



HOKKAIDO UNIVERSITY

Title	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
Author(s)	猪ノ原，次郎
Citation	応用倫理，13，18-33
Issue Date	2022-06-30
DOI	https://doi.org/10.14943/ouyourin.13.18
Doc URL	https://hdl.handle.net/2115/86432
Type	departmental bulletin paper
File Information	13_2_18-33.pdf



「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 *patency* 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 sharpen 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間置者か人間と天使の間置者（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、後方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при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ど？

〈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のおよび警察の活動、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はほぼ同時に——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ムティック・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喰うよりも「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ぬ。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、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(patency)——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・苦しみを感ずる有感存在(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取えてしない、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)はメタ倫理学的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。とりわけクークルバーク (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)ということになるだろう。

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ら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）、やはり倫理的には言うべきことはとくに無いのではないか。むしろ 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」を脱する仕方ではロボットの権利を擁護したい Gunkel などは、むしろ既存の人間の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／彼方のトラウマの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は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（人間と何らかの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的に考えると？ 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、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ブ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)、最近で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对象)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な位置は、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ブ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るだろうか。

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ア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实践理性への信に、

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「精神のおよび社会・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 anthropomorphizing 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、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るか？

大森とは別の筋道を通してではあるが、ここにも道德の始原を人間の自然的社會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、いかなる道德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、反動的態度 (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「個性性 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、世界に魂の氣息を「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、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、容易に解きたい古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、これはバトナムが1964年に出した結論にほぼ等しい (Putnam 1964: 668-91)。バ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ものである (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) の特に第三章と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的＝政治哲学的パースペクティブ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. & Anthi, 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
- クーケルバーク、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・ススバウム編『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