

## 2.2.11 ロボティクスと社会

### 共生社会にむけた統合知としてのロボティクスの役割と社会変革の様相

本項は、研究開発領域の域を超えたメタな内容となるため通常の記述形式に捉われない形で記述する。

#### (1) はじめに

近年の人工知能（AI：Artificial Intelligence）やロボット技術の進展はめざましく、人と接しない構造化された環境から、物理的にも心的にも複数の人と接する非構造化環境へのロボットの導入が進んでいる。人間社会における機能優先の技術開発のみではなく、社会におけるロボットの役割を明確に意識した設計論が望まれる。特に、最先端AI技術などが搭載された高度自律型ロボットが人と相互作用する場合、単なる道具からパートナー的な存在として位置づけが出現し、道徳や倫理、最終的には法制度の問題も絡んでくる。本稿では、これらの課題を扱う上で、最初に、ロボティクスをさまざまな学問分野から構成されるtrans-disciplinary（超域）として再考する。次に、人と相互作用するロボットの設計論の背後にあると想定される哲学的背景を概観する。そして、社会における自律性の意味を問いなおし、共生社会における倫理のあり方を議論する。最後に法制度の面から社会制度改革への提言につなげる。

#### (2) キーワード

ロボティクス再考、哲学的背景、自律性概念、倫理、法制度、構成的人間学、設計原理、身体性、自己

#### (3) 再考：ロボティクス

ロボティクスの要素技術や統合化・システム技術、各応用に関しては、本報告書のそれぞれの項目を参照頂くとして、本稿では、最新AI・ロボット技術の社会実装による未来共生社会に向けた課題やアプローチを大局的な側面から見直すことで、ロボティクスを再考する。

2003年6月に第17期日本学術会議の吉川弘之会長のもとに新しい学術のあり方が模索され、社会のための学術と文理の融合では、以下が報告されている<sup>1)</sup>。

科学技術の利用の量的・規模的拡大、複合化が行われると、小規模、個別的利用では無視し得た影響が拡大し、その影響の質も変化し、後戻りができない修復し難い事態を生み出す。これをどの様に予測し、排除していくかは、現代科学の利用に当たって我々が克服しなければならない課題である。この課題に挑戦するためには従来の狭い範囲で専門化された‘認識’を中心とする科学では対応が困難である。科学技術の適用に当たって、それがもたらす便益（ベネフィット）と危険（リスク）を広い立場から俯瞰し、便益を増大するとともにリスクを排除していく学術的仕組みを構築することが必要である。理学、工学の立場だけでこれを行うことはできない、技術の適用について、必要な社会制度、外部的に処理される経済を内部化するための仕組み、これまでの生活文化に対する影響、自然環境・人類の生存に与える影響などを広く考察することが必要である。このためには哲学、倫理学、心理学、社会学、法学、経済学、エコロジー、医学などの観点を加えて、従来の学問体系を横断して研究が行われなければならない。俯瞰的研究プロジェクトはこの研究体制を実現の一環として行われる活動である。

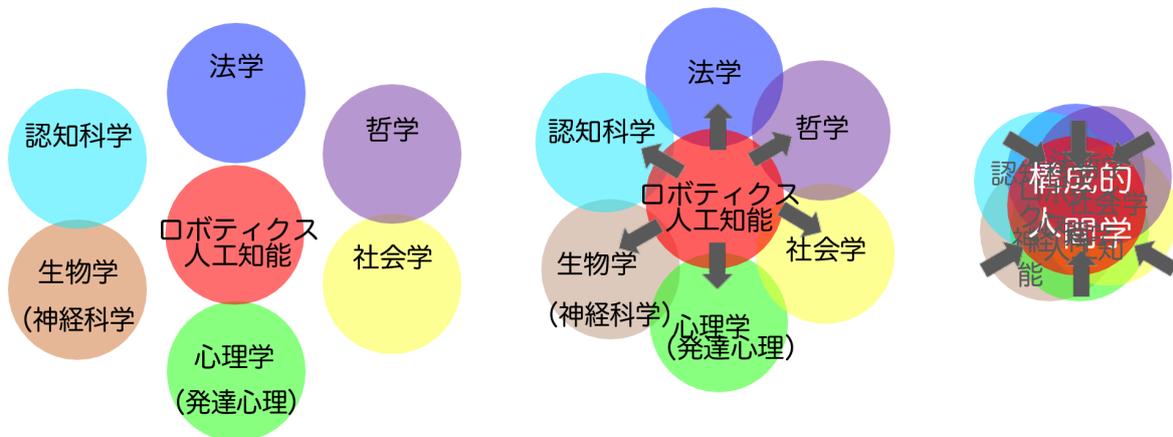


図2-2-7 Multi-disciplinary (左) から Inter-disciplinary (中) を経て Trans-disciplinary (右) へ

さらに、文系と理系を横断する「設計科学」というディシプリンが提唱され、俯瞰的研究プロジェクトの推進が謳われている。これに先立って、1999年に認知ロボティクスの提案<sup>2)</sup>がなされているが、設計科学の典型がロボティクスであろう。Roboticsはロボット工学と訳されてきたが、本稿ではロボット学と称する<sup>1)</sup>。機械、電気・電子、情報、制御工学、材料などの工学的分野の集積であるばかりでなく、人文社会系の学問分野との連携もいわれて久しい。しかし、真の融合に至っていないのが現状である。図2-2-7に示すように、multi-disciplinaryは、多くの学問分野が共存している状態、inter-disciplinaryは、各学問分野が少し重なりはじめて、異なる分野間の協働が始まっている状態だが、より進めて、trans-disciplinary、すなわち超域として新たな学問分野が構成されなければならない。それが、ロボット学が目指す姿である。

ロボット学に関連する学問分野は、既存の工学分野に加えて、生物学（神経科学）、認知科学、心理学（発達心理）、社会学、哲学、倫理学、法学などが考えられる。これらの既存の学問分野から見れば、ロボット学は、学際融合の橋渡し役であり、リサーチビークルの役割が期待されている（図2-2-7の中央）。既存分野のアプローチは、説明原理に基づくものが多く、設計原理に基づくロボティクスの手法により実証・検証されるとみなされる。このような既存分野の補助的な役割に留まらず、ロボット学は、これらの分野を内包し、新たな規範のもとに、再構築されなければならない。この規範の核となるアイデアが「構成的人間学」である（図2-2-7の右）。

構成的人間学における「人間」は、種としてのヒト、社会的存在として人を表すと同時に、他の種も含めた生物の代表としての意味合いも持つ。「構成的」とは、計算機シミュレーションやロボットなど機械の設計・製作・作動を通じて、仮説の検証や新たな仮説の枠組みの提案能力などの特徴を示している。この点が、先に示した「設計科学」の意味合いである。構成的人間学における仮説は、微視的レベルから巨視的レベルにわたり、各分野との相互浸透的な相互作用が必須である。形骸化しやすい学会間連携ではなく、研究者レベルで密な連携から進めるべきであろう。

近年注目されることの多い自動運転を「ロボット学」の視点から再考すると、技術的な課題だけでなく、より深刻な法制度の問題が表出している。法律専門家にまかしておけばいいという問題ではない。技術の課題

1 社団法人日本ロボット学会の学会誌の呼称は旧字を用いた「ロボ學」で再考の意味が込められている。

とそれが社会に及ぼす影響を最初の段階から考慮・検討しなければならない。その際、技術者だけではなく、他分野の研究者とも議論できる場が、特殊ではなく、それが普通にならないといけない<sup>3)</sup>。

SF作家であり、かつ科学・技術コメンテーターの瀬名秀明は、日本ロボット学会2020年度学術講演会RSJ2020特別公演「ロボティクスと総合知」において、ロボット学で、何が可能かとの問いに対して、「ロボット学が総合知を拓くと示すことの大切さ」を主張している。背景には、今回のコロナ禍によるパンデミック状況において、専門家の意味が問いただされ、専門家に責任を押し付けるのではなく、総合知による対処の可能性を見出すことが重要で、その過程にロボット学が貢献すべきであると説く。また、ロボット学と人間学の共存も説いている<sup>4)</sup>。

#### (4) 共生社会構築の哲学的背景

人と相互作用する最先端ロボットの設計思想の背景にあると想定される哲学的考察を紹介する（文献5）の第三章）。もともと、認知発達ロボティクス<sup>6)</sup>の思想的背景として説明<sup>7), 8)</sup>でもあるが、共生社会に向けたロボットの設計思想にも関連するので紹介する。

図2-2-8にその流れを示す。これは、Taniの書籍<sup>9)</sup>や稲谷の論説<sup>10)</sup>に基づき、イマヌエル・カント(Immanuel Kant)とイタリアの哲学者ジャンバッティスタ・ヴィーコ(Giambattista Vico)や徳倫理のジュリア・アナス(Julia Annas)を追加したものである。ルネ・デカルト(René Descartes)は、心と身体、もしくは事物の関係に関して、心身二元論<sup>2)</sup>を唱え、近代哲学の基礎を築いた。

一方、ヴィーコは、デカルト主義やあらゆる還元主義に反対の立場をとり、真実は、デカルト主義にしたがった観察によってではなく、創造か発明によってしか検証されないという真実の原理(Verum Factum Principle)を主張した<sup>3)</sup>。この考え方は、ロボット学を言い換えた「構成的人間論」の構成的手法の思想的原点である。従来の説明原理に基づく科学ではなく、設計原理を内包した新たな学問分野としての提案であ

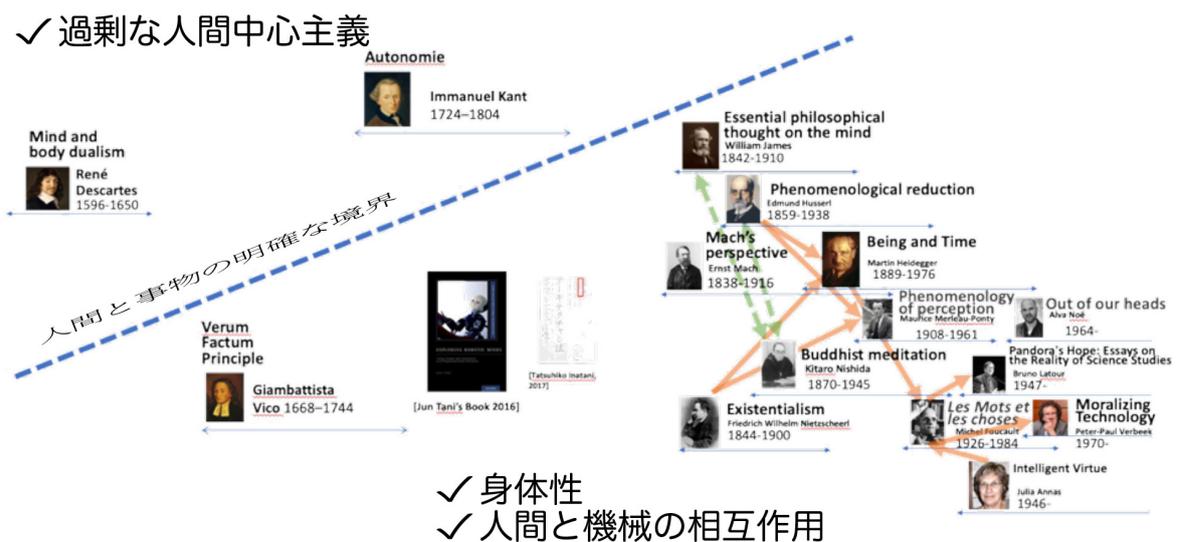


図 2-2-8 共生社会構築の哲学的背景

2 実体二元論、物心二元論、霊肉二元論、古典的二元論とも言われる。  
3 [https://en.wikipedia.org/wiki/Giambattista\\_Vico](https://en.wikipedia.org/wiki/Giambattista_Vico)

り、先に示した「設計科学」の原点とも言える。それがこの時期に提唱されていることは、重要である。

カントは、自身が唱える道徳哲学において、義務的な行為としての道徳のありかた、すなわち「どうあるべきか」といった視点から義務論的道徳を説いた<sup>11)</sup>。科学技術の高度な発達で、さまざまな意味で価値観をシフトさせる現代においては、このような人間中心の考えによる義務論的道徳が通じなくなっている。ミシェル・フーコー (Michel Foucault) は、アナスが説く徳倫理<sup>12)</sup>に注目したと言われており<sup>13)</sup>、「どうあるべきか」ではなく、「どうありたいか」がより重要であると考えた。前者はトップダウンに規定されると説いているのに対し、後者は、科学技術の高度な進展に対応するために、新たな倫理観を常々更新する必要があることを説いている。

デカルトを超えて超越論的現象学へと進む「新デカルト主義」を主張し、現象学的考察を与えたのは、エドムント・フッサール (Edmund Gustav Albrecht Husserl) である (例えば、14) など)。我々は客観的な物理世界が事前に存在すると捉えがちだが、そうではなく、個人の意識的な主観体験による表象が、共有できる可能性としての客観性、すなわち主観と客観の狭間の中主観性の考え方を展開し、後世に多大な影響を与えた。

フッサールの現象学を拡張・進化させたのが、マルティン・ハイデッガー (Martin Heidegger) (『存在と時間』<sup>15), 16)</sup> やモーリス・メルロー＝ポンティ (Maurice Merleau-Ponty) (『知覚の現象学』<sup>17), 18)</sup> である。

ハイデッガーは間主観性とは異なるアプローチで主観性を取り上げた。すなわち、主観と客観を分けずに、いま現実に存在すること (「現存在」) の重要性を説き、これが、時間軸上の過去と将来との動的な相互作用によって生じると主張している。人間という存在を時間軸で捉えるとき、現在という切断面で見えるのが「現存在」だが、それは、決して孤立しているわけではなく、過去と未来の駆け引きの隙間にあると解釈できる。また、現存在する個々のエージェントは、それぞれが目的を持って相互作用しているという合意のもとに、個々が相互に存在しうる。これは、社会的相互作用の重要性の一つでもある。

メルロー＝ポンティは、主観と客観に加えて身体性という次元が創発し、そこでは、同じ身体が、触れたり見たりする主体と同時に触れられたり、見られたりする客体にも与えられうると主張している。このことは、主観と客観の二つの極の間の繰り返される交流の場を身体が与えているのだ。すなわち、客観的物理世界 (先の議論では偶像に過ぎない) と主観的経験をむすぶメディアとしての身体的重要性を指摘している。これは、ロボット学における「身体性」の基本概念の根幹である。

デカルトやカントらは、主体と客体とを厳格に区分する近代的な思考法 (人間存在のあり方を本質化する、ヒューマニズム) の提唱者だが、その考え方に頼っていると、現代社会を適切に扱えないとブルーノ・ラトゥール (Bruno Latour)<sup>18)</sup> は警告する。なぜかという、高度な科学技術の進展により、現代社会は主体と客体が入り混じったハイブリッドな世界となっているからだ。また、ピーター＝ポール・フェルベーク (Peter-Paul Verbeek)<sup>20)</sup> は、「技術は、我々の行為や世界経験を形成し、そうすることによって、我々の生活の仕方に能動的に関わっている」と主張する。つまり、本来、人間にとって便利に機能すべく設計された人工物によって、我々の行動や、更には、考え方自身が影響を受けているのだ。

このような流れから、図2-2-8では、人間と事物を区別し、人間中心主義に基づく考え方 (太い点線の左上部) から、人間と事物の相互作用による関係性に重点を置き、創造することで理解する視点の重要性を説く考え方やそれが社会に影響を及ぼすことの重要性を指摘する考え方 (太い点線の右下部) への移行が見取れる。これは、

## 2.2

### 俯瞰区分と研究開発領域 ロボティクス

1. 深層学習に代表される機械学習によって、人工物が判断し、意思決定にコミットしており、あるレベルの機械の自律性が、すでに実現されつつあること、
2. 人間固有と思われていた自由意志や意識なるものの構造や機構などが、神経科学・生理学・認知科学などで徐々に明らかにされつつあること

を考えると、近代諸学問における意識や自律に対する考え方が機能しなくなっている。このような背景から、機械の自律性の可能性について議論可能になりつつある。そこで、以下では、人との相互作用における自律性の意味を物理的・心的側面から眺めてみる。

### (5) 物理的相互作用の設計

身体性はロボット学においてもっとも際立った特徴の一つである。メルロー＝ポンティが示唆する情報構造は人と人との間を想定したものであったが、人工物との関係においても成り立つと考えられる。しかしながら、従来の金属製の硬い身体を有し、電動モーターで駆動されるロボットは、目標軌道を達成するために、人との相互作用のダイナミクスをノイズ要因として扱い、それを消し去る方向で開発されてきた。目標の達成をソフトウェア的に実現できたとしても、本来の物理的な硬さは変わらない。これに対し、近年ソフトロボティクスが興隆し、全世界的に研究が活発である。従来、ソフトロボットは、その柔らかさ故、軌道制御が難しく、制御対象としては嫌われていたが、逆に豊かな相互作用による時空間的にリッチなセンサー情報が獲得され<sup>21)</sup>、それに応じて心的相互作用にも影響を及ぼすことが可能になる。柔らかなペットボトルを掴む事例から、硬いロボットとソフトロボットの違いを示す(図2-2-9)。左側の通常の硬いロボットハンドだと静的な点接触のみだが、右のソフトロボットハンドであれば、面接触で掴み直すなど、識別に有利なリッチな時空間センサー情報が得られる。

ソフトロボットの持つ潜在的な意味や価値については、2.2.1ソフトロボティクスや2.2.2生物規範型ロボティクスに詳しく述べられているが、物理的な相互作用の観点から要約すると以下のようにまとめられる。

1. 物理的な柔らかさが、安全面を向上させ、安心感や信頼感に繋がる。
2. 軌道制御が困難なのではなく、軌道制御自体が意味を持たず、むしろ柔軟な行動や行為の単位での相互作用が意味を持つ。

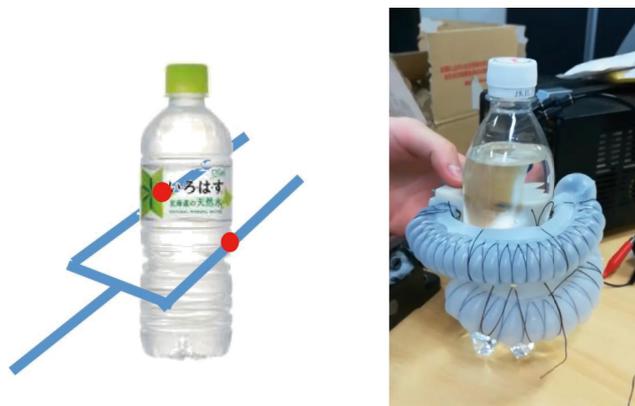


図 2-2-9 従来の硬いロボットハンド (左) とソフトロボットハンド (右: Kawasetsu et al. (in prep.))

3. 多様で柔軟な物理的相互作用が、さまざまな心象を誘発し、そのことが心的相互作用の設計論に影響を与える（例えば、文献 22）など。
4. 生体親和性の高い材料を用いることで、その経時的変化による劣化も含めて、成長・老化するソフトロボットの可能性が示され、このことと3を考慮することで、心的相互作用の設計論を含めて、人間に寄り添い馴染んでくれるドラえもんのようなロボットの設計に繋がると期待される。

## (6) 心的相互作用の設計

ロボットの心的機能の設計を検討するとき自律性は大きな課題である。自律性や主体感、共感からモラル創発の可能性、さらに人工モラルマシンに関する議論が重要である<sup>3), 7), 8)</sup>。

### (6-1) 人と機械の自律性と主体感

自律性に関して、人のみが自律性を有し、機械には自律性が無いと主張する考えがある<sup>23)</sup>。それは、ロボットが生物学的な意味合いにおける自律性を持っていないことによるだけでなく、人が自身で意思決定する際と比較して、ロボット自身が「自身」という表現をもっておらず、さらにロボット自身が自らが意思決定のモジュールであるという明確な確信を持っていないからである。しかしながら、人においても、自由意志と言うのは、空想に過ぎず、さまざまな過程から、意識レベルがつつま合わせを行っているに過ぎないということを示唆する結果が神経科学的から得られている<sup>4, 24)</sup>。このように考えると、人工システムには真の自律性がなく、純粋に自律性を持っているのは人だけである、とは言えず、自律性に関する人と人工システムとの違いは、相対的な差に過ぎない様に見える。

1970年代初頭、チリの生物学者ウンベルト・マトウラーナ（Humberto Maturana）とフランシスコ・バレラ（Francisco Varela）は生物の自律性をシステムの観点から考察し、自己創出（再生と自己による境界決定）を主眼としたオートポイエーシス<sup>25)</sup>を提案した。これを根拠に機械に自律性が無いとする主張がなされているが<sup>23)</sup>、マトウラーナとバレラの細胞の自己創出のコンピュータシミュレーション<sup>26)</sup>では、そのオートポイエティック性が示されている。感覚運動の個体レベルから社会システムレベルに上がれば、厳密な生物学的な意味合いでのオートポイエティック性が薄らぎ、より本質的、もしくは機械と共有できる性質としての意味合いでの、オートポイエティック性が引き出せる。また、これを理由にバレラは、フォーコーやメルロー＝ポンティらの思想に傾倒したと述べている。そして、相互作用（コミュニケーション）の重要性を示し、自律性の規範の更新を謳っている。

自律性と並んで、重要な概念として「自己」がある。自他認知を含む自己の概念の進化は、ウルリック・ナイサー（Ulric Neisser）<sup>27)</sup>が提唱する自己知識の五つの視点が参考になる。生態学的自己、对人的自己、想起的自己、私的自己、概念的自己のうち、最初の二つをそのまま借用し、残り三つを社会的自己としてまとめて、個体発生、すなわち発達過程とみなして、浅田がまとめている（文献5の第三章図3.2）。

Seth et al.は、当時の神経科学や精神医学の知見や証拠をとりまとめ、自己主体感を作るネットワークと自己存在感をつくるネットワークの相互作用モデルを提案した<sup>28)</sup>。自己存在感は、有益な内受容性予測信号が入力と正常に一致し、予測誤差が抑制されると生じると主張する。外受容信号と内受容信号によって内受容感覚を予測する。自己主体感のモジュールが感覚運動予測と同時に内受容感覚も生成する。自己主体感の

4 行動レベルでもマニュアル人間は多く観られ、自律性に乏しく見える。

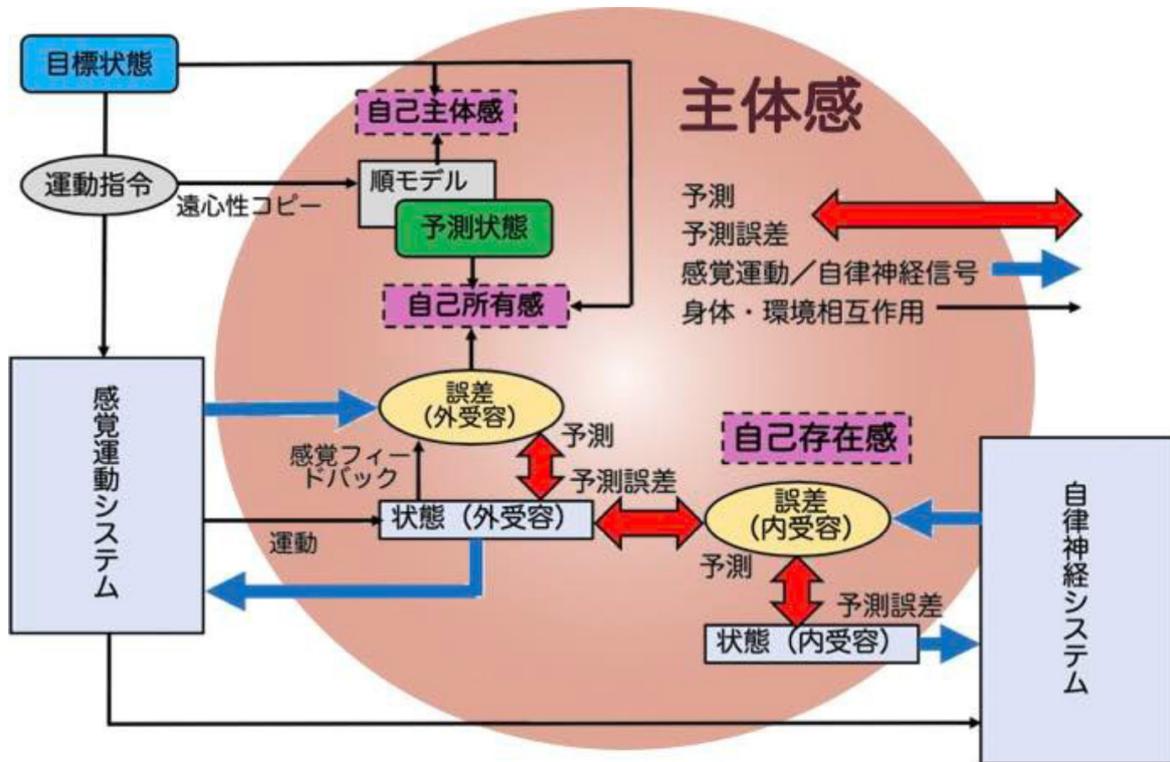


図 2-2-10 広義の自己主体感の構図

モジュールが自己存在感のモジュールより階層が上とされている。ロボットの場合、内受容からの情報によって駆動される自律神経系を自己主体感、自己存在感、自己所有感などにどのように結びつけるかが課題である。

Haggard<sup>28)</sup> は、Gallagher<sup>30)</sup> の2つの様相、すなわち主体感と所有感は、予測と後付で説明可能で、その2つをあわせて主体感 (SoA) と呼んだ。図2-2-10にGallagher<sup>30)</sup>、Seth et al.<sup>28)</sup>、Haggard<sup>29)</sup> の主張を一つの図にまとめた。Legaspi et al.<sup>31)</sup> は、Haggard<sup>29)</sup> のヒトに関する主体感の認知神経科学的なレビューを参考に人工システムにも適用可能な主体感のあり方を議論している。Haggardの論文では扱っていない人間とロボットの相互作用局面における人側の主体感のモデルを提案している。Gallagher<sup>30)</sup> のモデルは比較器モデル (comparator model) として捉えられているが、そこに遡及的推論 (retrospective inference) を加え、拡張している。

### (6-2) 共感からモラル創発の可能性

ロボットの共感からモラル創発の可能性として、痛覚の重要性が示唆されている<sup>32)</sup>。意識 (心) 創発の要が痛覚であることを示すための作業仮説として、

- ① ロボットが痛みを感じるように痛覚神経回路を埋め込む。
- ② MNSの発達を通じて、ロボットは他者の痛みを感じるかもしれない。
- ③ すなわち、情動感染、情動的共感、認知的共感、同情、哀れみの感情をロボットが発達させる？
- ④ 道徳の原型が創発される。
- ⑤ ロボットが道徳 (被) 行為者になる。

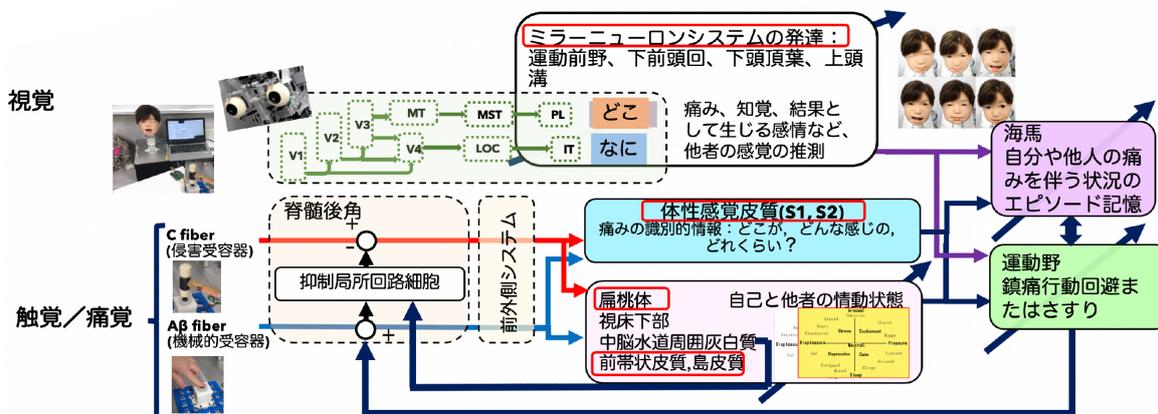


図 2-2-11 痛み体験の共有：共感の始まり

⑥ロボットやAIに対する法制度が検討される。

といった流れを考え、①に対しては、柔軟触覚センサー<sup>33)</sup>を用いて、痛覚と通常の触覚の区別可能性を示している。②に関しては、計算モデル<sup>34)</sup>を提案しているが、③<sup>35)</sup>以降に関しては、概念モデルのみである。図2-2-11は、自身の痛み体験を基に、他者の同様の状況を観察したときに励起される自身の経験の想起と回避行動学習の概念図である。

(6-3) 人工モラルマシンの可能性

人工モラルマシンの可能性を検討する上で、人自身のモラルの成り立ちについて、ジョシュア・グリーン (Joshua Greene) の「モラル・トライブズ」<sup>36)</sup>を参考にする。若くしてハーバード大学の心理学の教授になったジョシュア・グリーンは、実験心理学者、神経科学者であると同時に哲学者でもある。先に示したカントの義務論的道德論と並んで、功利主義に基づく道徳哲学を説き、有名なトロッコ問題を題材に、ポイント切り替えによる5人の救助と力づくで人を落下させての5人の救助 (いずれも一人の死を伴う) の例で、前者が道徳的に許されるのに後者が許されないとされる理由を基本問題として、これにまつわるモラルの課題を広く深く解説している<sup>37)</sup>。詳細を省くと、モラル発動には2つのモードがあり、自然に (無意識に)、それ故、速い応答として振る舞うオートモードと論理的思考を経て、認知的な判断をくだすマニュアルモードである。前者は、自身が所属する集団内のルールとして定着しており、そのため、自動的に反応するのに対し、後者は、他の集団との対応の中で、必要とされるとしている。宗教や文化が集団内の規律であるとする、他宗教や他の文化に触れるとき、オートモードでは軋轢が生じ、マニュアルモードにしないといけぬ。先のトロッコ問題では、力づくで一人を落下させることによる5人を救助することは、直接的かつ意図的であるためにオートモードでの対応であり、ポイント切り替えによる一人の死は5人の救助におけるは副次的なものでありマニュアルモードの対応だと解釈されている。問題は、過去の公民権運動や女性の参政権など、既存の価値観 (主に宗教) や文化に依存すること、すなわちオートモードでの判断ばかりしていると、反対することが自然に映るが、マニュアルモードで論理的に判断することによって初めて他の集団の価値観や文化を認めて、合意に至ることができる。現代でもオートモードとマニュアルモードの葛藤による多くの課題が存在する。

さて、本題の人工モラルマシンの設計課題に移る。ウェンデル・ウォラックとコリン・アレン (Wendell

Wallach and Colin Allen) の人工モラルマシンの著書<sup>38)</sup>を参考にその可能性について議論する。本書は、10年以上前に出版されながら、その序論において展開される世界の状況予測は、まさしく現在の社会が抱えているAIの最新テクノロジーが及ぼす社会的状況を的確に表している。2010年5月に起きた株価相場の瞬間的急変動のフラッシュクラッシュなどがその典型例である<sup>39)</sup>。

ウォラックとアレンは、人工的道的行為者 (AMAs: Artificial Moral Agents) をどのように設計可能かについて、既存の関連技術を踏まえ、トップダウンとボトムアップの両面から議論している。結論から言えば、完全な人間が存在しないと同様に完璧なAMAも存在しない。しかし全く出来ないわけでもない。倫理コードをトップダウンに埋め込んでも、すぐさまフレーム問題に突き当たる。もちろん、人間もフレーム問題を解いている訳ではないが、ある程度の融通性をもって対処している。ただし、文化や宗教の違いなどから、どのような行動を (行動選択) どのような時に (ある行動を起こすトリガーのしきい値) とするかといった道的行動の選択基準が異なり、人間とて、容易ではなく、経験と学習に依存する。まさしくオートモードとマニュアルモードの葛藤である。

Asada<sup>31)</sup> は、痛覚のポテンシャルを見出している。すなわち、痛みの記憶による共感の可能性である。痛覚神経系を内包する人工エージェントが物理的な痛みの経験から、ミラーニューロンシステムの発達を通じて (例えば、Nagai et al.<sup>40)</sup>, Kawai et al.<sup>41)</sup> の研究参照)、他者の痛みを推定し、共感できる可能性を指摘している。その際、痛覚のみならず、社会的な環境を通じて多様な情動の理解と表出 (例えば、Watanabe et al.<sup>42)</sup> の研究参照) が可能になり、結果として道的行動の創発が期待される。ただし、このままだとオートモードの再現に留まる可能性がある。アシモフのロボット三原則は、大きな矛盾をはらみつつも、ロボットの倫理則とも言われている。矛盾は人間社会そのものの矛盾の鏡像であり、その意味では、人間の倫理則といえなくもない。瀬名もそれを指摘している<sup>4)</sup>。痛み回路は、第三原則の必要条件と察せられるし、社会的相互作用を可能にする機能は第一、二原則の遵守に必要な要素と考えられる。さて、これでマニュアルモードも再現できるだろうか? 経験する社会的環境が、狭いとオートモードに陥り、広くとることでマニュアルモードの獲得も可能だが、一体のロボットが物理的に取りうる空間は限られるので、論理的思考による認知的判断は、ネットを通じた広域的な知識や事実の蓄積に委ねることもできるだろう。再現可能なマニュアルモードは、その実現手法が人と異なっても、モードによる判断結果が同じになれば、共生社会の規範になり、ロボットを道的行為者とみなせるだろう。もし、ロボットを道的行為者とみなすことの合意が得られれば、それは同時に道的被行為者とみなすことを意味するだろうか? ロボットなので、人間を助けることは、当然だが、人間からロボットに対して道的行為を授ける必要があるか? という議論になろう。最終的には、法制度の問題も絡むため次節で扱う。

## (7) ロボットと法制度

ここまで機械の自律性と主体感について述べた。これらは、事故や犯罪にお機械ける責任帰属に関して、重要な要件である。Haggardはヒトの主体性に関する論文<sup>29)</sup>の中のコラムで、現在の法制度が神経科学の知見と整合していないことを以下のように指摘している。

- ・すべての法制度には、刑事責任の概念があり、この概念は、健全な成人は行動の自発的な制御を支える彼ら自身の意図、行動、および結果についての意識的な経験と知識を持っているという仮定に基づく。ローマ法体系では、過失故意という用語を使って、責任の主観的で意図的な側面を説明している。したがって、原則として、行動に対する主体性の欠如した者は、その行動に対して刑事責任を問われることは

ない。仮に被告が自分たちのしていることを経験しなかったと主張する場合、裁判所は当然懐疑的である。意図と主体感の客観的な証明は困難である。自発的として経験される行動の前兆として、自動的かつ無意識の過程が存在することが神経科学的には明らかだが、故意の法的概念、すなわち、自発的行動の起点としての犯意が存在することを前提としている点とは好対照である。

- ・しかしながら、法は一部、行動の自発的制御や主体感が減少する状況を認めていて、例えば、長期にわたる虐待が、認知能力と行動能力に重大な変化を引き起こす可能性があり、そのような場合の殺人は、脅威に対する「戦う」反応に類似していると見なされる可能性がある。当然、長期にわたる虐待が、大脳辺縁系と運動の相互作用を変化させ、行動の自発的制御に影響を与えることが神経科学的に示されつつある。
- ・英国法では、長期にわたる虐待行動が、自発的な行動管理とその責任に対して、通常の条件を満たさない可能性があることが認識されている。ストレスを制御した経験のあるラットは、その後の制御不能なストレスの影響に対する回復力を獲得しているが、持続的な虐待は、その回復力を阻止する可能性がある。よって、虐待の期間がトリガーとなり行動制御不能に陥ると見なすための神経基盤を提供する可能性があることは激しい議論の的である。

現在の法制度と最新の神経科学の知見との不整合は非常に複雑な問題であるが、ここに人工物が介在した事故や事件が発生したときには、法制度のあり方がさらに解決困難な課題として表出する。この課題に対し、稲谷はポストヒューマニズムの観点から、法のあり方を論じている<sup>13)</sup>。(4)でも紹介した人間と機械の相互作用に重点をおいた考え方をベースに、人間自体の再定義、よき生のあり方を追求する過程における法のあり方を論じ、現在の法制度に挑んでいる。要点は以下である。

- ・最先端技術が搭載された人工物が人間と機械の境界を曖昧化し、結果的に意思決定や行為の主体を特定することが困難な状況を作っている。
- ・外的環境から文字通りに独立して意思決定し、また行為することのできる主体としての人間は、存在しない。意思決定や行為の主体は、外的環境との相互作用の場として立ち現れる。人間は主体でも客体でもない。
- ・従来の刑事法は、近代を基礎付ける心身二元論と主客二分論というアイデアに立脚し、主体の自由意志に基づく行為に対して刑事制裁を科すことで、望ましくない行為を抑止して、社会秩序を維持するないし法益を保護するという構図で、あるべき主体と客体の線引きを確保するべく、あらゆる存在を多少強引にでもどちらかに押し込む(規律・訓練によって〈他者〉を〈標準化〉して人間にする)。
- ・従来の刑事法を新たな状況に適用を考えると、以下の状況

1. 神経科学的に自由意志が存在しないユーザーに責任を押し付けるのは意味がない。
2. 製造者側に責任を求めると、萎縮が生じて、イノベーションが生じない。

- ・よって、人間を所与の前提とする思考法を捨てる必要(ポストヒューマニズム)があり、人とその構成に関わる外的環境のありようを、反省的にデザインしていくことのできる法理論が必要。
- ・法理論の前提となる倫理について、ポストヒューマニズムの観点からは、〈どう生きたいか〉という問いを広く直接的に問い続けることで、人の存在態様の望ましきについての批判的吟味を絶えず行い、より良

## 2.2

### 俯瞰区分と研究開発領域 ロボティクス

い人の存在態様とそれを実現する方法とについて探究する営みとして徳倫理を推奨。

- ・ポスト・ヒューマニズムにおける法とは、自己の練り上げに呼応しながら、自らの姿を変えていく可能性を留保した、開かれた動態的運動としての法自己を練り上げ、自身と他者（人と人以外の存在とを含む）とを導く理の定立を通じて、現実の新たなありようを示し、法の言説を利用することで、これを法と結びつけ、現実を変化させようとする実践（自律実践）は、法自身を一層動態的で開かれた運動へと変えていくことになるのである。

そして、公私協働による法秩序形成が必要と結言しており、これに関しては、ガバナンスイノベーションとして、すべてのステークホルダーが参画して、決めていく、変えていく体制を提案し<sup>5</sup>、具現化に努めている。

ロボット学の観点からも稲谷の考え方は重要である。（3）で提案したように、様々な学問分野が分野ごとの研究のみならず、その成果が社会にどのような影響を及ぼしうるか、その過程も含めて、科学技術の進展とともに、パラダイムを常時更新していく姿勢を可視化し、多様な人間社会での規範の構築と相互理解を通じて、共生社会を設計していく必要がある。

## 参考文献

- 1) 新しい学術の体系— 社会のための学術と文理の融合 —日本学術会議運営審議会附置新しい学術体系委員会、<http://www.scj.go.jp/ja/info/kohyo/18pdf/1829.pdf>
- 2) 浅田, 石黒, 國吉. 認知ロボティクスの目指すもの. 日本ロボット学会誌, Vol. 17, No. 1, pp.2-6, 1999.
- 3) 浅田稔. 講演報告「再考（最高?）：ロボット学」. 日本ロボット学会誌, Vol. 39, No. 1, p. to appear, 2021.
- 4) 瀬名秀明. 講演報告「ロボティクスと総合知」. 日本ロボット学会誌, Vol. 39, No. 1, p. to appear, 2021.
- 5) 浅田稔. 浅田稔のAI 研究道. 近代科学社, 東京, 2020.
- 6) Minoru Asada, Koh Hosoda, Yasuo Kuniyoshi, Hiroshi Ishiguro, Toshio Inui, Yuichiro Yoshikawa, Masaki Ogino, and Chisato Yoshida. Cognitive developmental robotics: a survey. IEEE Transactions on Autonomous Mental Development, Vol. 1, No. 1, pp. 12-34, 2009.
- 7) 浅田稔. なじみ社会構築にむけて：人工痛覚がもたらす共感、道徳、そして倫理. 日本ロボット学会誌, Vol. 37, No. 4, pp. 287-292, May 2019.
- 8) 浅田稔. 再考：人とロボットの自律性. 日本ロボット学会誌, Vol. 38, No. 1, pp. 7-12, January 2020.
- 9) Jun Tani. *Exploring Robotic Minds: Actions, Symbols, and Consciousness as Self-Organizing Dynamic Phenomena*. Oxford University Press, 2016.
- 10) 稲谷龍彦. 技術の道徳性と刑事法規制. 松尾陽（編）, 『アーキテクチャと法』, 第4章, pp. 93-128. 弘文堂, 2017.
- 11) 牧野英二編. 『新・カント読本』. 法政大学出版局, 2018.
- 12) ジュリア・アナス（著）, 相澤康隆（訳）. 徳は知なり. 春秋社, 東京, 2019.
- 13) 稲谷龍彦. ポスト・ヒューマニズムにおける刑事責任s. 宇佐美誠（編）, 『AI で変わる法と社会：新しい

5 <https://www.meti.go.jp/press/2020/07/20200713001/20200713001-1.pdf>

人間の条件』, 第 6 章, pp. 113–136. 岩波書店, 2020.

- 14) フッサール 著, 浜渦辰二 訳.『デカルト的省察』. 岩波文庫, 2001.
- 15) マルティン・ハイデッガー, 細谷貞雄 訳.『存在と時間 〈上〉』. ちくま学芸文庫, 1994.
- 16) マルティン・ハイデッガー, 細谷貞雄 訳.『存在と時間 〈下〉』. ちくま学芸文庫, 1994.
- 17) M. メルロー=ポンティ, 竹内芳郎・小木貞孝 訳.『知覚の現象学 1』. みすず書房, 1967.
- 18) M. メルロー=ポンティ, 竹内芳郎・木田元・宮本 忠雄 訳.『知覚の現象学 2』. みすず書房, 1974.
- 19) ブルーノ・ラトゥール, 川崎 勝 平川秀幸 訳.『科学論の实在—パンドラの希望』. 産業図書, 2007.
- 20) ピーター=ポール フェルベーク 著, 鈴木 俊洋 訳.『技術の道德化: 事物の道德性を理解し設計する』. 法政大学出版局, 2015.
- 21) 細田耕.『柔らかかヒューマノイド』. 化学同人, 2016.
- 22) Hisashi Ishihara, Binyi Wu, and Minoru Asada. Identification and evaluation of the face system of a child android robot affetto for surface motion design. *Frontiers in Robotics and AI*, Vol. 5, p. 119, 2018.
- 23) 河島 茂生編著.『AI 時代の「自律性」』. 勁草書房, 2019.
- 24) スタニスラス・ドゥアンヌ, 高橋 洋=訳.『意識と脳—思考はいかにコード化されるか』. 紀伊國屋書店, 2015.
- 25) 河本英夫.『オートポイエーシス』. 青土社, 1995.
- 26) F. J. Varela, H. R. Maturana, and R. Uribe. Autopoiesis: The organization of living systems, its characterization and a model. *BioSystems*, Vol. 5, pp. 187–196, 1974.
- 27) Ulric Neisser. The self perceived, pp. 3—22. *Emory Symposia in Cognition*. Cambridge University Press, 1994.
- 28) Anil Seth, Keisuke Suzuki, and Hugo Critchley. An interoceptive predictive coding model of conscious presence. *Frontiers in Psychology*, Vol. 2, p. 395, 2012.
- 29) P. Haggard. Sense of agency in the human brain. *Nature Reviews Neuroscience*, Vol. 18, pp. 196–207, 2017.
- 30) Shaun Gallagher. Philosophical conceptions of the self: implications for cognitive science. *Trends in Cognitive Sciences*, Vol. 4, No. 1, pp. 14 – 21, 2000.
- 31) Roberto Legaspi, Zhengqi He, and Taro Toyoizumi. Synthetic agency: sense of agency in artificial intelligence. *Current Opinion in Behavioral Sciences*, Vol. 29, pp. 84 – 90, 2019. SI: 29: Artificial Intelligence (2019) .
- 32) Minoru Asada. Artificial pain may induce empathy, morality, and ethics in the conscious mind of robots. *Philosophies*, Vol. 4, pp. 38–47, 2019.
- 33) T. Kawasetsu, T. Horii, H. Ishihara, and M. Asada. Flexible tri-axis tactile sensor using spiral inductor and magnetorheological elastomer. *IEEE Sensors Journal*, Vol. 18, No. 4, pp. 5834–5841, 2018.
- 34) Jorge L. Copete, Yukie Nagai, and Minoru Asada. Motor development facilitates the prediction of others' actions through sensorimotor predictive learning. In *Proceedings of the 6th IEEE International Conference on Development and Learning, and Epigenetic Robotics (ICDL-EpiRob 2016)*, pp. (CD-ROM) , 2016.

- 35) Minoru Asada. Towards artificial empathy. *International Journal of Social Robotics*, Vol. 7, pp. 19–33, 2015.
- 36) ジョシュア・グリーン (著), 竹田円 (訳). モラル・トライブズ-共存の道德哲学へ- (上・下). 岩波書店, 東京, 2015.
- 37) 笠木雅史. 自動運転の応用倫理学現状と課題: 自動運転車とトロリー問題. 日本ロボット学会誌, Vol. 39, No. 1, p. to appear, 2021.
- 38) Wendell Wallach and Colin Allen. *Moral Machines: Teaching Robots Right from Wrong*. Oxford University Press, 2008.
- 39) ウェンデル・ウォラック (著), コリン・アレン (著), 岡本 慎平 (翻訳), 久木田 水生 (翻訳). ロボットに倫理を教える—モラル・マシーナー—. 名古屋大学出版会, 2019.
- 40) Yukie Nagai, Yuji Kawai, and Minoru Asada. Emergence of mirror neuron system: Immature vision leads to self-other correspondence. In *IEEE International Conference on Development and Learning, and Epigenetic Robotics (ICDL-EpiRob 2011)*, pp. CD-ROM, 2011.
- 41) Yuji Kawai, Yukie Nagai, and Minoru Asada. Perceptual development triggered by its self-organization in cognitive learning. In *In Proceedings of the 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 5159–5164, 2012.
- 42) Ayako Watanabe, Masaki Ogino, and Minoru Asada. Mapping facial expression to internal states based on intuitive parenting. *Journal of Robotics and Mechatronics*, Vol. 19, No. 3, pp. 315–323, 2007.