可能性である。実際、米国のロボット工学者である R. アーカンは、2014 年、子どもの形をしたセックスロボットを小児性愛者の予防的治療に使うことを提案した。すなわち、小児性愛の欲望を感じる人は、子どもの形をしたロボットと性的行為を行うことで、子どもに危害を加えずにそうした欲望を処理できるのではないか、ということだ[14]。そしてその後で、そうした欲望をセラピストなどに報告したり、ロボットの使用を減らそうと努力したりすることもできる。こうして、このような治療は、子どもを守るだけでなく、禁じられた行為への欲望をもつ人々のセラピーや治療にも役立つと考えられる。

言うまでもなく、こういった予防的治療としてセックスロボットを使うことの有用性は、簡単に実証できるものではない[15]。むしろ、そうした欲望を肯定することによって、加害の予防効果を低める可能性すら考えられる。この点を踏まえると、治療におけるセックスロボットの可能性をどのように実証できるか（あるいはそもそも実証しようとすること自体が倫理的に許されるのか）については慎重に考えなくてはならない。また、治療として使える可能性があったとしても、未成年を表象するセックスロボットは原理的に非倫理的であると言えるかもしれない。しかし、それでもなお、セックスロボットは別の何らかの形で医療的な役割を担うことができるかもしれない。たとえば、何らかの性的トラウマから回復するために自分でコントロールできるロボットを使用したり、自らの性的生活を管理できない認知症の患者がセックスロボットを使用したりする可能性を考えると、セックスロボットのベネフィットを無視することはできないだろう。こうした視点から、セックスロボットの倫理的価値を肯定的に論じることもできるかもしれない。

14　プリッグ（Prigg 2014）のような筆者はこの治療モデルを依存症の患者のメタドン治療と比較している。この比較が妥当であるかどうかに関してここでは深入りしないが、少なくともこの例えは理解可能であるように思われる。
15　これは様々な医療団体によって報告されていることである。（Cox-George その他 2018 参照）

3. 未来の予期されている諸問題

これまでのまとめから見て取れるように、現代社会においては、セックスロボットは人間の性をどのように表象しているのか、ということと、そうした表象はどんなインパクトを与えるのか、ということがすでに問題になっている。しかし、こうした二つの問いは、これからの未来に向けて我々が考えるべき重要な論点を示唆していると言える。すなわち、我々は技術的進歩を視野に入れて、これからセックスロボットとどのように関わっていくべきか、という根本問題を積極的に検討すべきである。そこで、本節では、この問題が実際の社会に影響を与える前に検討しておくべき、セックスロボットの設計に関わる二つの重要な課題を紹介しておきたい。

1）セックスロボットにおける「欺瞞」の問題

セックスロボットに限らず、近年、いわゆるソーシャルロボットにおける欺瞞の問題が活発に議論されている。「欺瞞（deception）」の問題は、意識または内的な感情などをもたないにもかかわらず、あたかも生きているかのように振る舞うことで、ユーザーの同情・共感を不正に引き出すロボットの存在可能性に出発点がある。そうした可能性があるとするならば、人間のユーザーはロボットに対し、一方的に感情（友情・恋愛感情・親しみなど）を抱くこともあるだろう。すると、ユーザーはより充実した人間関係を無視したり、感情を返すことのできないロボットに裏切られたと感じ、ショックを受けてしまったりするかもしれない。この立場の代表的論者であるS. タークル（Turkle 2011）は、現代人が相互理解および何らかの共感を含意する本来的な人間関係からますます距離を取り、ロボットや技術との付き合いに集中しすぎているという点に、ある種の危機を見出している。

上述したように、これはソーシャルロボット一般に関する問題ではあるが、その役割を果たすために人間と同じ姿であることが必要とされるセックスロボットの場合、こうした欺瞞の問題はさらに次の課題を内包するだろう[16]。すなわち、ユーザーとの会話を行う能力を有するだけでなく、外見を個々人の希望に合わせてモデル化できるロボットが、他のどんな人間よりもはるかに望ましいパートナーになる日が来るのではないか、という懸念である。現時点では、セックスロボットの会話能力は限られている。そして外見上も、セックスロボットはいわゆる「不思議の谷（uncanny valley）」から脱出できていないと言えるだろう。しかし、繰り返しになるが、セックスロボットの進歩は驚くほど早い。セックスロボットは、ユーザーの趣味趣向を学習したうえで「理想的な」会話ができるようになる可能性を有しているし、「不思議の谷」を出られたとしても、カスタマイズ可能なルックスを保つはずである。そうなった場合、実際に共感したり、相互理解を得たりすることのできる人間のパートナーよりも、内的な感情・意識をもたないセックスロボットとしか時間を過ごさないユーザーが数多く現れる可能性がある。もっと言えば、こうしたセックスロボットは人類の終わりを意味すると考えることもできるかもしれない[17]。すなわち、優れたセックスの能力と

16　本稿ではこの問題に立ち入らないが、ロボットはなるべく共感されないように設計するべきであると論じることができる。シューツ（Scheutz 2012）に従えば、そのために人間と同様な形をはじめ、ジェスチャーなどをしないようにロボットを設計すべきである。しかし、セックスロボットは実際の人間が行う性的行為を模した行為を行う必要がある以上、非人間的な形で設計することは容易でないと考えられる。

17　『The Age of Artificial Intelligence: The Documentary』というドキュメンタリーのインタビューでは、R. ウズカイはこの問題をわかりやすく解説する（15:00 ～ 21:00）。

カスタマイズされた外見をもつロボットが作られた場合、もはや人間同士の間に性的行為は起こらないことになってしまうのではないだろうか。

むろん、こうした問題意識には疑いの余地がある。たとえば、内的な感情を前提とする「本来的（authentic）」な人間関係の優位性を否定し、エージェント同士のやり取りを重視するいわゆる関係論的なアプローチを取ることができる[18]。実際、ある種の関係論的なアプローチをとるダナハー（Danaher 2017, 2020A, 2020B）は、セックスロボットの内面を問う必要は一切ないとしており、行動主義的（behaviorist）な観点を取るだけで十分だと考えている。つまり、あるロボットが恋人として振る舞えれば、それがユーザーを本当に愛しているかどうかはあまり重要ではないということだ[19]。むろん、あるロボットが十分にユーザーのニーズに応えることができなければ問題が発生する。しかし、ダナハーによれば、これは倫理の問題ではない。また、仮にある人間がロボットをパートナーとすることで充実した生活を送ることができるならば、そもそも確認不可能なロボットの内面が本当にどうなっているかを問う必要はない、とダナハーは主張する。

セックスロボットはより充実した人間関係からユーザーの目をそらす、という疑念に対するもう一つの根本的な問題として、人間同士の関係と人間とロボットの関係の違いが挙げられる。先ほど述べたように、セックスロボットは人間のユーザーの好みを反映することができるかもしれない。また、ユーザーが求める性的行為を拒否しないという意味で、理想的な恋人に思われるかもしれない。しかし、こうしてユーザーの欲望を一方的に満たすだけのセックスロボットは、恋人になくてはならない「抵抗の可能性」を欠いていると考えられる。実際、我々が一人の人間と付き合うとき、我々の求めている性的行為が拒否されることもあるし、我々の意見や世界観が否定されることもある。しかし、これは必ずしも悪いことではない。むしろ、自分とは異なる他者と関わることによって新しい視点を得ることもできるし、予想しなかった刺激を与えられることもあるだろう。この点に鑑みると、セックスロボットは人間にとって十全な意味で恋人になることはできないように思われる。つまり、人間同士の恋愛・セックスは人間とロボットのセックスとは異なると考えられるし、ロボットによっては与えられない刺激をもたらすであろう人間同士の付き合いが消えるとは考え難い[20]。

2）セックスロボットにおける奴隷制問題

これまでの議論では、セックスロボットは意識、あるいは何らかの内的な感情および心的な状態をもたないという前提の下で議論を進めてきた。しかし、この前提には疑いの余地があると言わざるを得ない。なぜなら、ロボットの内面を見て、意識があるか否かを確認することはできないからである。また、たとえ現代社会におけるロボットは（おそらく）意識をもっていないとしても、意

18　ここで見落としてはならないのが、こうした「関係論的なアプローチ」とは一つの統一化された立場ではなく、内的な意識よりもエージェント同士のインタラクションを重要視する方向性を指す、ということである。（Coeckelbergh 2011 参照）

19　もちろん、この主張を疑う余地がある。序論で言及したジェームズは、人間の虚栄心を理由に、本当は何も感じていない Automatic Sweetheart で満足できる人はいないと主張している。

20　たとえば、「Real Doll」の CEO である M. McMullen が指摘するように、セックスロボットは必ずユーザーの命令に従うため、人間同士の付き合いにあるはずのいわゆる「緊張感」を再現することができない。それゆえに、McMullen はセックスロボット・セックスドールが「退屈」であると断言する。（Gurley 2015 参照）

識を備えたロボットの存在は少なくとも想像できる。しかし、そうしたロボットがいつから、どのように意識をもつようになるかを予測することは極めて困難である[21]。そこで、次の疑問が湧き上がる。我々は、人間の性的欲望を満たすためだけの奴隷のような存在を生み出しているのではないだろうか。あるいは、我々はもうすでに奴隷を生み出してしまっているのではないだろうか。

むろん、この問いは大げさに思われるかもしれないし、意識をもつロボットは甚だ遠い未来の問題に過ぎないと考えられるかもしれない[22]。しかし、この問いは非常に重要な別の問題に関連している。たとえば、意識をもたないと思われてはいるものの、あるロボットが「No」という言葉を学習し、ユーザーの要求を拒否するようになったケースを想定しよう。あるいは、自分が「No」と言ったのにユーザーにレイプされた、と明確に伝えることができるロボットが現れたというケースも考えられる。そうした場合は、セックスロボットの主張を無視してもよいのだろうか。むろん、すでに述べたように、ロボットには主観的な意識が欠けている以上、快苦を感じないという点で、ロボットの発話を無視してもよいと言えるかもしれない。ただし、そこで問題になるのは、我々がどういった枠組みの下でセックスロボットに意識がないと判定できるか、という点である。実際、グンケル（Gunkel 2018）が指摘するように、歴史上、あるマイノリティグループが本当に何も感じない人間以下の存在と決めつけられることは多々あった[23]。ゆえに、なぜロボットに意識がないものと判断してその発言を無視することが許されるのか、ということが問題になるだろう。

言うまでもなく、これも疑わしい議論ではある。このようなコミュニケーション能力を備えるロボットが果たして現れるかどうかは明らかでない。また、人間や動物のように意識をもつ生き物と異なるロボットは、あくまで何も感じないモノとみなすことができるかもしれない。しかし、いずれにせよ、これからロボットを設計していく中で、ロボットをどれほど人間らしく作るかについては検討する必要があるだろう。限りなく人間に近い、あたかも意識をもつかのように振る舞うロボットを設計することができたとしても、そうしたロボットは作らないほうが良いと言えるかもしれない。

まとめ

現時点では、セックスロボットの能力は限られているが、本稿では、検討すべき様々な倫理的問題を明らかにした。現時点でとりわけ問題になっているのは、セックスロボットの表象である。セックスロボットは、女性の身体を非現実的な形で表象するだけでなく、セックスにおける同意を軽視した態度を促すと考えられるかもしれない。それだけでなく、徳倫理的な観点から考えれば、そもそも同意のないセックスの表象を楽しむべきではないとも考えられるだろう。このように、リチャー

21　むろん、確認できると述べる論者もいる。たとえば、S. シュナイダー（Schneider 2019）が指摘するように、いわゆる「チップテスト（Chip Test）」で非生物的な物質が意識を生み出せるかどうかを試せるかもしれない。「チップテスト」とは、意識を備えているはずの主体の細胞をコンピューターチップと入れ替えたうえで、その主体にまだ意識があるかどうかを聞く、といった単純なテストである。言うまでもなく、シュナイダーのチップテストの正当性を疑う余地がある。ただし、ここで見落としてはならないのが、こうしたテストが決定的ではないとしても、ロボット意識の有無の判断基準として使えるかもしれない、ということである。

22　あるいは、ブライソン（Bryson 2010）のように「ロボットは奴隷であるべきだ」と言えるかもしれない。しかし、ブライソンは意識をもっているロボットを想定しない以上、彼女の議論は本節の目的とずれるためここでは取り上げない。

23　（Gunkel 2018 115-120; Coeckelbergh 2012：16-20 参照）

ドソンのような論者は、セックスロボットは現代社会の女性たちにとって売春と同程度に有害であるとして、この種のロボットに反対している。

とはいえ、将来に向けて、セックスロボットの治療的な使い方も真剣に考察するべきであると言えるだろう。現時点では、経験的なデータがないためその効果については何とも言えないし、セックスロボットの治療的な効果を実証することも容易ではないが、それでもなお、セックスロボットは様々な医療的用途に活用できるかもしれない。これはもっぱら、何らかの障がいあるいはトラウマを抱えている患者に当てはまることである。しかし、さらに言えば、セックスロボットは誰もが一定程度の快適な性生活を享受するための分配手段にもなれるかもしれない。セックスの相手を見つけることに困難を感じる人にとって、セックスロボットはウェルビーイングを高めると考えられるだろう。

最後に、セックスロボットが今の限界（いわゆる不気味の谷）から出てしまったときには、様々な問題が起こりうる。ますます人間に近づいてくるであろうセックスロボットが、いつか単なる奴隷にされてしまう可能性はないのだろうか。この類の問題がどのように生じうるか、そしてどれほど現実的なのかについてはさらなる検討が必要だろう。そうしなければ、我々倫理学者は現代ロボット工学の進歩についていけなくなってしまうのだろう。

※ 本稿の内容は、*Journal of Applied Ethics and Philosophy vol.13*, 2022（Center for Applied Ethics and Philosophy, Hokkaido University）に掲載された研究ノート Richard Stone “Living in the Age of the Automatic Sweetheart: A Brief Survey on the Ethics of Sexual Robotics” にもとづいている。

また、本稿は「JST-RISTEX「科学技術の倫理的・法制度的・社会的課題（ELSI）への包括的実践研究開発プログラム「人工主体の創出に伴う倫理的諸問題を分析・討議するプラットフォームの構築に向けた企画調査」」による研究成果の一部である。

参考文献

Bryson, J. J.（2010）“Robots should be Slaves,” *Close Engagements with Artificial Companions: Key Social, Psychological, Ethical and Design Issues*, 8, pp. 63-74.

Coeckelbergh, M.（2011）“Are Emotional Robots Deceptive?” *IEEE Transactions on Affective Computing*, 3（4）, pp. 388-393.

――――（2012）*Growing Moral Relations: Critique of Moral Status Ascription*, Palgrave Macmillan.

Cox-George, C., & Bewley, S.（2018）“I, Sex Robot: the Health Implications of the Sex Robot Industry,” *BMJ Sexual and Reproductive Health*, pp. 161-164.

Danaher, J., Earp, B., & Sandberg, A.（2017）“Should We Campaign Against Sex Robots?” in *Robot sex: Social and Ethical Implications* ed. Danaher and McArthur, MIT Press: pp. 47-72.

Danaher, J.（2017）“Should We be Thinking about Sex Robots?” in *Robot Sex: Social and Ethical Implications* ed. Danaher and McArthur, MIT Press, pp. 3-14.

――――（2018）“Regulating Child Sex Robots: Restriction or Experimentation?” *Medical Law Review*, 27（4）, pp. 553-575.

――――（2020）“Sexuality,” in *The Oxford Handbook of the Ethics of AI*, pp. 403-420.

Devlin, K.（2018）*Turned on: Science, Sex and Robots*, Bloomsbury Publishing.

Di Nucci, E.（2011）“Sexual Rights and Disability,” *Journal of Medical Ethics*, 37（3）, pp. 158-161.

――――（2017）“Sex Robots and the Rights of the Disabled,” in *Robot sex: Social and ethical*

implications ed. Danaher and McArthur, MIT Press, pp. 73-88.

Elder, J（2020）“AI-Powered Sex Robots are Selling Well during Lockdown – Which worries Some Experts, Who say That They can Introduce Some Surprising Regulatory Problems,” *Business Insider*. https://www.businessinsider.com/ai-sex-robots-are-selling-well-realdoll-regulated-2020-6 Accessed December 20, 2020

Eskens, R.（2017）“Is Sex With Robots Rape?” *Journal of Practical Ethics*, 5（2）.

Gouvela, S.（2020）“The Age of Artificial Intelligence: The Documentary,” Released December 2, 2020. https://www.youtube.com/watch?v=zMrt6ALaio0 Accessed August 1, 2021.

Gunkel, D. J.（2018）*Robot rights*, MIT Press.

Gurley, G.（2015）. “Is this the dawn of the sexbots（NSFW）,” *Vanity Fair*, https://www.vanityfair.com/culture/2015/04/sexbots-realdoll-sex-toys Accessed December 21, 2020

Halley, J.（2016）“The move to affirmative consent,” *Signs: Journal of Women in Culture and Society*, 42.1, pp. 257-279.

Herzfeld, N. L.（2017）“Religious Perspective on Sex with Robots,” in *Robot Sex: Social and Ethical implications* ed. Danaher and McArthur, MIT Press, pp. 91-102.

James, W.（1908/1955）“The Pragmatist Account of Truth and its Misunderstanders,” in *Pragmatism*, New York: Meridian.

Jecker, N. S.（2021）“Nothing to be Ashamed of: Sex Robots for Older Adults with Disabilities,” *Journal of Medical Ethics*, 47（1）, pp. 26-32.

Levy, D.（2009）*Love and Sex with Robots: The Evolution of Human-Robot relationships*, New York.

Mackenzie, R.（2014）“Sexbots: Replacements for Sex Workers? Ethicolegal Constraints on the Creation of Sentient beings for Utilitarian Purposes,” in *Proceedings of Advances in Computer Entertainment 2014 ACE'* 14 Workshops（article 8）, New York: ACM. doi:10.1145/2693787.2693789

MacKinnon, C. A.（1993）*Only words*, Harvard University Press.

McArthur, N.（2017）“The Case for Sexbots,” in *Robot sex: Social and Ethical Implications* ed. Danaher and McArthur, MIT Press, pp. 31-46.

Menegus, B.（2016）“Sex Robots May Literally Fuck us to Death,” *Gizmodo*, Published December 19, 2016. https://gizmodo.com/sex-robots-may-literally-fuck-us-to-death-1790276123?utm_campaign=socialflow_gizmodo_twitter&utm_source=gizmodo_twitter&utm_medium=socialflow Accessed August 1, 2021.

Peeters, A., & Haselager, P.（2021）“Designing Virtuous Sex Robots,” *International Journal of Social Robotics*, 13（1）, pp. 55-66.

Prigg, M.（2014）“Could Child Sex Robots Be Used to Treat Paedophiles? Researchers say sexbots are inevitable – and could be used like ‘methadone for drugs addicts’,” Daily Mail Online. https://www.dailymail.co.uk/sciencetech/article-2695010/Could-child-sex-robots-used-treat-paedophiles-Researchers-say-sexbots-inevitable-used-like-methadone-drug-addicts.html Accessed December 18, 2020.

Richardson, K.（2016）“The Asymmetrical ‘Relationship’ Parallels between Prostitution and the Development of Sex Robots,” *ACM SIGCAS Computers and Society*, 45（3）, pp. 290-293.

Sakairi, E.（2020）“Medicalized Pleasure and Silenced Desire: Sexuality of People with physical Disabilities,” *Sexuality and Disability*, 38（1）, pp. 41-56.

Scheutz, M.（2012）“The Inherent Dangers of Unidirectional Emotional bonds between humans and Social Robots,”in *Robot Ethics: The Ethical and Social Implications*, ed. Patrick Lin, Keith Abney, and George A. Bekey, MIT Press, pp. 205-221.

Schneider, S.（2019）*Artificial you: AI and the Future of your Mind,* Princeton University Press.

Sparrow, R.（2017）“ ‘Robots, Rape, and Representation’ ,”*International Journal of Social Roboticism,* 4, pp. 465-477.

――――（2021）“Sex robot fantasies,”*Journal of Medical Ethics*, 47（1）, pp. 33-34.

Stone, R.（2022）“Living in the Age of the Automatic Sweetheart: A Brief Survey on the Ethics of Sexual Robotics,”*Journal of Applied Ethics and Philosophy,* 13, pp. 21-30.

Turkle, S.（2011）*Alone Together,* Basic Books.

Uszkai, R.（2019）“A Theory of（Sexual）Justice,”*Információs Társadalom* XIX, 4, pp. 133-146.

Wotton, R.（2020）“Paid Sexual Services for People with Disability,”in *The Routledge Handbook of Disability and Sexuality*.

神崎宣次（2021）「これからのロボット研究のための倫理」『日本ロボット学会誌』、vol. 39 No. 1、18-21 頁

西條玲奈（2013）「性愛の対象としてのロボットをめぐる社会状況と倫理的懸念」『社会と倫理』、28 号、37-49 頁

研究ノート

中国国内の AI 倫理研究の現状について

侯乃禎（北海道大学大学院文学院）

要旨

本稿は、2017 年 7 月から 2021 年 1 月までの中国における AI 倫理研究に関する注目すべき話題や研究を整理して紹介するものである。まず「次世代人工知能発展計画」以来の中国における AI 倫理研究の背景を紹介する。とりわけ 2019 年に中国人工知能学会が設立した人工知能倫理道徳委員会の主任である陳小平についての記事を紹介する。次に、AI の主体性という話題をめぐるいくつかの中国の研究者たちの主張をまとめる。それは、今現在の AI が主体性を持つかどうかということについての研究と、将来の AI が主体性を持つ状況およびその状況に伴う危険性についての議論に分かれている。最後に、AI 倫理に関する中国の特色ある内容として、中国思想の研究者たちの AI 倫理についての議論を取り上げる。その節には、刘紀璐の「儒家のロボット倫理」という論文をめぐる議論と、AI と人の関係という話題をめぐる中国思想の研究者たちの議論が含まれている。それにより、まだ十分に日本に発信されていない中国国内の AI 倫理研究の動向を示す。

Abstract

This paper presents the notable researches and topics on AI ethics in China from July 2017 to January 2021. First, it introduces the background of China's AI ethics research since the *New Generation Artificial Intelligence Development Plan*. This part especially organizes the relevant news of Chen Xiaoping, who is the Director of AI Ethics Committee which established by the Chinese Association for Artificial Intelligence in 2019. Second, it sorts out the claims of some Chinese researchers on the subjectivity of AI. Two aspects were mentioned in this part: The research on whether AI has subjectivity today, and the discussion on the issues accompany with AI's subjectivity in the future. Finally, it takes up the discussion of AI ethics with Chinese characteristic by researchers of Chinese thought. This part contains not only the discussion of JeeLoo Liu's *Confucian Robotic Ethics*, but also the discussion of the relationship between humans and AI by researchers with Chinese thought. In this way, the paper shows the AI ethics research trends in China that has not yet been fully known by Japan.

はじめに

2017 年 7 月 8 日に公表された「次世代人工知能発展計画」（中:《新一代人工智能发展规划》）で、中国において AI 技術を発展させるために倫理規範が AI 戦略の重要な「保障措置」（中: 保障措施）であると、中国国務院は強調している。それ以来、中国における AI 倫理に関する研究は盛んになってきた。本稿は、膨大な数のそれらの研究に対する統計的・包括的な調査ではなく、2017 年 7 月から 2021 年 1 月までの中国における AI 倫理研究に関する注目すべき話題や研究を整理して紹介するものである [1]。

まず「次世代人工知能発展計画」に関連する AI 倫理研究の背景と、人工知能倫理道徳委員会の主任である陳小平についての記事をまとめることによって、2017 年 7 月から 2021 年 1 月までの中国における AI 倫理研究の背景を紹介する。次に、その時期における AI 倫理研究についての三つの話題を取り上げて整理する。一つ目は、AI が人間のような主体性を持った存在であるかという問題に関する議論である。この AI 倫理研究における一般的な問題に対して、中国の研究者はどのような主張をするのかということについて整理する。二つ目と三つ目は、中国研究界における JeeLoo Liu（刘紀璐）の「儒家のロボット倫理」という論文をめぐる議論と、AI と人間の関係に関する話題である。とりわけ、これらの話題についての中国思想の研究者の論文を中心に取り上げる。これにより、AI 倫理研究についての中国研究者の独創的な主張を紹介し、まだ十分に日本に発信されていない中国国内の AI 倫理研究の動向を示す。

1. 中国における AI 倫理研究の背景

「次世代人工知能発展計画」には、「倫理」という言葉が 15 回ほど出てくる。その中では、AI 技術を発展させるための法律・法規と倫理規範が、中国において AI 戦略の重要な「保障措置」として明言されている。中国電子技術標準化研究院の「人工知能倫理リスク分析報告」によれば、「次世代人工知能発展計画」で倫理をそれほど強調するのは、AI が社会に与える倫理的影響に注意を払うためだけではなく、倫理体系および倫理規範を制定して AI の安全性と信頼性、そして制御可能な発展を確保するためである。

また、中国電子技術標準化研究院は「人工知能標準化白書 2018」、「人工知能標準化白書 2019」、「人工知能倫理リスク分析報告」において、AI 技術の安全性と AI 倫理の問題を重要視している。とりわけ「人工知能倫理リスク分析報告」では、アルゴリズムに関するプライバシー保護、差別、濫用、解釈可能性、責任といった具体的な問題に注目している。また、その報告は、アシロマ AI 原則、IEEE が提唱した AI 倫理綱領、日本人工知能学会の『人工知能学会倫理指針』などを参照した上で、「人間根本利益原則」と「責任原則」を提唱している。「人間根本利益原則」は、AI 研究は人間の利益を実現するためのものであるということを強調する。「責任原則」は、AI のアルゴリズムの説明可能性や検証可能性や予測可能性を強調し、そして AI を設計する際の責任の帰属先を強調する。

1　本稿が取り扱うのは、主に「中国学術文献オンラインサービス（China National Knowledge Infrastructure : CNKI）」に収録されている AI 倫理についての中国語の文献と、インターネットで公開されている中国語の新聞記事である。

さらに、2019 年 6 月 17 日に中国国家次世代人工知能ガバナンス専門委員会は、「次世代人工知能ガバナンス原則――責任ある人工知能の発展」を公表した[2]。「ガバナンス原則」には、責任ある人工知能の発展というテーマをめぐり、調和・友好、公平・公正、包摂・共有、プライバシー保護、セキュリティコントロール、共同責任、開放・協力、俊敏なガバナンスの 8 項目が記されている[3]。そして、「ガバナンス原則」に呼応して 2019 年 8 月 29 日に上海コンピューター青年工作委員会は、2019 年の世界人工知能大会で「中国青年科学者 2019 人工知能イノベーションガバナンス上海宣言」[4] を発表した。「上海宣言」には、プライバシー保護、多様性・公平、技術の安定性、アルゴリズム透明性、説明責任、環境に優しいシステム、人間の幸福への利益、人道的アプローチの 8 項目が記されている。

中国の企業も、AI 倫理の問題を研究している。例えば、2019 年 7 月 11 日にテンセント研究院とテンセント AI Lab は、「知能時代の技術倫理観――デジタル社会の信頼を再構築する」という研究報告において、技術信頼、個人幸福、社会持続可能という三つのレベルを含む AI 研究の倫理観を示している。

そのほか、中国人工知能学会は、2019 年 5 月 26 日に人工知能倫理委員会を設立した[5]。現在、人工知能倫理委員会の具体的な成果はまだ確認できていないが、今後の中国の AI 倫理についての重要な研究センターになると思われる。そのセンターの主任である陳小平に関する 2019 年 12 月 11 日の記事がある。陳小平個人ないし人工知能倫理委員会の立場を示すために、以下ではその内容を整理する。

その記事において陳は、今の中国社会における AI 倫理研究に対して憂慮するべきことが主に三つあると述べ、それらに対する彼の意見を提示している。

一つ目は、AI 技術の発展によって、将来 AI が逆に人類を統治するという「技術の制御不能」である。陳によれば、「AI が人類を統治する」未来は、AI 研究の専門家が言ったことではないし、その情況が必ず生じるという確実な証拠もないので、短期的にはありえないと思われる。長期的には、人間はそのリスクを直視する必要がある。しかしその不確定な未来に対して、最も重要なのは主観的な憶測ではなく、それに対する客観的な科学研究である。

二つ目は、AI 技術の未熟および倫理的制御ができないといった原因によって使用者に危害が及ぶ「技術の誤用」である。例えば、データプライバシーや安全性や公平性などの問題がある。これらの問題に対して陳は、現在の科学界、産業界、および規制当局の AI 倫理の問題に対する理解がまだ不十分であり、対応する規範がないので、実際に技術倫理の問題が発生していると述べている。彼によれば、一般に歴史上、新たな産業が生まれた当初は様々な問題が生じる。これらの問題は技術と産業のさらなる発展によってのみ解消できる。問題があるからといって産業の発展を遅らせる

2　以下「ガバナンス原則」と略記。その英語版は以下のサイトで公開している。Chinadaily.com.cn. Governance Principles for the New Generation Artificial Intelligence--Developing Responsible Artificial Intelligence. URL=<http://www.chinadaily.com.cn/a/201906/17/WS5d07486ba3103dbf14328ab7.html>（最終閲覧日 2021 年 3 月 24 日）

3　この一文は以下のウェブサイトの日本語の記事を参照した。中国网、中国が次世代人工知能ガバナンス原則を発表、URL=<http://japanese.china.org.cn/business/txt/2019-06/19/content_74899392.htm>（最終閲覧日 2021 年 3 月 24 日）

4　以下、「上海宣言」と略記。なお、「上海宣言」は技術本位であり、AI に関する倫理問題を技術問題に還元する還元主義である、という批判がある（杨 . 2020: 138）。

5　中国の人工知能倫理委員会の設立日について、以下のウェブサイトの記事を参照した。博古睿研究院，全球视野下的人工智能伦理论坛 " 成功举办，URL=<https://www.berggruen.org.cn/activity/12>（最終閲覧日 2021 年 3 月 24 日）

のは、ただ問題の解決を遅らせるだけである。

三つ目は、AI の普遍的な応用に伴う深刻な社会的影響である。例えば、AI 技術を応用することによって多くの人の仕事が奪われ、短期間内に再就職できなくなるという「応用の制御不能」である。これについて、今の多くの企業における大量の単純作業、機械的な仕事は人間性に反するため、そのような仕事に従事する人手が足りなくなってきたという状況が存在すると陳は述べている。こうした問題および失業の問題は、産業転移と生産効率の向上によって生じた問題である。従って今の状況に対して、AI 技術を応用するのはもはや避けられないことである。

陳によると、AI 倫理研究が目指すのは、積極的に AI 技術および非技術的革新を促進し、既存のおよび新たな社会的、倫理的問題を解決し、より良い世界に向かって進む未来である。ゆえに、AI 倫理の「基本的な使命」は、「人工知能に倫理的なサポートを提供して、人間の福祉との調和と共存増進することである」。それは、AI が「良いことをするように促す」ことと、AI が「悪いことをするのを防ぐ」という二つの含意を含んでいるのである。

AI 倫理の「基本的な使命」に従って、陳は「人工知能倫理の全景動態観（中： 人工智能伦理的全景动态观）」という人工知能倫理の研究機構を設立するための指導原則を述べている。それはまず、指導原則では、上述した AI 倫理の基本的な使命に基づき、AI の倫理原則と、研究および施行細則を策定することが求められているが、これらの原則や細則の策定によって目論まれているのは、「研究」、「応用」、「需要」のループを形成することといえる。このとき、ループは次のように説明できるだろう。すなわち、ループとは、原則や細則に基づいた「研究」を積み重ね、「研究」で得られた成果を商業化するという「応用」を実現し、「応用」から新たな社会的「需要」が誘発され…という円環運動を指す。ここでの「研究」については、科学技術の研究部門と倫理の研究部門とを区別して設立することが必要であると陳は述べている。前者は、倫理的規定に符合する形で技術を研究する部門であり、後者は倫理的規定の策定および改善に従事する部門である [6]。

2. 主体性について

上記のような背景のもとで、2017 年後半から中国における AI 倫理を主題とする多くの論文が発表されてきた。その中には、AI 倫理の一般的な話題に関する研究があり、中国的な特色のある研究も存在する。この節では、主体性という AI 倫理研究の一般的な話題を手がかりにして、中国におけるいくつかの論文の内容を整理する。まず、今現在の AI が人間のような主体性を持った存在であるのかという問いについて、以下の研究を紹介する。

段偉文（2017）は、今の AI によって提示された「主体性」は、意志的能動性、自己意識および自由意志に基づくのではなく、ただ機能的な模倣であると述べている。彼によれば、そのような AI はただ、人為的な設計によって人のように振る舞わせるものなので、それを「疑似主体」と呼ぶべきである。また、彼は AI と人間は対立するという二元的な関係を批判し、コントロール（C）

6　2020 年 9 月に刊行された論文（陳 . 2020: 86-87）において、陳は上述の「人工知能倫理ダイナミクス体系」に基づいて「人工知能倫理体系構造」を提出している。そのなかでは、「研究」に関して、科学技術の研究部門と倫理の研究部門を区別して設立することは言及されず、経済的利益だけを重視する「伝統イノベーション」から、経済的利益と社会的利益の両方を重視し、異なる文化の伝統を包容できる「公義イノベーション」への移行を重視することが述べられている。

- AI（A）およびその設計者（D）- 一般的なユーザー（U）という複合的な関係こそが AI によって引き起こされる倫理的関係の基本モデルであると主張している。このモデルに基づいて彼は、責任あるイノベーション（中：负责任的创新）と主体の権利保護を強調している。責任あるイノベーションとは、AI 技術者は設計する際に一般的なユーザーだけではなく、社会全体ないし全人類に対する責任を考えるべきだ、ということである。主体の権利保護とは、人工的なエージェントが引き起こす問題の責任を明確にするためには、エージェントの制御の責任と設計の責任とを区別すべきだ、ということである。

刘永安（2021）は、意識の問題は人工的道徳的行為者（artificial moral agent）[7] の可能性の決定的な条件であると指摘した上で、現在の AI は道徳的主体にとって重要な現象学的意識欠いているので、まだ道徳的主体ではないと述べている。彼によれば、まず、AI は人のような生物的有機体ではないので、現象学的意識を生み出す実体的な基礎を欠いている。さらに、精神や意識を表現するような AI の振る舞いは単に機能的であり、AI の外に存在する三人称的視点からしか観察できない。人間の意識と比較すると、AI は自己反省の一人称視点および感情移入という道徳的主体になるために極めて重要なものが欠けている。また、感情は人の現象学的意識および道徳意識にとっては不可欠なものなのに、それを AI は備えていないと彼は論じている。

戴益斌（2020）は、道徳的主体という点では、AI（知能ロボット）は心的状態を持てないので道徳的主体にはなれないが、道徳的被行為者（moral patient）という点では、道徳的被行為者の地位を取得する可能性があると述べている。彼の言う道徳的被行為者のモデルは、レヴィナスの「他者の顔」の学説に基づいて展開されたものである。すなわち、AI をレヴィナスの言ったようなわれわれの自己が依存する「他者」としてではなく、われわれの自己に依存する「準他者」として考えることによって、AI の道徳的被行為者の地位を説明できると彼は述べている。しかし、それは AI が言語を習得した上で、積極的に人と向き合うことができる場合（例えば、ケアロボットが老人を看護する場合）、または人が AI と有意義な関係を持っている場合（例えば、言語を習得した知能ロボットに人間の情が移る場合）のみである。

さて、上述では現在の技術に基づいて、AI が人間のように主体性を持った存在であるのかということについての研究を整理したが、将来の AI は主体性を持つ状況およびその状況に伴う危険性についても、いくつかの論文が存在する。以下ではそれらを紹介する。

陳小平は中国科学報の記事（2020）で、中国科学技術大学が開発した相互コミュニケーション型ロボット「佳佳」の実験において、一種の新たな人間の経験が出現したと述べている。その実験の被験者たちは、「佳佳」と人を混同するのではなく、「佳佳」は人でも物でもないことをはっきりと感じた。この事実に対して陳は、「非人間的非物質的」（中：非人，非物）の三番目の存在物が出現する可能性があると述べている。また、同年 9 月の論文で陳は、近い将来、感情的なインタラクティブロボットのような特定の AI 製品は、一部の大衆に受け入れられる可能性があると言っている。この現象が出現する原因は、「科学や哲学の仮定とは異なり、人々は通常、ロボットによって示される感情が実際の人の感情であるかどうかを気にしないし、人とロボットの感情の本質的な違

7　道徳的行為者（moral agent）の中国語訳は、「道德行为者」（道徳的行為者）のほかに、「道德主体」（道徳的主体）もある。例えば、本文が扱う戴益斌（2020）と刘紀璐（2018）の論文には、「道德主体」という訳が使われている。

いを注意深く区別しない」（陳．2020：86-87）ことにあると彼は述べている。本質的な違いにあまり関心を払わないといった傾向は、専門家にとってとうてい支持できるものではないが、近い将来、この傾向を持つ人をわれわれは受け入れるのか、あるいは傾向を正すのかと陳は疑問を呈している。

杜厳勇（2016）は、AI 技術の安全性を確保するために、AI 製品の倫理を設計すること、AI 技術の適用範囲を限定すること、および AI の知能レベルを限定することを提案している。なぜなら、たとえスティーヴン・ホーキングの言ったような AI が「人間を越える」状況がなくとも、AI 技術の発展に伴う様々な予測できないリスクは存在するからである。

王天恩（2020）は、ニック・ボストロムの研究に基づいて、AI 技術の発展は人間全体の存在に壊滅的な脅威をもたらす可能性があり、それは人間にとっての実存的リスク（existential risk）であると指摘している。AI による実存的リスクに対して、AI に道徳的規定を内蔵させることにも、AI に人間の価値観を植え付けることにも、理論上の困難と実践上の困難が存在する。例えば、われわれ個人はどのような価値観を持つべきか、という問題はすでに十分複雑であり、AI にどのような人間の価値観を植え付けるべきかという問題はより一層困難である。王は、将来のスーパー AI を利用することでこうした難題に対応できると考えている。そして彼によれば、AI 技術に伴う実存的リスクは、伝統的倫理をそのまま適応することによって解決できる問題ではなく、世界を構築する創世倫理の問題である。なぜなら、現在のわれわれはまさに神のキャラクターを演じて人工物の世界を構築しているからである。従って、われわれは自分を創造主のような次元で、全面的に AI 技術の開発およびそれに伴う責任と結果を考えるべきであると彼は提案している。

趙汀阳（2018）は、AI が近い将来に引き起こす危険と遠い将来に引き起こす致命的な危険を分析している。彼によると、近い将来に発生する危険は三つある。一つ目は自動運転車のパラドックスである。それは、完全な自律性を持つ自動運転車が交通法規に違反する歩行者に遭遇した場合、自動運転車は乗客と歩行者のどちらを保護するのか、という問題である。その要点は、乗客と歩行者の両方を保護できるような自動運転車の技術がなければ、自動運転車のためにどのようなルールを選択してもわれわれが満足できないということである。二つ目は AI 技術に伴う失業問題である。未来の AI が多くの富を生み出し、多くの人間労働が不要となるという状況になれば、人は自主的に労働したくても何もすることがなくなってしまう。それは人生の意義の消失を意味する。三つ目は、人が他人を必要としない状況および AI 武器の出現のことである。さらに、遠い将来、言語を発明することができ、システム全体を反省する能力を持つスーパー AI が出現すれば、それはまさに神のような存在であり、そしてそれは人間にとっての致命的な危険であると彼は言っている。なぜなら、そのとき、世界における人間の地位が失われるからである。その危機を防止するために、彼は AI に自滅のプログラムを設置することを提案している。

韓敏と趙海明（2020）は、もしシンギュラリティを突破した強い AI が出現すれば、そのとき AI と人間との関係は、「主体と客体」の関係ではなく、「主体と主体」という間主体的関係になると考えている。また、その未来においては、AI と人間との間主体的関係は、インターネットだけではなく、完全にデジタル化されたバーチャル空間に通じて現れると彼らは予想している。

3. 儒家のロボット倫理をめぐる議論

中国では中国思想、とりわけ儒家思想の研究者が AI 倫理を考察した研究がいくつか出てきている。本節は、JeeLoo Liu（刘紀璐）の「儒家のロボット倫理」（Confucian Robotic Ethics）という論文をめぐる議論を紹介する。

Liu の「儒家のロボット倫理」はもともと 2017 年に英語で発表されたものが、2018 年に中国語に翻訳されて学術雑誌に掲載された後、それを批判する中国思想の研究者の論文が現れた。Liu はこの論文において主にトロッコ問題を通じて、アシモフのロボット工学三原則、カント倫理学、功利主義、儒家倫理を、人工的道徳的行為者であるロボットの倫理原則としての優劣という観点から比較した。具体的な内容は以下の通りである。

アシモフのロボット工学三原則：

Liu は、アシモフのロボット工学三原則の複雑な倫理的シナリオに対処する上での不十分さを指摘している。例えば、ロボット工学三原則に従うロボットは、ある人を助けるために他の人を犠牲にするようなトロッコ問題に遭遇した場合、何を選択しても必ず人間を犠牲にするので、「ロボットは人間に危害を加えてはならない。また、その危険を看過することによって、人間に危害を及ぼしてはならない」[8] というアシモフの第一原則に違反する。その場合、ロボットは行動不可能な状態になってしまう可能性がある。

カント倫理：

Liu は、カント倫理学に基づいて以下の二つのロボットの原則を述べている。

DR1：ロボットは、選択されたオプションが原則として他のロボットの普遍的な法則となるような仕方でのみ行為する必要がある。

DR2：ロボットは、人間性を単なる手段としてではなく、常に目的として扱うように行為しなければならない。

それから Liu は、このようなロボットの倫理原則の問題点述べている。

「DR1」について、それは標準的なトロッコ問題（Foot. 1967：2）に対して明確な解決案を出すことができないので、トロッコ問題のような選択に向かうときロボットは何も行動しない可能性がある。

「DR2」について、カントの定言命法は人という自律的な理性的主体を前提にしているが、ロボットはそれら自身のために立法することはない同時に、自由意志も持たないので、自律的な理性的主体ではない。また、「DR2」に従う道徳的行為者であるロボットを造るわれわれの行為は、道徳的行為者を目的ではなく手段として扱うので、カント倫理によればそれは不道徳である。

8　アシモフのロボット工学の第一原則の日本語訳は、以下の訳本から引用した。アイザック・アシモフ『われはロボット』、小尾芙佐訳、ハヤカワ文庫、1983 年、5 頁。

功利主義：

Liu は、行為功利主義の内容を抽出することによって次のロボットの原則を述べている。

UR：ロボットは、利用可能な一連の行動の結果を考慮する際に、関係するすべての人間に対して最大の利益を生み出すか、より大きな害を防ぐかのいずれかの行為を選択する必要がある。

この「UR」について Liu は以下の二つの問題点を指摘している。

「UR」に基づいて設計された自動運転車は、より多くの損害を避けるために、車自体と車の乗客を犠牲にする可能性がある。そのとき、たしかに大衆はこの種の車がもたらした結果を正義として肯定するかもしれない。しかし、乗客の安全が保証されない限り、おそらく大衆はこの種の自動運転車を買わないだろう（Bonnefon et al. 2016：1574）。

そして功利主義は全体の利益のために個人の利益を侵害する、というよく指摘された問題が「UR」にも存在している。さらに、ロボットには人間のような他人を犠牲にすることに対する心理的障害はないので、「UR」に従うロボットは、躊躇なく全体の利益のために個人を犠牲にする行動を取ると思われる。

儒家倫理：

Liu は、『論語』における忠、恕、仁という三つの美徳に基づいて次の三つのロボットの原則を述べている。

CR1：ロボットは何よりもまず、割り当てられた役割を果たす必要がある（忠 ; loyalty）。

CR2：他のオプションが利用可能な場合、ロボットは他の人間に最大の不快感または最も低い選好を与えるような行動をとるべきではない（恕 ; reciprocity）。

CR3：ロボットは、「CR1」および「CR2」に違反しない限り、道徳的改善を追求する他の人間を支援する必要がある。ロボットはまた、誰かの計画が彼らの邪悪な資質を引き出したり、不道徳を生み出したりするとき、彼らへの支援を拒否しなければならない（仁 ; humanity）。

「CR1」が強調するのは、ロボットはそれ自身の役割によって規定された社会的責任を果たし、役割を超えたことをしない、という「忠」の美徳である。Liu によると、これは第一義的に重要な原則である。「CR1」のメリットは、「UR」に従う自動運転車の乗客を犠牲にする問題を解決できることにある。なぜなら、乗客の安全を確保することは自動運転車の役割なので、「CR1」に従えばロボットは乗客を犠牲にしないからである。「自動運転車は、乗客の安全運転を確保する義務を果たし、スクールバスの壊滅的な被害や複数の歩行者の死亡を防ぐために、木にぶつかって乗客を犠牲にするような行動をとるべきではない」（Liu. 2017：19）。

「CR2」は、「己の欲せざるところは人に施すことなかれ」という「恕」の美徳をロボットに組み込む形で実現する原則である。「CR2」の原則はアシモフの第一の原則よりも優れていると Liu は信じている。なぜならこの原則は、負の価値（negative values）を考慮することができ、ロボットは許容される行動範囲をより柔軟に考慮できるようになるからである。同時に、「己の欲せざるところは人に施すことなかれ」という儒家の否定的な黄金律に基づく「CR2」はロボットに独善的な行動をさせないので、「他人に主観的な選好を押し付ける」（subjective imposition of preferences）問題を回避できる。

「CR3」は「己立たんと欲して人を立て、己達せんと欲して人を達す」と「君子は人の美を成す、

人の悪を成さず」から抽出した「仁」に基づく規則である。「CR3」が要求するのは、ロボットは人間にとって何が良いのか、または人間が何を達成すべきかを独善的に決定しない、ということである。同時に、人の命令が悪のためであるとき、ロボットはその人を支援することを拒否する。

標準的なトロッコ問題に対して、儒家倫理に従うロボットが運転士あるいは車掌であれば、自身の役割を果たすため線路の分岐器を使って一人の命を犠牲にして五人を救うが、それ以外の役割であればロボットは何もしない。また、トロッコ問題の歩道橋のバージョン（Thomson. 1976：207-208）に対しては、能動的に人を橋の下に突き落とす行為は「CR1」と「CR2」と「CR3」の全てに違反するので、ロボットは絶対にしない。このように、儒家ロボットは、標準的なトロッコ問題では一人の命を犠牲にして五人を救い、歩道橋のバージョンでは何もしない、という多数の人が受け入れられる形でトロッコ問題に対処する。

「私たちの社会に自動制御の人工的道徳的行為者がいる予見可能な将来には、それが行動する場合と行動しない場合の両方が人間に害を及ぼし、望まぬ結果をもたらすときは、行動するよりも行動しないことを選択してもらいたい、と私たちは望むだろう」（Liu. 2017：26）。このように、役割を超えた行為をロボットにさせない、ということを Liu は極めて重視している。

中国の学術界では、Liu の論文を批判する二つの論文が存在する。以下ではそれぞれの内容を紹介する。

まず呉童立（2020）は、Liu の議論が基づいている思考実験、すなわちトロッコ問題と自動運転車の問題は道徳的な状況を抽象化・単純化しすぎており、いくつかの重要な要因を無視する可能性があると批判している。例えば、自動運転車の場合では、自動運転車が運転者や同乗者よりも歩行者の妊婦や子供の安全を優先することはありうる。しかし「車の乗客を優先に保護する」という Liu 主張には、そのようなことが考慮されていない。

さらに呉によると、Liu の論文の立場は一種の「理想的規則主義」である。それには二つの意味がある。まず、すべての道徳的状況に対処するための普遍的かつ合理的ルールシステムを設計できる。また、このルールシステムが厳密に使用されている限り、ロボットは道徳的危険を回避できる。「人間とは異なり、AI には感情的な干渉、自己境界、利益の訴えがなく、絶対的かつ合理的に指示に従うことができる」と「理想的規則主義」を支持する人は考えていると呉は述べ、続けて、そのような人は「ゆえに教科書のようなルールシステムで AI を規制するのは自然なことである」という態度を取ると呉は言っている（吴. 2020：54）。彼によれば、「理想的規則主義」が前提とするのは、人間は責任を負うことができる道徳的主体の資格を持っているのに対して、AI は道徳的主体の資格を持っていない、ということである[9]。

また、呉によると、Liu の論文における儒家思想に対する解釈には問題がある。なぜなら、「忠」と「恕」と「仁」を規範化する Liu の解釈は、それらの持つ豊かな意味を解消するという点で不適切だからである。「忠」と「恕」と「仁」はすべて、個体と共同体との動的な相互関係、つまり共同体における具体的な文脈に応じて行為することを強調する思想だが、Liu はそれらを規範化して

9　この立場と逆に呉は、「独立して状況を判断し、ルールを調整して使用することで決定を下すことができる AI は、真の自律性を持っており、道徳的な主体として認定されるべきである」（吴 . 2020: 56）と述べている。つまり、AI は人間のような自律性を持っているわけではないが、それ自身によって予測して行為することができる複雑なルールを持っているので、やはり自律性のある道徳的主体と認めるべきであると彼は考えている。

解釈した。従って、「忠」は「役割に忠実であること」と定義された。「恕」すなわち、「己の欲せざるところは人に施すことなかれ」は禁止される行為のリストとして解釈された。「仁」は「善いことをする際に人間に許可を求めなさい」ということと「悪を助けることを禁止する」こととに還元された。このように、Liu の解釈は儒家の倫理思想を単純化しすぎたと言える。

方旭東（2020）は、彼の論文で Liu の論文に批判的に言及している。彼によると、Liu が提案したロボットに適用する儒家倫理は AI に独善的な行動をさせないことだとすれば、AI を極めて低い知能レベルに規制しなければならない。従ってそのような AI は人工知能ではなく、知能を持っていないただの機械である。

さらに方は、Liu が提案した儒家倫理は、役割を果たすことを強調する「忠」を一番に重要な原則と見なすので、「道徳に無関心」な結果を引き起してしまう可能性があると指摘している。そのような「道徳に無関心」な行動は、「仁」または「良心」に基づく道徳的行動を称賛する儒家の思想と衝突する（方．2020：785）。例えば、儒家の古典である『孟子』においては、赤ん坊が井戸の中に落ちようとしているのを見たならば、誰でも驚いて同情心（人に忍びざるの心）がわき、助けるようとするということが書かれている。

4. AI と人間の関係について

中国では、とりわけ儒家思想の研究者が AI 倫理を考察する研究には、AI と人間の関係という話題も活発に議論されている。この話題についての研究には、「人と獣の区別についての弁論（人禽之辨）」と「惻隠の心」に基づいて AI と人間の区別に着目する論文や「万物一体論」という古い中国思想に基づいて AI と人間の共存関係を主張する研究もある。それぞれの話題に関する研究の主張を整理することは、AI 倫理に対する中国思想の研究者の主張を理解するために有益であり、また中国的な特色のある AI 倫理研究を見いだすことができるだろう。

2018 年孔学堂冬季弁論大会は、「儒家の「人間と獣との区別（中：人禽之辨）」はロボットに効果があるのか」というテーマで開催された。その弁論大会の内容は論文化され発表された。以下ではそれらの論文の主張を簡単に紹介する。

呉根友（2019）は、「人間と機械との区別（中：人机之辨）」を議論する意味は、科学技術による人間の疎外およびより大きな不公平の出現を防ぐことにあるとして、技術の進歩は人々の全面発展のためであると主張している。

戴茂堂と左輝（2019）は、人間に比べて、ロボットは自己意識がないので、自由意志がないと論じた上で、ロボットは理性的かつ感情的な反応をすることは不可能であり、道徳的な意識を確立することは不可能である。なぜなら、道徳自体は人間の自由意志の自律に基づいているからである。このように、彼は人間とロボットとの区別を論じている。

董平は（2019）、「人間と獣との区別」というテーマを検討する目的が人間の主体性および生命の尊さを強調し、尊厳のある生活と人生の目標の実現することであれば、「人間と機械との区別」を議論する目的も同じであると言っている。それゆえ、AI 技術の発展はこの目的のもとで行われるべきだと彼は主張している。

上述した 2018 年孔学堂冬季弁論大会についての論文以外に、儒家の「同情心」（惻隠の心）に依

拠して人間とロボットとの区別を考察する論文もある。

付長珍（2019）は、「人間は道徳的な主体として、共感能力に基づく責任感を持っている。これはロボットにはないものである」（付．2019：42）と述べている。なぜなら、ロボットの感情は恐らくさまざまな状況に対する反応にすぎず、自然な感情ではないからである。つまり、感情に対する反応と自然な感情という経験が区別されている。人間が両方を持っているのに対し、ロボットは前者だけを持っている。この点により、人間とロボットを区別することができると付は考えている。

上記の中国思想に依拠して AI と人間との区別を議論する研究の他に、中国の伝統的思想である「万物一体論」に基づいて、AI と人間との共存的な関係を提唱する研究も存在する。

端的に言えば、「万物一体論」とは、人間と世界の関係を理解するための伝統的な中国思想である。この思想は、道家の古典『荘子』や儒家の古典『中庸』に遡る。また、宋朝の儒学者である王陽明は、血縁親族への同情心を、鳥獣、草木、石などの自然的存在物まで拡張し、世界の万物の運命と人間の運命との関連を述べている。王陽明のこの主張は、以下の二つの論文の中にも具体的に言及されている。

朱承（2020）によると、「万物一体論」の視点は、「人間」を出発点としながら、「万物」と人自身の生存、発展ないし運命を一体と見なす視点である。彼は「万物一体論」に基づいて、われわれは AI と人間の一体性を強調すべきだと主張している。彼によると、一方で、AI は人間を模倣することによって造られたものなので、それを人間の意志の延長と見なすことができる。他方、AI は人間にとっての客観的な対象物なので、それを人間と対象世界の融合と見なすことができる。

彭国翔（2020）は、「万物一体論」に従うならば、AI はただ人間によって制御される、人間のための道具である、という人間中心主義を取るべきではないと主張している。彼から見ると、人間中心主義は意識のある AI を人間にとっての脅威だと見なす AI 脅威論の前提である。最初からこの態度に固執すれば、われわれは AI と共存する未来の可能性を否定するに違いない。AI 脅威論に囚われるより、われわれは、意識と感情を持つ AI を人間のような生き物（human-like-creature）と見なすべきである。そうすることによって、AI は必ずしも人間にとって脅威ではなく、人間の友人になる可能性があると彼は述べている。

結　語

中国が AI 技術を社会的範囲で大幅に応用している現状から、中国の AI 倫理研究に対して興味を持つ人は少なくないと思われる。本稿は、「次世代人工知能発展計画」2017 年 7 月以来中国で盛んに論じられている AI 倫理の研究についての議論を、主体性という問題に関する議論と、儒家のロボット倫理をめぐる議論、そして AI と人間の関係に関する議論に分けて整理した。これにより、AI は人間のような主体性を持った存在であるかという一般的な問いに対する中国研究者の主張のみならず、AI 倫理研究という現代の学問についての中国思想の研究者の独特な考えも示すことができた。本稿は中国国内の AI 倫理研究に対する全面的、包括的な調査研究ではなく、あくまでその一部の話題を紹介するものであるが、まだ十分に知られていない中国国内の AI 倫理研究の現状およびその動向を日本に紹介するものである。今後は、世界で盛んに論じられている最先端の AI 倫理研究の成果を参照した上で、中国における AI 倫理研究の価値を示すことに努めたい。

※ 本稿は「JST-RISTEX「科学技術の倫理的・法制度的・社会的課題（ELSI）への包括的実践研究開発プログラム「人工主体の創出に伴う倫理的諸問題を分析･討議するプラットフォームの構築に向けた企画調査」」による研究成果の一部である。

参考文献

Bonnefon, Jean-François, Rahwan, I., & Shariff, A.（2016）“The Social Dilemma of Autonomous Vehicles,” *Science*, 352（6293）, pp. 1573-1576.

Foot, P.（1967）“The Problem of Abortion and the Doctrine of Double Effect,” *Oxford Review*, 5, pp. 1-6.

Liu, J.（2017）“Confucian Robotic Ethics,” Conference: The Relevance of Classics under the Conditions of Modernity: Humanity and Science, Hong Kong: The Hong Kong Polytechnic University.

Thomson, J. J.（1976）“Killing, Letting Die, and The Trolley Problem,” *The Monist*, 59（2）, pp. 204-217.

博古睿研究院，全球视野下的人工智能伦理论坛 " 成功举办，2019-08-29.
https://www.berggruen.org.cn/activity/12、最終閲覧日 2021 年 3 月 24 日

陈小平 . 人工智能伦理建设的目标、任务与路径：六个议题及其依据 . 哲学研究，2020（09）：86-87.

杜严勇．人工智能安全问题及其解决进路［J］．哲学动态，2016（09）：99-104.

段伟文．人工智能时代的价值审度与伦理调适［J］．中国人民大学学报，2017，31（06）：98-108.

董平．“人禽之辨”与“人机之辨”：基础与目的［J］．船山学刊，2019（02）：11-14.

戴茂堂，左辉．人何以为人？—— 从“人禽之辨”到“人机之辨”［J］．船山学刊，2019（02）：15-19.

戴益斌．人工智能伦理何以可能？—— 基于道德主体和道德接受者的视角［J］．伦理学研究，2020（05）：96-102.

付长珍．机器人会有“同理心”吗？—— 基于儒家情感伦理学的视角［J］．哲学分析，2019，10（06）：34-43.

方旭东. 儒家对人工智能伦理的一个可能贡献 —— 经由博斯特罗姆而思［J］. 中国医学伦理学, 2020, 33（07）：778-788.

国家人工智能标准化总体组，人工智能伦理风险分析报告．2019-04-26.
http://www.cesi.cn/201904/5036.html、最終閲覧日 2021 年 3 月 24 日

韩敏，赵海明．智能时代身体主体性的颠覆与重构 —— 兼论人类与人工智能的主体间性［J］．西南民族大学学报（人文社科版），2020，41（05）：56-63.

胡珉琦，中国科学报，AI 与情感，2020-7-23.
http://news.sciencenet.cn/htmlnews/2020/7/443181.shtm、最終閲覧日 2021 年 3 月 24 日

金融电子化，【中国科学技术大学教授、中国人工智能学会人工智能伦理专委会（筹）主任 陳小平】人工智能伦理全景动态观，2019-12-11.
https://www.sohu.com/a/359758385_672569、最終閲覧日 2021 年 3 月 24 日

刘纪璐，谢晨云，闵超琴，谷龙．儒家机器人伦理［J］．思想与文化，2018（01）：18-40.

刘永安．人工道德主体是否可能的意识之维［J］．大连理工大学学报（社会科学版）：2021（01）：1-6.

彭国翔．人工智能最终一定是人类的威胁吗 —— 一个儒家的视角［J］．道德与文明，2020（05）：28-35.

上海科技，创新人工智能治理，青年科学家发声《上海宣言》，2019-08-30.
http://www.shkjdw.gov.cn/c/2019-08-30/517552.shtml、最終閲覧日 2021 年 3 月 24 日

腾讯科技，腾讯发布人工智能伦理报告 倡导面向人工智能的新的技术伦理观，2019-07-11.
https://tech.qq.com/a/20190711/004971.htm、最終閲覧日 2021 年 3 月 24 日

吴根友．儒家的“人禽之辨”对机器人有效吗？［J］．船山学刊，2019（02）：1-4.

吴童立．人工智能伦理规范是什么类型的规范？——从《儒家机器人伦理》说开去［J］．文史哲，2020（02）：51-59.

王天恩．人工智能存在性风险的伦理应对［J］．湖北大学学报（哲学社会科学版），2020，47（01）：1-8

杨庆峰．从人工智能难题反思 AI 伦理原则［J］．哲学分析，2020，11（02）：137-150+199.

赵汀阳．人工智能“革命”的“近忧”和“远虑”——一种伦理学和存在论的分析［J］．哲学动态，2018（04）：5-12.

朱承．万物一体视域下的人工智能［J］．学术交流，2020（05）：21-29+191.

アシモフ、アイザック（1983）『われはロボット』、小尾芙佐訳、ハヤカワ文庫

中国網（2019）「中国が次世代人工知能ガバナンス原則を発表」http://japanese.china.org.cn/business/txt/2019-06/19/content_74899392.htm、最終閲覧日 2021 年 3 月 24 日

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 14 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 14 号への掲載は 2022 年 9 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）の電子媒体を本センター事務局宛にメールで送付すること。
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. 13
2022 年 6 月 30 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/