



Title	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？
Author(s)	草野, 原々
Citation	応用倫理, 13, 34-44
Issue Date	2022-06-30
DOI	https://doi.org/10.14943/ouyourin.13.34
Doc URL	https://hdl.handle.net/2115/86433
Type	departmental bulletin paper
File Information	13_3_34-44.pdf



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ごすだけではない。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る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して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して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たい。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なになるだろう。

- (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)

して見る設計的态度、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态度である。人々は、普段はロボットに対して志向的态度を向けているが、ロボットが悪さをしたときには、設計的态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、人間は罪をつぐなつた他者に対しては「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、つまり志向的态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、設計的态度を向けるだろう。

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（志向的态度をとっている）。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（志向的态度）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、社会的情動を持たなくなるだろう。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）設計的态度をとるといふことがあげられる。

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ブ的に存在している」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、「意志的な不信の宙ぶり」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ぶりにするというものだ。

フィクションの哲学では、「意志的な不信の宙ぶり」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、いったん不信を宙ぶりにしよう」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、ダンスに小指をおつけたとき、一瞬ダ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的な怒りだと言えよう。法廷でダ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的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、強弱のグラデーシオンのなかにある。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、映画を見ていると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と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出すことができる。

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出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」として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が（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」と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」と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、ここでこ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思われていたものが、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かもかもしれない。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。