

Kantian Dog Argument とロボット倫理

—ロボットの道徳的地位の擁護に関する批判的考察—

Kantian Dog Argument and Robot Ethics

—A Critical Examination of the Defense of the Moral Status of Robots—

清水 颯

**Abstract**

Do robots have moral status? If they have, should we give them moral consideration for their own sake? This paper focuses on Coeckelbergh's paper, which argues that robots have an indirect moral standing from a relational approach. In particular, I critically evaluate Coeckelbergh's argument, which presents a Kantian dog argument as its starting point. I will then argue that robot ethics, which takes Kantian dog argument as its starting point, is incompatible with arguments in defense of the moral status of robots.

**(1) 研究テーマ**

ロボットは道徳的地位をもつだろうか。あるいは、ロボットは道徳的配慮に値する存在なのだろうか。人間とロボットの共生が想定される社会においては、我々はロボットをたんなる道具として扱うのではなく、何らかの意味で道徳的に配慮することが必要になるかもしれない。しかし一方で、ロボットは機械であり、人間（や動物）のような心的状態を持つと考えるのは難しい。そうであれば、ロボットは道徳的地位をもちえないようにも思われる。

さて、道徳的地位とは何だろうか。従来、道徳的地位の問題としては、「道徳的行為者性(moral agency)」と「道徳的被行為者性(moral patiency)」の二つが扱われてきた。道徳的行為者性の問題においては、その存在が自律的な意志や意識をもち責任を負うことができる主体であるかどうか問われ、道徳的被行為者性の問題においては、その存在が道徳的配慮に値するがゆえに尊重されるべき存在であるかどうか問われてきた<sup>(1)</sup>。本稿が問題にするのは、ロボットが道徳的被行為者としての道徳的地位をもつかどうかである。

**(2) 研究の背景・先行研究**

ロボットが道徳的配慮に値する存在であるか、すなわち道徳的地位をもつかどうかという問題へのアプローチは様々である。ロボットや AI の倫理を

先導する倫理学者の一人である M. Coeckelbergh によれば、ロボットの道徳的地位をめぐる見解は、「特性アプローチ(properties approach)」と「関係論的アプローチ(relational approach)」という二つの立場に大別できる。特性アプローチは、例えば意識や痛みを感じる能力などの道徳的に重要な内在的な(intrinsic)特性がある存在は、道徳的配慮に値する存在であるから道徳的地位をもつ、という主張をする。Coeckelbergh は、D. Gunkel とともに、意識や感覚などの内在的な特性に基づいて道徳的地位の要件を定めてきた特性アプローチを批判する<sup>(2)</sup>。批判のポイントは、人間であれロボットであれ、その内的状態について確信することはできないという懐疑である。そこで、Coeckelbergh らは新たなアプローチとして関係論的アプローチを提唱し支持する(cf. Coeckelbergh 2021: 339)。

特性アプローチと対照的に、関係論的アプローチは、「機械の特性に基づいて道徳的地位を機械に付与するのではなく、その地位がロボットとの関係の中でどのように形成されるかに焦点を当てる」ものである(Coeckelbergh 2014: 71, Coeckelbergh 2021: 339)。Coeckelbergh は、特性アプローチに基づいた直接的な道徳的地位を問題にせず、関係論的アプローチから帰結する道徳的地位に焦点をあて、それを「間接的な」道徳的地位と呼ぶに至る(Coeckelbergh 2021: 340)。この「直接／間接」の区別は、カントの徳論で展開される「間接義務」の議論を参照し、それをより広義に道徳的地位についても適用することで提案されたものである。それゆえ、Coeckelbergh は、間接的な道徳的地位を擁護するための推論を助けるアイデアとして、「Kantian Dog Argument」という議論を提示し、それを出発点とする(ibid: 341)。

では、Kantian Dog Argument とは何か。Kantian と名がつくように、それはカントに触発された議論であるが、その出自は『道徳の形而上学』の「徳論の形而上学的基礎」(以下『徳論』)や『倫理学講義』の一部で展開された、「動物に関する間接義務」についての議論である<sup>(3)</sup>。カントは実際に、人間以外の存在をむやみに残虐に扱わないことや、それに感謝の気持ちをもつことは、間接的に義務であることを認める。間接的である理由は、それがあくまで人間に対する義務だからである。道徳感情を弱め、悪徳な性格が形成されることによって、人間が人間に対して負う義務に違反するため、それを間接的な根拠として、我々は動物などの非ヒト存在を配慮する必要がある。当然、カントはロボットと人間の共生などを想定していないため、ロボットを配慮する必要があるかどうか、答えを与えてくれるわけではない。しかし、Coeckelbergh は上記の Kantian Dog としてロボットを想定することで、同じ理由でロボットも配慮するべきであると提案する。これを Kantian Dog

Argument として定式化したのは Coeckelbergh だが、カントの間接義務を参考にして、ロボットへの振る舞いが人間の性格特性に影響を与えるという理由から、ロボットに危害を加えることを控えるよう主張する論者には、K. Darling や A. Gerdes、R. Sparrow などもある。

Kantian Dog Argument によれば、ロボットはそれ自体として道徳的地位をもつわけではないが、人間との関係を媒介する仕方でも間接的に道徳的地位を獲得できると考える。この主張を擁護するために、Coeckelbergh は四つの条件（一つが満たされればよい）を提示する。その筆頭に挙げられているのが、次のような徳倫理的な条件である。「もしロボットに悪いことをするような人間がいたら、他の人間からは性格が悪い、悪い人間（徳倫理でいうところの徳がない、人間に悪いことをする危険性がある）と見られるだろう」(ibid: 341)。この条件は、上記で述べた Kantian Dog Argument に直接基づいた議論であり、本稿ではこの第一の条件に焦点を当てる<sup>(4)</sup>。カント的な議論では、理性的存在者のみが直接的な義務の対象であり直接的な道徳的地位をもつが、『徳論』においてカントは、人間への影響を考慮して動物への行為を義務に関連付ける。Coeckelbergh も引用しているように、カントは「動物の苦痛に対する人間のうちなる共感が鈍くなり、そのことによって、他の人間との関係における道徳性に非常に役立つ自然的素質が弱められ、そのうちに根絶されてしまう」ことを理由に、動物への残虐な扱いを禁止する (MS VI: 443)。これは、間接義務として知られる議論の基本的な枠組みである。

この議論を徳倫理的に再構成すると、次のようになる。「もし私たちが犬に対して残虐に振る舞う習慣がついた場合、つまり、この悪徳を培い、性格が凶悪になった場合、私たちは人間に対して凶悪になる可能性が高くなる」(Coeckelbergh 2021: 343)。カントはあくまでも動物についてのみこの議論をしているが、Coeckelbergh は、人間の徳に注目する場合、ロボットにも同じ議論が適用できなければ首尾一貫していないと指摘する。これは人間の有徳な性格についての議論であるが、Coeckelbergh はこの議論を関係論的な仕方 (relational way) で定式化する<sup>(5)</sup>。それによれば、「何が有徳なのかは、私自身の評価だけでなく、他者の評価にも依存する」ため、「ロボットは徳に損害を与えるような行為をする可能性のある人間の徳に対する懸念に基づいて、道徳的地位を与えられうる」(ibid)。

このように Coeckelbergh は、関係論的アプローチと、人間に対する残虐さにつながるかもしれないという「道徳的リスク」への懸念に基づいて、ロボットを配慮すべき存在として位置付ける (ibid: 353)。しかし、Coeckelbergh も指摘するように、この結論は定言的なもの (categorical) では

なく仮言的なもの(hypothetical)である(ibid: 354)。ロボットを含めた非ヒト存在の直接的な道徳的地位が論争的である場合には、カントを源流とする間接的な議論が有効であると示したに過ぎない。しかし、カントの『徳論』に基づいた **Kantian Dog Argument** が、ロボットの間接的な道徳的地位を擁護する関係論的アプローチを発展させるための出発点になりうるのだろうか。例えば、動物倫理の議論においては、カントの動物に関する義務は動物の道徳的地位を間接的にでも擁護することを意図したものではなかったと評価されることが多い<sup>(6)</sup>。むしろ、非ヒト存在は人間の道徳的地位のための道具的で手段的な地位しかもたないはずである。また、カントの枠組みでは、人間への影響を考慮して動物などの非ヒト存在への道徳的な振る舞いを規定するが、そこで具体的な他者との関係性が問題になるのだろうか。次節では、人間の徳からロボットへの道徳的配慮の問題を考える徳倫理的なアイデアは、カントを出発点とする場合、どこに向かうべきなのかを提示したい。

### (3) 筆者の主張

本節では、Coeckelbergh のように **Kantian Dog Argument** を出発点とする議論は、関係論的アプローチに基づいてロボットの間接的な道徳的地位を擁護するためには有効ではないことを主張する。Coeckelbergh は、**Kantian Dog Argument** を援用して、ロボットが間接的な道徳的地位を持つことを擁護するが、この議論はロボットが人間の徳を促進する手段として保護されるべきであるという点に焦点を当てており、ロボット自体が道徳的配慮の対象となるべき存在であるとは述べていない。それゆえ、**Kantian Dog Argument** に基づいて道徳的地位を擁護しようとするアプローチは、結局のところロボットの道徳的地位を根本的に擁護するものではないように思われる。**Kantian Dog Argument** によれば、ある存在がどのような理由であれ道徳的配慮の対象となる可能性がある場合、その存在に道徳的地位が認められると見なしているように思われるが、この理解は、その存在の固有の道徳的価値を認めるものではない。そこで、**Kantian Dog Argument** を出発点とするならば、ロボットにはそれ自身のために配慮されるべき道徳的地位はないとみなしたうえで、ロボットがそれとは別の理由で配慮される可能性があるのではないか、という問題設定にするべきだと提案する。

まずは、カントの見解をもう少し詳細にみていきたい。これまで見てきたように、Coeckelbergh も正しく指摘しているが、カントの主張の焦点は、人間の有徳さにとって重要である自然の素質、すなわち道徳感情を損なわせることを防ぐための間接的な手段として、動物への残虐な扱いを禁止すること

である。この議論が展開されるのは、「道徳的反省概念の両義性」、すなわち自己に対する義務を別の存在に対する義務と取り違えることについて説明される挿入章である(MS VI: 442)。つまり、カントはここで、動物に対する義務があるように見えるかもしれないが、それは自己に対する義務であることを示している。ここで想定される自己に対する義務とは、「完全性の義務」であり、有徳な性格を完成させる義務を意味する(cf. MS VI: 385-7)。動物を残酷な仕方で扱うと有徳な性格が損なわれる可能性があり、自己に対する義務に違反するため、そのような扱いは控えるべきなのである(7)。

ここで重要なのは、道徳的配慮の対象は誰なのか、という問題である。あくまで動物に関する義務は、自己に対する義務に関連する手段として導出されるため、義務がそれに対して向かうのは人間(とりわけ自己)である。それゆえカントの間接義務の議論では、動物に関する(regarding/in Ansehung)義務はあるが、動物に対する(toward/gegen)義務はない(MS VI: 443)。そしてカントは、「それはいつでもただ人間の自己自身に対する義務にすぎない」と結論づける(ibid)。ロボットについても同様に、人間のみが義務として配慮されるべき対象であり、その手段としてロボットへの残酷な扱いの禁止が導出される。それゆえ、**Kantian Dog Argument**を出発点とする場合、ロボットが道徳的配慮の対象でありうるのは、あくまで人間のためであることがわかる。しかも、ここでは自己の有徳さを完成させるための陶冶プロセスのみが焦点であるから、ロボットとの具体的な関係やインタラクション、それによる他者からの評価はとりわけ関係のないものである。それゆえ、**Kantian Dog Argument**と関係論的アプローチの結びつきは自明ではない。例えば、Coeckelberghが**Kantian Dog Argument**と呼ぶ議論と同じ個所を参照し、徳倫理的な観点からロボットへの残酷な扱いを控えることを主張するSparrowは、関係論的アプローチにはコミットしていないし、それゆえロボットの道徳的地位を擁護することを意図しない(cf. Sparrow 2017)。

**Kantian Dog Argument**は、非ヒト存在が、人間の徳を中心とした道徳的配慮の一部に入りうることを正当化するためには有効な出発点だろう。しかし、それは関係論的な観点から道徳的地位を与える議論ではないため、ロボットの道徳的地位を擁護する議論の出発点としては有効ではない。少なくとも、**Kantian Dog Argument**に触発された徳倫理的な議論は、関係論的アプローチの枠内で間接的な道徳的地位を擁護する議論から独立させる必要がある(8)。

また、**Kantian Dog Argument**と呼ばれるカント的な間接義務の議論では、ロボットの道徳的地位を擁護し、それらが道徳的に配慮されるべきであるこ

とを主張することはできない。カント的な議論では、ロボットはそれ自身のために道徳的に配慮されるべきものではない、すなわち道徳的地位をもたないことを前提にしたうえで、人間の徳に依存する仕方でも配慮を要求するからである。間接義務の議論では、それ自身が道徳的に配慮されるべき存在ではないとしても、別の理由から配慮される必要性があることを示唆するだけで、それ自身が道徳的に配慮されるべき道徳的地位をもつことを擁護できない。つまり、**Kantian Dog Argument** は、本来道徳的地位のない存在に対する道徳的な振る舞いを規定する議論であり、それらに道徳的地位を与えるというのはややトリッキーである。あくまでカント的な議論を出発点とするなら、その存在が道徳的配慮の対象としての道徳的地位をもつとは別の理由から、すなわち自分の徳を損なわせないという理由から配慮の対象範囲に含まれることは可能だが、それだけで道徳的地位を与えることは避けるべきであると考えている。なぜなら、道徳的地位が問題にしていることとは、「その存在がそれ自身のために道徳的に配慮されるべきか否か」だからである (cf. Kamm 2007, Jaworska, Agnieszka and Julie Tannenbaum 2023)。

Coeckelbergh は関係論的アプローチと **Kantian Dog Argument**、徳理論的議論をつなげて、道徳的地位をあまりに薄く使用しているように思われる。例えば、この議論は「物を大切にしない人は徳がない」という議論と平行であると考えた場合、椅子にも間接的に道徳的地位が与えられるべきである、という主張がなされるはずである。これは、どこまで道徳的地位に値する存在とみなせるのか、という厄介な問題を抱え込む。しかし、本稿が明らかにしたように、カント的な着想に基づいた徳倫理的な議論はロボットに道徳的地位を与えることなく、別の理由からそれを道徳的配慮の対象に含めることができる。少なくとも、**Kantian Dog Argument** を出発点とする徳倫理的な観点からロボット倫理を検討するならば、その存在に道徳的地位がなくとも道徳的配慮が要求される可能性があるという方向で語る方が望ましいと言える<sup>(9)</sup>。本稿が論じてきたように、少なくとも **Kantian Dog Argument** を出発点とする徳倫理的な議論であれば、ロボットは道徳的地位をもつから配慮すべきだ、という実践的な含意は導くことはできない。しかしこの議論は、それ自身が配慮に値する存在でないとしても、その存在に対する倫理的振る舞いが規定される可能性を示唆している。

#### (4) 今後の展望

本稿は主に Coeckelbergh の **Kantian Dog Argument** を出発点とする関係論的アプローチに基づいて、徳倫理的な観点からロボットの道徳的地位を擁

護する試みを批判的に検討してきたが、今後は関係論的アプローチの展望と徳倫理的議論の展望をより広いスコープから見る必要があるだろう。本稿でも多少触れたが、ロボットに道徳的地位はないと考える立場から徳倫理的な議論を採用する Sparrow や、関係論的アプローチからロボットの道徳的地位を擁護しつつ徳倫理的な観点を採用しない Gunkel など、それぞれの立場の主張が独自の仕方で多様に展開しているように見える。ここで論点になるのは、それぞれの立場が考える「道徳的地位」や「道徳的配慮に値すること」の概念的な位置づけである。しかし、現状のロボット倫理の議論では、ターミノロジーが整理されていないこともあって、立場によってそれらの概念がずれており、それぞれの観点が複雑に交叉している状況にある。本稿の貢献の一つは、その立場上の概念的な混乱を整理することにあっただと言えるが、今後は、既に多くの蓄積がある動物倫理や医療倫理における道徳的地位に関する議論などを参考にしつつ、より包括的な観点からロボットと人間の共生についての倫理を考えていく必要があるだろう。

#### 注釈

(1) ロボット倫理において行為者性(agency)と被行為者性(patency)がどのように扱われてきたかについては、Froridi 2013、猪ノ原 2022 を参照した。

(2) Gunkel は、レヴィナス的な他者論を根拠にして、ロボットが他者として現れるのであれば、それを根拠にロボットに権利を与えるべきであると主張する (cf. Gunkel 2018: 95-96)。

(3) カントの著作に関してはアカデミー版カント全集を用い、慣例に従って、巻数をローマ数字、頁数をアラビア数字で示す。MS は『道徳の形而上学』(Die Metaphysik der Sitten) の略称である。

(4) 紙面の都合上、残り三つの条件には触れることはできないが、それらを整理すると、「2. 関係的なケア感情に基づく条件」、「3. 共同行為に基づく条件」、「4. 予防的な議論に基づく条件」である(Coeckelbergh 2021: 341)。

(5) 2010年代の Coeckelbergh は、人間とロボットの具体的な関係性やインタラクションそれ自体を基礎としていたが、2020年代に入り、その関係を通じた人間への影響に焦点を移しているように思われる。これは、立場変更とまでは言えずとも、内在的な展開と評価することは可能だろう。

(6) カントが動物に対する義務はないと述べて動物の道徳的地位を否定したことはよく知られる見解であり、それは多くの動物倫理学者を不満にさせたと評価される(cf. Müller 2022: 59)。

(7) カントは他人に対する共感などの道德感情が損なわれないような扱いができれば、動物それ自体が積極的な配慮の対象であるとは考えない。それゆえ、『徳論』の同個所で、動物を使役することや、目的達成のために遂行される苦痛の少ない動物実験を行うことでさえ許容されている。

(8) カントの間接的な議論がロボット倫理でいかに位置づけられているのかを明確にすることを意図して書かれた Flattery 2023 では、関係論的アプローチが擁護しようと試みる間接的な道德的地位が議論の前提から外されている。これも、関係論的アプローチと Kantian Dog Argument の相性の悪さを物語っているように思われる。

(9) ある存在が道德的地位を持たずとも、それが私たちの道德的配慮の一部となりうることは、道德的地位に関するより一般的な議論の中でも成立している (cf. Kamm 2007: 299)。

## (5) 参考文献

- Coeckelbergh, M. (2010). "Moral Appearances: Emotions, Robots, and Human Morality," *Ethics and Information Technology*, 12, pp. 235-241.
- Coeckelbergh, M. (2014). "The moral standing of machines: towards a relational and non-cartesian moral hermeneutics," *Philosophy & Technology*, 27(1), pp. 61-77.
- Coeckelbergh, M. (2021). "Should We Treat Teddy Bear 2.0 as a Kantian Dog? Four Arguments for the Indirect Moral Standing of Personal Social Robots, with Implications for Thinking About Animals and Humans," *Minds & Machines*, 31, pp. 337-360.
- 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.
- Flattery, T. (2023). "The Kant-inspired indirect argument for non-sentient robot rights," *AI Ethics*, Online publication date: 05-July-2023, <https://doi.org/10.1007/s43681-023-00304-6>.
- Froderi, L. (2013). *The Ethics of Information*, Oxford University Press.
- Gerdes, A. (2016). "The Issue of Moral Consideration in Robot Ethics," *Acm Sigcas Computers and Society*, 45 (3), pp. 274-279.
- Gunkel, D. (2018). "The Other Question: can and should Robots have Rights?," *Ethics and Information Technology*, 20, pp. 87-99.



- Jaworska, A., and Julie T. (2023). “The Grounds of Moral Status,” *The Stanford Encyclopedia of Philosophy* (Spring 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.). <https://plato.stanford.edu/archives/spr2023/entries/grounds-moral-status/>.
- Kamm, F.M. (2007). *Intricate Ethics: Rights, Responsibilities, and Permissible Harm*, Oxford University Press, New York.
- Kant, I. (1908). *Kants Gesammelte Schriften*, Hrsg. von der Königlichen Preußischen Akademie der Wissenschaften und Nachfolgern, Bd.VI, *Die Metaphysik der Sitten* (MS), Berlin. [樽井正義、池尾恭一訳 (2002). 『カント全集 11 人倫の形而上学』、岩波書店]
- Müller, N. D. (2022). *Kantianism for Animals: A Radical Kantian Animal Ethic*, New York: Palgrave Macmillan.
- Sparrow, R. (2017). “Robots, rape, and representation,” *International Journal of Social Robotics*, 9(4), pp. 465–477.
- Sparrow, R. (2020). “Virtue and vice in our relationships with robots: Is there an asymmetry and how might it be explained?,” *International Journal of Social Robotics*, 13, 23–29.
- 猪ノ原次郎 (2022). 「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」、『応用倫理—理論と実践の架橋—』(北海道大学応用倫理・応用哲学研究教育センター発行オンラインジャーナル)、vol. 13、18-33 頁。

(北海道大学)