# 「心を持ったロボットをつくる」というプロジェクトはどのようなものでありうるか？[1]

井頭昌彦

## 序

本稿では、まず、「心を持ったロボットを作る」というプロジェクト、およびその中で哲学がなしうる寄与について論じる。その際、心の哲学に関する1つの見解が提示されるとともに、それがどのような正当化構図のもとで妥当性を問われうるかについて検討がなされる。そして、本稿の最後では、哲学研究の――それほど斬新というわけではないが――新たなあり方を示唆する、というメタ哲学的課題にも取り組む。

本稿は、実際に「心を持ったロボットが実際に完成した」という朗報（?）を報告するものではないし、それを実現するための方法を決定的な論証によって示しているわけでもない。つまり、本稿は、何らかの成果を報告する種の論考になっていないという意味で、「試論」的な性格が強いものとなっている。とはいうものの、上記プロジェクトのような取組みに際して留意・検討すべきことの示唆、哲学者が注力すべき課題の提示、融合研究であるがゆ

えに正当化のための議論構造が変化しうる事実への注意喚起など、いくらかでも有益な情報提供をすることができるのではないかと考えている。

## 第一節　心の哲学の役割と「直観的理解」の重要性

心を持ったロボットをつくるには何をしなければならないのか？どんな機能を実装し、どんな外見を持たせ、どういった素材からつくらねばならないのか。こういったタスクに取り組むためには、「心とは何か」「心を持つとはどういうことか」という問いへの取組みを避けることができない。なぜなら、こういったことがわかっていなければ、あるいはそれについて一定の作業仮説がなければ、「心を持ったロボット」の実現のためにつくり込むべき機能の検討を進めることはできないからである（後述するように、この問いは上記タスクをクリアするのに**先立って答えられている**必要は必ずしもないが、少なくともそれと**平行して問われる**べきである、と筆者は考えている）。

さて、この「心とは何か」「心を持つとはどういうことか」という問いは、哲学の領域でながらく問題にされてきた問いでもあった。そして、この点において、本プロジェクトに哲学が寄与しうる側面がある。本節では、まず、この問いに関して二十世紀後半の「心の哲学」において展開されてきた議論の概略史を描き、様々な理論の妥当性評価基準の一側面――具体的には「心についての直観的理解との合致」が持つ重要性――を描き出すことにより、この問いがどういう性質のものであるかについての一つの描像を提示する。これは上記タスクに関して展開される次節以降の議論の準備となる。

デカルト的二元論から始めよう。この立場は「心は物質とは根本的に異なる領域に属す実体である」「心的状態

の正確な内容は一人称的視点からのみアクセス可能である」という二つの考えから構成されるものである。これは一見、自然な理解であるように思われるが、心に関して我々が有している様々な**直観的理解**[2]と折り合いが悪いという難点を持つ。たとえば、我々は通常は他者の心を理解できると考えているし、心的事象は物理的事象を引き起こしうる（心身因果）と考えているが、デカルト的二元論では心に関する我々の理解に深く根ざしたこれらの考えは否定されてしまうのである。

これらの難点を抱えるデカルト的二元論が支持を失った後に登場したのが、心的状態を何らかの行動傾向と同一視する（哲学的）行動主義である。この立場は、心的状態を公共的に観察可能な行動や発話と同一視することによって、デカルト的二元論の難点として指摘された直観的理解の取り込みを可能にした。しかし、他方で、「同じ心的状態でも文脈次第で行動へのあらわれかたが違うことがある」という別の直観的理解をうまく説明できないという難点を指摘され、広く支持されるには至らなかった。

その後、神経科学の発展を背景に行動主義に取ってかわったのが心脳同一説（タイプ同一説）である。これは、心的状態を特定の神経科学的状態タイプと同一視する立場であり、外部から観察できない神経科学的状態を持ち込むことによって、行動主義の難点として指摘された直観的理解を取り込むことに成功した。しかし、他方で、人間と全く同様の神経科学的組成を持たない存在――異星人や簡易サイボーグを含む――には原理的に心の存在を認められなくなる（しかしそれは我々の直観に反する）、といった難点を指摘され、批判されてきた[3]。

こういった経緯を経て登場してきたのが機能主義と解釈主義という二つの立場である。機能主義は、基本的には「心的状態の内容および同一性は感覚刺激・他の心的状態・振る舞いに対してそれがもつ因果関係によって規定される」とする立場であり[4]、他方の解釈主義は「心的状態の内容およびその同一性は第三者からの解釈実践によって

規定される」とする立場である。これらの立場はヒトの神経科学的組成に依存しない形で心的状態を規定することで排他的な見解をとらずに済むため、我々の直観的理解への適合性という点で一定のメリットを持つが、他方でいくつかの難点も指摘されている。たとえば、機能主義（のあるバージョン）に対しては、我々が通常心的状態の保持者と認めないようなもの（ある仕方で組織化されている人間集団など）に対しても心的状態を認めることになってしまう、という点で直観との齟齬が指摘されることがある。また、解釈主義（のあるバージョン）に対しては、心的性質に本来的な因果機能を認めつつ「〜の心的状態にあったことが、〜の物理的出来事を引き起こした」という心身因果を認めることが困難になる、という問題点が指摘される。つまり、現在において重要な検討対象と見なされている機能主義と解釈主義もまた、心に関して我々が有している（別の）直観的理解との不整合性を根拠に批判されているのである。

さて、こういった概略史――かなり偏った、かつ多くの論点を取りこぼした不十分な概略史であることは否定しない――を踏まえた上で指摘できることの一つは、理論の妥当性評価の場面において「我々が心に関して持っている直観的理解」が担っている重要性である。具体的に言えば、それは次のような役割を果たしていると言えるだろう。

- 我々は心に関して様々な直観的理解を持っている。
- これらの直観とうまく整合しない理論は攻撃の対象となる。
- そういった理論は問題の直観と両立する別の理論にしばしば取って代わられる。

このように考えるなら、心の哲学の論争史から一つの理論評価基準として「理論と直観的理解との合致」を抽出することができるだろう[5]。なぜこれが重視されるかというと、たとえば「心についての説明」と称して提示された

理論的定義があるとして、それが心に関する直観的理解から大幅にズレていたり、直観的理解の多くを否定するものであった場合には、その定義を「心についての説明・解明」として理解することができなくなるからである（このことは、たとえば「国家という概念を《カエルの心臓の上部にある突起物》を指す用語として定義する」という主張に対する違和感を想像することで理解されるだろう）。こういった考察を経由することにより、心の哲学における理論構成の（一つの）目的ないし規範理念を「理論と直観的理解との合致」という形で与えることができるかもしれない（図1）。実際、心の哲学において理論の修正や交代が行われる際には、しばしば理論と直観的理解の不整合が指摘されてきたのである。

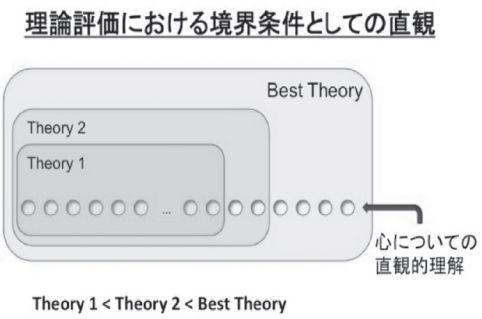


図1.「哲学理論の評価」と直観の関係

しかし、仮に「理論と直観的理解との合致」が目的の一つだとすると、その達成方法を「理論改訂」や「理論交代」といった《理論サイドの修正》に限定する必要はないはずである。ここで念頭に置かれているのは「直観的理解の改訂」という選択肢である。もちろん、心に関する直観的理解は心を巡る我々の実践に深く根ざし、それを「心についての実践」として成立させているものであるから、容易に変えられる／変えてよいものではないが、ここでは「直観的理解は原理的に改訂不可能であるわけではない」ということさえ押さえておけばよいだろう。この点については、「原理的な改訂不能な言明」の存在を否定する（認識論的全体論等の）諸議論を想起するなら[6]、あるいは「鯨は魚である」「地面は動かない」「～は人権を持たない」といった直観的理解が実際に改訂されてきたことを想起するなら、その説得性に関して

一定の理解を得ることができるだろう。

以上で、心に関する我々の直観的理解は（改訂不能性を伴った絶対的立脚点ではないとしても）それとの合致が理論評価の一つの境界条件になっている、という構図が理解された。以下では、この構図を背景とした上で、「心をもったロボットをつくる」というプロジェクトにどういうアプローチが可能であるかを考えてみよう。

## 第二節　一つの切り口としての「痛み」と、言語実践の階層的発展モデル

さて、「心と持ったロボットをつくる」にはどうすればよいか、と実際に考えてみると、これほど取り組みにくい課題もないだろう。一つの大きな問題は、心というものの複雑な諸側面のどこに着目すべきか、というものである。近年では、意識・感覚・情動・思考といった心の諸側面が概念的に区別され、それぞれ異なった特性を持つものとして別個に分析が行われている。さて、ロボットはこれら全ての側面を備えなければ「心」を持ったことにならないのか、それとも一部が実現できればよいのか？あるいは、成人並の検出・生産・操作能力を持たねばならないのか、ごく限られた能力で十分なのか？

こういった問いに思考実験や論証だけによって答えるのは困難であるし（その一つの理由は「ある種の機能を備えたロボットを前にした時の我々の反応」を思考実験のみによって十分に予測することが困難だからである）、それゆえ、何らかの仕方で「心を持つための必要条件」をあらかじめ確定してから、その実現に向けて動き出す、というやり方は有効なものではないように思われる。むしろ、できそうなことからはじめて、成果を確認しながら随時方針や状況理解を修正していく、というやり方のほうが議論の着実な前進が見込まれるであろう。そこで、本稿

では、意識・感覚・情動・思考といった諸側面を全て備えた「フルスペックの心」を一挙に実現することを目指すのではなく、実現への目処が比較的立ちやすい[7]「感覚」、なかでも「痛み」に焦点をあて、「痛みを感じられるロボットをつくる」というダウングレードした形で当座の課題を設定することにしたい。そして、それがある程度実現できてから、その是非（それで心を持ったロボットができたことになるのか）を改めて評価しなおし、場合によっては方針を再考する、というアプローチをとるのである（感覚の中でも「痛み」に焦点を当てる一つの理由は、第三節で述べるように、ロボット等に「心」を認定する際に大きな障壁となる「**チートの可能性**」を乗り越える処方を与える見込みがあるから、というものである[8]）。

しかし、このように課題を絞ってみたとしても、それにむけた取組みが一気に容易なものとなるわけではない。たとえば、「痛みを感じられるロボットをつくる」という課題を遂行するために必要な「痛みセンサ」を開発する、という場面を考えてみよう[9]。しかし、そこで作製されるべきセンサはヒトの侵害受容器の検出能力を完全に模倣する必要があるのか（ヒトの痛みセンサは機械刺激・温度刺激・化学刺激など多様な刺激を検知できる）、それともその一部を検出できればよいのか？あるいは、検出能力の問題が解消されたとしても、そのセンサを取り付けただけで「痛みを感じられる」ようになったと言えるのか？それとも、さらなる追加メカニズムを備える必要があるのか？仮に追加メカニズムが必要だとして、どのような機構をどのような理由で実装する必要があるのか？これらは問題の一端に過ぎないが、「心」から「痛み」へと課題をダウングレードしたとしてもなお、回答困難な無数の難題が残されていることは理解できるだろう。

もちろん、「痛みとは何か」ということをあらかじめ規定しておくことができるならば、上記の問いに対して是非の判断を下すことができる。しかし、具体的に考えてみると、「痛みを規定する」ということ自体が容易ならざ

るタスクであることがわかる。おそらく最初に着想されるだろう一つの方向性は「痛み＝特定の神経科学的状態」という規定であるが、この路線には、一人称報告と矛盾が生ずるケースをどう処理するかという問題、あるいは、神経科学的組成を共有しない異星人は原理的に痛みを持てなくなるといった難点[10]が指摘されうる。また、国際疼痛学会（ＩＡＳＰ）の「痛みの定義」に付された注釈において、純粋に神経科学的状態のみによって痛みを定義するのではなく、主観的判断や言語報告の重要性が認められているという点にも留意すべきであろう。

……多くの人々は、組織損傷あるいはそれに相応した神経生理学的原因がないのに、痛みがあると言う。このような経験と組織損傷による経験とは、通常区別できるものではない。もし彼らが、自分の体験を痛みと思い、組織損傷によって生ずる痛みと同じように報告するなら、それを痛みと受け入れるべきである。この定義は、痛みを刺激と結びつけることを避けている。侵害刺激によって侵害受容器及び侵害受容経路に引き起こされる活動が痛みであるのではない[11]……。

あるいは、一人称的な主観的現象として痛みを規定する路線が考えられるかもしれない。「痛いときのあの感じ、あれこそが痛みだ」というわけである。しかし、もし痛みがそういった「主観的感覚」に尽くされているとすると、たとえば「（第三者が外部から覗き込んで見て取ることができないという意味で）私秘的なものを指示するための言葉をどうやって教示するのか」という言語学習上の問題や、「私秘的な感覚の生起メカニズムをどうやって特定し実装するのか」という困難な課題を抱え込む羽目になるだろう。

結局、「痛みとは何か？」という問いにあらかじめ答えを与えておいてから、何を実装すべきかを考える、とい

う路線でうまくやるにはかなりの難題をクリアしなければならないと思われる。こういった困難な状況を打開する、あるいは少しでも前に進めるための一つの手がかりとして、本稿では「痛み」という《語》に着目するというアプローチを検討してみたい。すなわち、「痛みとは何か？」という直接的な問いかけから、「我々は『痛み』という語で何を意味しているか？」という問いへと焦点をずらすのである。

これは一見思われるほど論点をズラしているわけではない。たとえば、机のカドに足をぶつけてうずくまっている山田氏を気遣う次のようなやりとりを考えてみよう――「どうしたの？」「足が痛いんだ」「かわいそうに」。このやりとりにおいては、「痛い」という言葉が意味できる以上の情報は伝達されないだろう。つまり、痛みに関して我々が論じられる事柄――本稿の議論も含む――と「痛み」という語が意味しうる事柄とは密接に関わっているのである。こういった事情を踏まえるならば、「痛みとは何か？」「そのモノに痛みは生じているのか？」といった問いは「痛み」という語が意味しうる情報の範囲内で問題にされるべきだ、という考えに一定の説得力を見て取ることができよう。また、第一節で見たように、ある概念の有り様を解明する哲学的理論の妥当性評価が当該概念に対する我々の直観的理解との適合性によって与えられると考え、かつ、その直観的理解が我々の概念把握に依拠していると考えるならば、「痛み」という語の用いられ方およびそれによって表現される概念把握に焦点を当てるこのアプローチが、痛みに関する理論構築に際して重要な意義を持つことは容易に理解されるだろう。いずれにせよ、以下ではこの路線で「痛みロボット」プロジェクトに対してどこまで光を投げかけることができるか、検討してみることにする。(12)

さて、「我々は痛みという語で何を意味しているか？」という問題への取組みにあたっては、「学習可能性」という観点が一つの手がかりになる。我々が「痛み」という語を使用し、それを用いて他者とコミュニケーションをと

れるようになるためには、我々はその語の使い方を学ばねばならない。「痛み」という語を用いる実践に参入する際に、学習者はどのような情報に依拠し、どのような事柄と対応づける仕方でこの語の使い方を学ぶのか。こういったことを概観することによって、「我々は痛みという語で何を意味しているか?」という目下の問題に答えるための手がかりが得られるだろう。

言語学習に関してまず押さえるべき事柄は、教示者および学習者が共に参照する情報は公共的・間主観的に確認できるものだ、ということである。この事情は（一見私秘的なものを指すために用いられると思われる）心的な語彙についても同様である。たとえば、子供の手に針が刺さっているのを確認して「痛かったねえ」と使用例を示す場合や、犬の手に刺さったトゲを指し示しながら「ワンワン、痛いね」と使用例を示す場合、教示者は子供や犬の心の中を覗き込んで主観的な《痛み》が生じていることを確認してから教示を行っているわけではないだろう。むしろ、「どういう場面で『痛い』という語を用いるべきか」に関する公共的状況を参照しながら教示を行っているのである。

次に、通常は心的状態に関する報告には一人称特権があると考えられているが、言語習得期においては心的な状態に関する一人称報告であっても訂正されることがある、という点も指摘しておくべきだろう。上で述べたように、「痛い」という語の使用法の教示において参照されるのは間主観的にアクセス可能な状況であるが、そういった教示が完了し「痛い」という語を習得したと見なされるのは、そういった適切な状況で（のみ）その語を用いることができるようになった段階においてである。逆に言えば、学習の途上にある者が不適切な状況下で——たとえばくすぐられているときに——「痛い」という言葉を用いた場合には、それが一人称報告の体裁をとっていたとしても、誤用として訂正の対象となりうるのである（「『痛い』じゃなくて『くすぐったい』でしょ」など）。

ただし、間主観的状況の重要性を強調することは、もちろん、第三者から観察されない主観的感覚なるものの存在を一切否定することではない。もし、内観によって把握される主観的感覚のようなものが存在しないなら、痛みが典型的に認定される公共的状況を参照しなければ痛みの報告がなしえない、ということになりかねないが、それは明らかに行き過ぎだろう。というのも、実際には、我々はそういった公共的状況を参照することなく痛みの報告をすることがあるからである（不意に痛みを感じて痛みを表出した後に、手にトゲが刺さっているのを確認する等）。したがって、間主観的状況の重要性を認めつつ、公共的状況を参照しない内観ベースの報告が成立することも認められるような説明がもとめられるのだが、それは概略的には次のようなものになると思われる。たしかに、「痛み」という語が学習される際に教示者と学習者がともにアクセスできる参照情報は間主観的なものであるが、学習者自身が痛みの主体であるときには、そういった状況の成立と平行して生じている主観的に参照可能なデータ・内部センサ状態などがありうるだろう。そして、子供が「痛み」という語の学習を完遂し、適切な状況で「痛み」という語を使えるようになった後でなら、そういった状況で典型的に生ずる内観データ・内部センサ状態をカテゴライズすることが可能になるのであり、その結果として、この種の（痛み主体のみがアクセスできる）主観的情報を介した――そして公共的状況を参照しない形での――権威ある痛みの報告が可能になるだろう。つまり、痛みの第一次的な同定基準を公共的状況と見なすことと、「主観的情報のみに基づいた痛み報告」の可能性を認めることとは、十分に両立可能なのである。

これまでに述べられてきた描像[13]をいったんまとめておくならば、以下のようになるだろう。まず、言語学習は教示者と学習者が共有できる情報を用いて行われる。そして、この事情は心的な語彙に関しても同様であり、心的語彙を学習する際に参照されるカギ情報は主観的感覚ではない。その一方で、一度当該語彙の使い方が習得されたな

ら、同時発生していた主観的情報の方を用いて一人称報告が可能になる、というわけである。

ここで重要なのは、「痛み」という語の習得が進んで行く過程で、この語の用法を規定する使用条件が追加されたり「軸足」がシフトしたりすることがある、ということである。「痛み」という語に関する第一次的な言語実践では「ある特定の間主観的状況のもとでこの語を用いること」が適切にこの語を使用するための条件であった。しかし、この言語実践を十分に習得した者にとっては、そういった条件下において頻繁に検知される主観的感覚・内部センサ状態等との対応づけを介することで、間主観的状況を参照しない「主観的感覚に基づいた痛み報告」という言語実践が可能になり、さらに適切な間主観的状況がない場合での一人称報告がなされた場合に後者が優先されるようになると、今度は「一人称特権を持った痛みの報告」が成立するようになるのである。つまり、ここでは、「痛み」という語に関する言語実践が進展する中で、この語の使用条件の軸足が「公共的状況」から「主観的感覚」へとシフトしているのである。実際、IASPの定義において、痛みの原因が確認できない状況下であっても一人称報告に定位して痛みを認定する可能性が示唆されていることを見れば、（学習の過渡期には十分に成立していなかった）一人称報告優先型の言語実践の存在は認めざるをえないであろうし、それゆえ「軸足のシフト」によって言語実践の変化を説明するというこの路線にも意義が認められるだろう。

さらに、こういった使用基準に関する「軸足のシフト」という考えを援用することで、痛みが神経科学的に定義されるものとして語られる言語実践の成立経緯についても、大まかなイメージを描くことができるかもしれない。たとえば、しばしば脳機能イメージング研究においてなされるように、痛みの生起を被験者に報告させ、その際の脳状態を撮像することにより、痛みが生じているときに特有の活性化パターンが同定された後、次第にそういった活性化パターンが「痛みの生起の認定基準」として用いられるようになる、という一連の流れを想定してみよう。

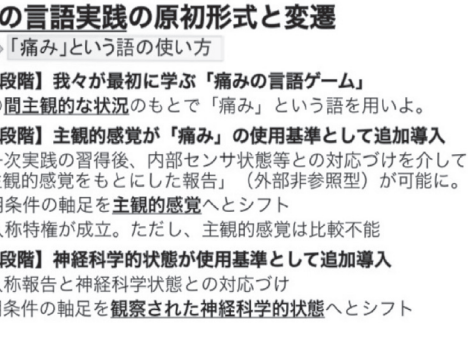


図２．言語実践の階層的発達

ここで創出された新たな言語実践においては、仮に適切な間主観的状況が不在でありかつ痛みの一人称報告がない場合であっても、痛みの生起が認定されうるだろう。つまり、このケースでは、上述のケースと同様に、おおむね共変関係にある情報を利用し、使用基準の軸足を切りかえる、という作業がなされているのである。

ここで示されたのは、間主観的状況を参照しながら痛み認定を行う第一次的な言語実践が「痛みの言語実践系列への参入ゲート」[14]として機能しつつも、当初の認定基準と概ね共変関係にある別の情報を同定し、そちらに使用条件の軸足をシフトさせていくことで、少しずつ異なった基準で「痛み」という語を用いる多様な言語実践系列が生み出されて行く、という階層的発達モデルである（図２）。極めて大雑把なものであることを認めつつも[15]、このモデルが「痛み」に関する異なった言語実践のパターンを区別し、その相互の発展関係について（それほど的外れではない形で）意味ある描像を与えることができていると考えるならば、これに沿って以下のような仕方で「痛み」に関する様々な見解を分類できるだろう。

1. 「公共的情報に基づく痛みの帰属」という第一段階の言語実践に定位し、痛みが解釈可能性から独立に自存することを否定する解釈主義的アプローチ。
2. 一人称権威を十全な仕方で認める第二段階の言語実践が痛みの本質を捉えているとする（デカルト的二元

論に類似した）内観主義的アプローチ。

3. 第三段階の神経科学的定義をもって「痛み」に関する事実が尽くされるとする立場。タイプ同一説的、あるいは機能主義的なアプローチ。

4. いずれの言語実践パターンにも特権性を認めず、場面ごとに異なる適用を認める複数主義的プラグマティズム[16]。

筆者個人としては、存在論的負担（＝それと両立しうる形で整合的な「世界」概念を描き出すことの困難さ）が大きくなることを認めつつも4の立場を支持することに魅力を感じているが、目下検討中のプロジェクトに関しては、「いずれの立場が正しいかをあらかじめ議論によって決定してから、それに従ってロボットに実装すべき機能を決める」というアプローチではなく、「一番実現しやすそうな立場に定位し、それが正しいと仮定した上で機能設計・実装を試み、実機の完成度がある程度上がって様々な実験的テストに供することができるようになってから、それらのテスト結果も踏まえた上で元々仮定されていた立場の妥当性を再考する」というアプローチをとることにしたい（このようなアプローチの採用は、哲学的考察を進めるという観点からは後退しているような印象を与えるかもしれないが、私見では、このようなやり方でも十分に哲学的議論に貢献することができるし、自明な直観に基づく思考実験のみに依拠した議論が停滞してしまっている状況を打開するには、むしろ有効な一手段である、と考えている）。

さて、実現が一番容易な形で「痛み」を描いている言語実践はどれだろうか。2については一人称的認知を実装することの困難さ（そもそも「主観的な現れ」とは何かがよくわかっていない）から、また、3については刺激のインプット経路はともかく内部処理やアウトプットへと至るメカニズムについての解明が十分に進んでいないこと

や痛みという事象に相当する範囲を確定することの困難さから、作業仮説としての採用を見送ることにしたい。他方、残された選択肢である1については、痛みの生起に関連する状況について我々が一定の知識を有していること、少なくとも人間に痛みが生ずる典型的なケースについては具体的な状況特定が可能であることから、利用できる情報の多さに鑑みて、まずはこのルートに定位して機能設計・実装を試みることとしたい。

## 第三節　チートの可能性（あるいは「擬人化の壁」）

さて、最初はどうアプローチしてよいかわからなかった「心を持ったロボットをつくる」というプロジェクトは、これまでの議論の中で少しずつ選択肢を絞ることにより、痛み概念を解釈主義的な観点から捉えた上で、それを実装したロボットの作製を目指す、という形で再定式化されるに至った。課題は少しずつ明確になってはきたが、それでもなお、多くの疑問が残されている。本節では、そういった疑問の中から、（1）解釈主義的言語実践に定位して痛みを理解した場合、具体的にはどのような仕組みや振る舞いの機能をロボットにつくり込むことになるのか？（2）解釈主義とはそもそも命題的態度に適用される学説であって、「痛み」には適用できないのではないか？（3）表面なインプット／アウトプット関係のところでは痛みを感じているように**見える**モノを作製できたとしても、「そのように見えるだけで本当は痛みなど感じていないのではないか」という疑いは消えないのではないか？　という三つのものを取りあげ、個別に解消を試みておきたい。

まず、（1）について簡単に触れておこう。たとえば、乳幼児や動物の痛みを認めるのであれば、「典型的に成人に痛みを生じさせるような状況の全て」に対して「成人と同様の痛みの反応」をしなければならない、という条件

設定はあまりに過剰であると思われる。実際、前者に関して言えば、痛みの閾値に有意な個人差があることや生物種ごとに侵害受容器の機能や配置が異なりうることを考慮するなら、同じ刺激に対して痛みの生起に差異がありうるということはごく自然に理解できるだろう。また後者に関しては、そもそも反応表現に関する身体的制約が異なっている可能性を考慮するならば「画一的な痛みの反応」を要請することが不合理であることは明らかである。

しかし、このように考えてくると、「適切な状況で適切な振る舞いをする」という認定基準自体が実質的内容を持てなくなるのではないか、という疑問が生じてくるかもしれない。この疑問はもっともなものであり、「状況と反応」からなる諸パターン群のうちから「痛みの生起」であるものを抽出する文脈独立的な基準を明確化することは困難であると筆者も考える。この論点に関して本稿では、痛みの概念と密接な繋がりを持つと考えられる「道徳的配慮」の概念を手がかりにする、という方向性で問題の解消ないし軽減を試みたい。別の言い方でいえば、痛みの主体の側の性質に焦点を当てるのではなく、「痛みを認定し、それに対して配慮する」という観察者側の性質の方に焦点を当てて、「そのような配慮実践を引き起こす強い傾向を持つもの」という観点から実装すべき反応や振る舞いのパターンを構想していく、ということである。この方針を説明する上では、先に言及した問題点（2）についての考察が適切な導きの糸を提供してくれる。それゆえ、ここで問題（1）についての検討をいったんペンディングして、（2）の問題について本プロジェクトでのアプローチの説明に移行することにしよう。

さて、（2）であるが、たしかに［Dennett, 1987］や［Davidson, 1984］で提示される解釈主義的立場の代表例は、痛みのような感覚・意識状態を説明対象とするものではない。たとえば、デネットの「志向的スタンス」という考えは、心的状態を行動の予測・説明のためのものと見なした上で、これらを帰属させることで行動が合理的に予測・説明できるものを「志向的システム／心を持つもの」と規定する（たとえば、台所に行き冷蔵庫を開けると

いうある人物の行動を「ビールを飲みたい」という欲求や「冷蔵庫にビールがある」という信念の帰属により説明できるケース）。しかし、この説明に登場する心的状態は信念や欲求といった「内容を伴った心的状態」（「命題的態度」ともいう）であり、知覚や感覚・意識は基本的には説明対象に含まれないとされるのである。

しかし、解釈主義という考え方の一つの魅力が、私秘的・主観的である（ように思われる）心的状態を第三者がアクセスできる日常的な公共空間に引きずり出すことを可能にする点に存するのであれば、同じ議論を命題的態度以外の心的状態にも展開できるかどうか見てみるというのは悪くない取組みだろう。問題は、命題的態度のケースに見いだされた「行動予測」や「合理的説明」といった外的基準に類するものを知覚や感覚の場合にどのような仕方で設定するか、その際にどのような概念連関に訴えるか、という問題だが、これについてはすでにいくつかの先行研究の中で注目すべき取組みがなされている。たとえば、[Mölder, 2010] は、デイヴィドソンやデネットの見解を下敷きにしつつその問題点のいくつかを解消した「帰属説（ascription theory）」という立場を提示し、信念と区別されたものとしての知覚状態の特徴を非概念性・不可侵性・事実性という形で規定した上で、その帰属条件を予備的な仕方で定式化している。また、[Robbins & Jack, 2006] は、デネットの志向的スタンスをモデルとしながら、「Xに対して現象的スタンスをとる」ということを「Xを現象的システムとして理解すること、すなわち現象的経験[17]の座（locus）として扱うこと」（p. 69）として規定した上で、「何かを［現象的］経験の座として理解することは、関連する状況（苦しみを示している等）においてその何かに対して共感的に反応することであり、道徳的な関心を払う価値あるものと見なすことである」（p. 76）とする基本的な考えを示している[18]。

本稿では、基本的にはロビンスらの「現象的スタンス」という考えに沿って、「（現象的経験の一種とされる）痛みの保持」と「道徳的配慮実践の存在」との間に密接な概念的連関があるという考えを採用する[19]。そして、あるシ

ステムが「痛みを感じられるシステム」であるための条件を「配慮実践の確立・普及」という形で設定することにより、「我々が典型的に痛みを見て取るような状況に反応し、痛みに典型的な振る舞いをすることができる機能」ではなく、そういった反応や振る舞いを通して「我々の配慮実践を惹起するような機能」をこそ、実装目標に設定するのである。

この現象的スタンスという構想によって、解釈主義的な方針を命題的態度以外の心的状態にも適用する道筋は確保されるため、(2)の問題は一応回避できたと考えられる。そして、この構想の導入によって、先にペンディングしていた(1)の問題へも多少の光が当てられることになる。たしかに、痛みが認定されうる状況は極めて広範なものだが、それら全てをカバーできるだけの機能を実装する必要はない。むしろ、「痛みの保持」と「道徳的配慮実践」の間に密接な概念連関があるのだとすれば、道徳的配慮実践が惹起されるだけの状況検出能力と振る舞いパターンさえ組み込めば十分であると考えられる。また、道徳的配慮実践を介在させることにより、機能選定に関しても新たな情報が得られることになる。というのも、どういう場合に配慮実践が惹起されやすいかについてすでに我々自身がもっている常識的な知識を利用することができるし、それに加えて、近年注目を集めている社会脳研究等[20]を参照することにより、有望な機能の候補に関して多くの示唆が得られるだろうからである。

もちろん、こういった知見を導入したとしても、確定的な必要条件や十分条件を特定するには至らないかもしれないが、プロジェクトの性質からすれば、そういった条件確定は必要ない。むしろ、行われるべき作業は、「どういった特性が道徳的配慮実践を惹起しやすいか」に関する知見を参照しつつ、実際にそういった特性のいくつかを具体的に実装してみて、それが期待通りの反応を観察者側に引き起こすかどうかを確認し、随時(機能追加も視野に入れた)ヴァージョン・アップを試みる、というものであろう。そうした取組みのうちで徐々に目指すべき効果

が実現できるようになってきているならば、前提として採用していた方針をそのまま信頼し続ければよいし、いつまでたってもうまくいかないならば基本前提の見直しに着手する、という形で進めていけばよいのである（むしろ、実機制作を組み込むことの一つのメリットとして、概念的考察「だけ」では答えを得ることが困難で十分な前進を期待できないような論点について、哲学的議論にも利用可能なデータを少しずつ創り出して行くことによって議論を前に進めることができる、という点を指摘することができるだろう[21]）。

こうして、「現象的スタンス」という構想によって（1）と（2）の疑問については解消のめどが立った、あるいは解消に向けた活動を始める準備が整ったということができよう。しかし、もう一つ残されている疑問（3）についてはどうだろうか。これは、「痛み」を巡る我々の様々な実践にうまく接合できるようなロボットをつくったとしても、「痛みがあるように**見えるだけ**で本当は違うのではないか」という疑問であり、要するに「チートの可能性」をどう解消するかという問題が提起されているのである（実際、こういった疑念は授業アンケートやロボット工学者達との議論の際にしばしば表明されるものである）。

しかし、実はこの点についても、「現象的スタンス」の導入によってかなりの程度疑念を和らげることができるのではないか、と筆者は考えている。たしかに、「適切な状況におかれたときにそれらしい動きをする」「『イタイ』という音声を発する」といった機能の実現をゴールに設定するならば、チートの可能性――「そういうふうに動くように作られているだけでは？」という疑念――を払拭するのは難しいかもしれない。しかし、そういった検出能力や振る舞いの実現に加えて「配慮実践の確立・普及」をゴールに設定した場合にはどうなるだろうか。つまり、ロボットに針が刺さったときに「イタイ」というだけでなく、我々がその状態にあるロボットに対して配慮実践をとるようになり、かつそういった対応の仕方が普及する所までを目指すのである。この場合、チートの可能性を本

気で主張すること――つまり、痛がる振る舞いを適切な仕方で示すロボットがおり、その状態をケアする実践を自分自身が行いながらも、「本当に痛みがあるわけではないんだけどね」と留保をつけること――はかなり難しくなるように思われる。むしろ、ある対象の痛みの振る舞いに対して本気で配慮実践を行っている人は、その行いそのものによって、当該対象に痛みを認定していることになるのではないだろうか[22]。

先に述べたように、現象的スタンスという考えが示唆しているのは、ある種の典型的な振る舞いだけでなく、その振る舞い主体に対する道徳的配慮実践がなされることも、痛みという概念に対して密接な概念的繋がりを持っている、ということである。そして、前述のように（振る舞いにとどまらず）道徳的配慮実践を惹起する機能の実現をも達成目標に含めることが意味しているのは、痛みに関する複雑な概念ネットワークを断片化せずにできるだけそのまま移植・実装することで、当該ロボットが我々の痛み実践により深く根を張れるようにする、ということなのである。こういった試みによって、「痛みを典型的に引き起こすinputに反応し、痛みの典型的なoutputと見なされるような振る舞いをするだけのロボット」を目標に掲げた場合よりも、チートの可能性に対する疑念を大幅に縮減できるのではないかと期待するのは十分に合理的なことであろう。このことが認められたならば、あとは、どういう「配慮実践惹起機能」を実装すれば「チートの疑い」をどの程度縮減できるかという試行錯誤のプロセスによってヴァージョン・アップを行っていけばよいのであり、「改良するごとに調査を行い、またそこからさらなる改良のための示唆を得つつ、機能設計にフィードバックをかけていく」というサイクルを確立することで、どういう配慮実践惹起機能を実装すれば真正の痛み認定に至れるのか――あるいは配慮実践惹起機能を練り上げることでは痛み認定には至れないのか――という問いに答える効率的な道筋を与えることができると考えられるのである。

## 第四節　メタ哲学的構図の確認

これまで本稿では、「心を持ったロボットをつくる」というプロジェクトの輪郭を与えるために、痛み感覚の実現によって心を実現できるとする仮定、心的状態に関する解釈主義モデル、現象的スタンスの採用など、哲学的に十分正当化されているわけではないいくつかの前提をおきながら、議論を進めてきた。しかし、十分に立証されていない学説をロボット工学的実践の中に組み込むという異分野融合的な試みには、哲学的に重要な一つの意義があると思われる。それは、ロボット工学的な成果物（本プロジェクトに関してはそのロボットを用いて創出・普及させられる直観的理解も含まれる）によって、その成果物を得る際に前提されていた哲学的学説を遡及的・間接的に正当化する、という構図が可能になることである。つまり、なんらかの哲学的仮説を正当化するための方策の一つとして、実機を制作してその効果を検証するというロボット工学的アプローチが可能だということである。本プロジェクトに即して描き直すなら、諸哲学的主張を作業仮説としてたてることでロボットを作製し、それが最終的に「心を持ったロボット」として認定されるに至るならば、当該哲学仮説がさかのぼって間接的に正当化されうる、という構図がここで提示されているのである。[23] これは本稿の議論が示唆する一つの特徴的な「哲学的仮説の正当化ルート」であり、哲学的仮説の正当化のために利用できるリソースの拡張効果があると考えられる。

もう一つ、別の「正当化ルート」の可能性にも言及しておきたい。「心を持ったロボットをつくる」というプロジェクトが本稿で提示されたような仕方で進められた場合、その成否を判断する基準としては、「作製されたロボットに対する配慮実践が確立・普及するかどうか」「痛みの主体と認められるかどうか」というものがありうるだろう。ここで仮に、様々に機能をつくり込んでみたものの、ロボットを「道徳的配慮の対象」と見なすことに対する

直観的な抵抗がどうしても拭いきれず――せいぜい「ごっこ遊び」等の擬人化実践が成立するくらいで――結局配慮実践も痛み認定も確立・普及しなかった、という事態を想定してみよう。このケースではたしかに何かがうまくいっていない、少なくとも作製方針の背後にある理論的仮説群と直観的判断が整合していないわけであるが、問題の所在と対処法はどこに見いだされるべきであろうか。

自然と思われる選択肢は、作製方針の背後にある理論的仮説（解釈主義や現象的スタンス等）のいずれかを放棄する、というものであろう。しかし、筆者は、「直観的な抵抗」の方を修正することによって《理論と直観の整合性》を回復させる、という方策についても可能な選択肢に含めておきたいと考えている。たとえば、同じく非人間的対象である愛玩動物と機械的ロボットの間に《配慮対象とすることへの抵抗感》の違いがあったとしても、仮にその差異が言語学習期において痛み帰属実践や配慮対象実践のデフォルト的対象であったか否かの違い――怪我をした犬を見ながら「ワンワン痛いねえ。よしよし」と教示される一方で、ロボットは「壊れちゃうでしょ」といったnon-patient扱いを強化するような言い回しのデフォルト的対象として参照されるなど――に起因するものであったならば、その抵抗感の差異を改訂不能な本質的差異と見なすべき強い理由は存在しないように思われる。むしろ、心的語彙の言語習得期に参照されるデフォルト対象にロボットを組み込んでやる、といった仕方で教示方法を改善することにより、直観的な抵抗感を弱めることができるかもしれないのである[24]。

ここで述べられたことがらは多分に空想的な想定に基づくものであり、我々の概念実践や直観の実際のあり方についての記述としては十分な根拠を持たないが、それでも、「ロボットを『道徳的配慮の対象』と見なすことに対する直観的な抵抗」の方を修正するという方策が単に「往生際の悪い不合理な抵抗」であるわけではない、という点を理解する助けにはなるであろう。つまり、直観の改訂によって哲学的仮説の正当性を確保する、というルート

は原理的に不可能なものというわけではないのである（実際、筆者自身は、ロボットにつくり込む機能を向上させることと平行して、こういった言語学習期に確立される直観を改訂しようとする試みを進めることが、「心を持ったロボット」の実現にむけたもっとも効率のよいアプローチであると考えている[25]）。

以上の検討から、「心を持ったロボットをつくる」というプロジェクトに対する本稿の議論は、「哲学的仮説の正当化ルート」に関して、やや拡張された仕方で展開されていることが理解されるだろう。一つは、ロボット工学的なアウトプットの成功に立脚して、そこで仮説として前提されていた哲学的仮説を間接的に正当化する、というルートを認めている点であり、もう一つは、理論と直観の不整合を解消するための方策として「（能動的な仕方での）直観の改訂」という選択肢を考慮に入れている点である。

これら二つの正当化ルートは、哲学において典型的に見られる「理論的主張に対する標準的な正当化手続き」から逸脱しているように見えるかもしれないが、実は、第一節で提示された「理論と直観の合致を目的とする」という構図のもとに包摂することが基本的には可能である。標準的手続きからの「逸脱」は、第一のルートに関して言えば、実機を用いることによってはじめてなしうる認知調査などを介して純粋な思考実験では確保できない直観的判断を創り出す点に、また第二のルートに関して言えば、広範に見られる直観的判断そのものを（実機投入に次世代教育等を組み合わせるなどして）能動的に作りかえる可能性を考慮に入れている点に存している。しかし、本稿での取組みは、上記のような拡張によって「直観的理解」の範囲を伸縮させる可能性があるとしても、理論と直観の合致をもって着陸点とする構図そのものは従来型の手続きと共有している。それゆえ、本稿の取組みは従来の正当化手続きから許容できないほど大幅に逸脱した不適格なものではない、と筆者は考えている。

## 結びにかえて

第二節の冒頭で、「心を持つための必要条件」をあらかじめ議論によって確定してから、その実現に向けて動き出す、というやり方はとらないと述べた。本稿で支持されているのは、むしろ、実機を用いることによってなしうる調査・実験によって「我々の側の反応」を確認しながら、「心を持ったロボットをつくる」ために何が必要であるかを一歩ずつ明らかにしていく、というのが本稿のアプローチである。こういった取組みが示しているのは、心の哲学における議論を他の諸学における研究と接続可能なものとして位置づけなおす「自然化」への道筋であり、哲学的諸立場の妥当性を巡る論争を（ゼロベースの思弁的思考実験のみに頼らざるを得なかった状況に比して）一つ前に進めるために使えるであろう「諸立場によらず共有できる新たな情報」を創りだす可能性である。《理論と直観の合致による正当化》という従来の構図を維持しつつ、こういった情報によって「直観的理解」の範囲を伸縮させる可能性を示唆する点において、本稿は、「心を持ったロボットをつくる」というプロジェクト構想の提示を主題としながらも、心の哲学に関する特定の立場――感覚状態にも適用されうるものとしての解釈主義――の正当化に向けた一階の取組み[26]、およびそういった議論全体が位置づけられるべきメタ哲学的構図の再考という二階の取組みをも含み込んだものであった、ということができよう。

### 註

（1）　本稿の作成にあたっては、石黒浩氏、笠木雅史氏、小山虎氏をはじめ、多くの方々から有益なコメントを頂くことができ

た。各氏は必ずしも筆者の見解に同意しているわけではないことを明記した上で、ここに記して謝意を表したい。

（2）「直観的理解」を定義するのは困難だが、さしあたり、（i）我々が思弁や推測に頼らずに即座にくだすことができる判断であり、かつ（ii）焦点となっている概念の把握に由来する一定の正当性をそれ自体として確保している判断である、という形で理解しておきたい。これは、「心ってそういうもんだよね」というような概念的理解の輪郭を形づくるものであり、提案された定義が「心の定義」になっているか否かを判定するための基準として頻繁に用いられる。（直観概念の特徴づけと哲学的理論の正当化作業におけるその役割については［笠木、二〇〇九］にレビューがある。）

（3）ただし同一説に対する再評価の動きも存在する。［太田＆山口、二〇一〇］を参照。

（4）「機能主義」は理論的洗練の途上で様々なヴァリアントを生み出しており、哲学的立場としての妥当性を精査するにはこれらの一つ一つを個別に論ずる必要があるが、本稿の主題であるプロジェクト遂行指針には関わらないため、ここでは検討しない。

（5）もちろん、科学的知見との合致、内的整合性、そこから帰結する存在論的見解の真っ当さ、理論的・実践的課題への有効性などといった別の評価基準もある。

（6）［丹治、二〇〇九］を参照。なお、筆者は［井頭、二〇一〇］において分析性概念を擁護しているが、そこでの議論は「原理的な改訂不能な言明」の存在を主張するものではない。

（7）「感覚」や「痛み」には心の哲学における最大の難物であるクオリアの問題があるのだから、「実現への目処が立ちやすい」という判断は根本的に問題状況を見誤っている、と思われる読者がいるかもしれない。しかし、次節以降で展開される「現象的スタンス」に関する議論によって、この問題はある程度処理可能であると筆者は考えている。

（8）この点で本プロジェクトは「思考能力の実現」を目指してきた従来型知能ロボティクスの取組みと一線を画すものになっている。なお、類似の着眼点を持つ先行研究としては［柴田、二〇〇八］および〈http://siva.w3.kanazawa-u.ac.jp/index.html〉が参考になる。

（9）これは実際にある工学系の研究室で検討された取組み――構想されたセンサは機械刺激の一部のみを検出するものであった――であるが、本文で述べられた諸問題を解消する目処が立たずに、現在その取組みはストップしている。

（10）［Lewis, 1980］を参照。なお、この後者の論点に関しては「火星人特有の《神経科学》を構築すればよいではないか」という形で反論が提起されるかもしれない。ここでは紙幅の問題で詳述はできないが、神経科学的情報のみによって「痛

み」を定義するというこの路線を擁護しきることは困難であると筆者は考えている。というのも、火星人の《神経科学状態パターン群》のうちで「痛み」の状態にあるパターンとそうでないものとを区別するためには、結局「ある状態にある火星人に痛みが生じているか否か」を先行して判断し、それに基づいてパターン群の仕分けを行う必要があるからである。同様の議論は本稿の主題であるロボットのケースでも成立する。

(11) 〈http://www.iasp-pain.org/AM/Template.cfm?Section=Pain_Definitions〉にある IASP Taxonomy から。また、[Anand & Craig, 1996a] [Merskey, 1996] [Anand & Craig, 1996b] [Shapiro, 1999] なども参照のこと。

(12) ただし、本稿では、いかなる意味でも「言語報告できない者は痛みを持ちえない」という見解は支持されていない。

(13) ここで提示されている描像は [Wittgenstein, 1953] [Sellars, 1956/1997] [Brandom, 1997] の議論を大雑把に組み合わせたものに過ぎない。なお、一人称的にのみアクセス可能な心的状態を理論的措定物と類比する形で理解するというアイデアについては [Sellars, 1956/1997] の四八―六三節を参照せよ。

(14) この意味において、第一次的言語実践は根源的で原初的なものであり、比喩的にいうならば、ある種のデフォルト性を備えた〈意味のふるさと〉であると考えられよう。

(15) 本文では、紙幅の関係及びモデルの導入意図との関係が希薄であることから言及を避けたが、事実としての発達プロセス・言語習得プロセスの実態とマッチしたものとしてこの種のモデルを提示するならば、発達心理学や認知科学の知見を取り入れてより詳細なモデルに仕上げる必要があるだろう（特に、公共的情報に定位する第一次的なパターン習得と、一人称的情報に依拠するパターン習得を独立のフェーズとして扱っている点は、実態と大きくかけ離れているだろう）。しかし、このことは本論での主張の妥当性に大きく影響を及ぼすものではない。本文の議論の文脈において重要なのは、一人称的用法の単独での確立・習得可能性の否定であり、一人称的用法の習得が三人称的用法に依存し、それに統制される形で秩序化されるという点の指摘である。なお、このモデルの提示にあたっては「言語が『流通しつつ変化し』ながら連続的に存続する」という事象に関する [丹治、一九九六、二〇五] の考察も参照されている。

(16) なお、この立場と [井頭、二〇一〇] で提出された「多元論」の立場とは明確に区別されるべきである。前者は痛み概念に関する複数の概念図式・概念実践の併存を認める立場であり、後者は概念図式群を事実記述的なものとそうでないものとに仕分けする一元論的基準の存在を否定する立場である。あるいは、[井頭、二〇一〇、二二〇―五] で言及した「水平的／垂直的多元論」の区別を用いて差異を明確化してもよいだろう。また、関連して [丹治、一九八五、一七〇] にあ

る《概念の過剰規定》に関する見解も参照されたい。

(17) 現象的経験とはクオリアを伴うとされる痛みや気分などの意識経験である。

(18) なお、ロビンスらは、あるものに対して現象的スタンスをとることは「そのものを道徳的配慮の対象として扱う」ということに尽きるものではないとし、「本能的な共感能力」や「他者との緊密な関係」といった他の概念との連関についても言及している。

(19) ただし、本プロジェクトへの援用という観点から見れば、ロビンスらによる現象的スタンス規定には若干の修正が必要になる。以下では、(a) 現象的システム認定と配慮実践との論理的関係、および (b) 客観性の確保、という二点について簡単に触れておく。

まず (a) について。もちろん、「現象的経験の保持」を理解する上でそれと概念的に連関した「道徳的配慮実践」が導入されることに異論はないし、痛み等の現象的経験に関する認定条件の具体化が求められている現状ではむしろ有望なアプローチである。しかし、先に述べたように解釈主義の一つのメリットが「私秘的・主観的である（ように思われる）心的状態を公共空間に引きずり出す」ことにあると考えるならば、「Xに対して現象的スタンスをとる↓Xを道徳的配慮の対象とする」という概念連関だけでは不十分であると思われる。なぜなら、この一方向的な含意関係からは、「Xに対する道徳的配慮実践の確立」という公共的情報から「Xは現象的経験を持つ」を導出できないからである。そこで、本稿では、用語を拝借してロビンスらの取組みに敬意を表しつつ、「Xを道徳的配慮の対象とする↓Xに対して現象的スタンスをとる」という含意関係を組み込んだ「改訂版現象的スタンス」に立脚して以降の議論を進めていくことにする。

次に (b) だが、「痛み」と「道徳的配慮」の間の密接な関係に訴えることで痛みの公共化を目指すという路線が基本的に妥当なものであったとしても、これだけではデネットの「志向的スタンス」の議論がもっている様々な長所の全てを模倣するところまでは至らない。一例として、デネットの志向的スタンスは「志向システムである＝志向的スタンスからの予測が十分なレベルで成功する」という規定を与えることによって、あるシステムが志向的システムであるかどうかを（「**実際に**誰かが志向的スタンスからの解釈を行うこと」に依存しないという意味で）「客観的な事実」として認めることが可能になっている、という点を挙げることができる。このような「（偶然的な解釈実践の生起に依存しないという意味での）客観性」を「現象的システム性」に関して確保するためには、配慮実践の偶然的生起ではなく、「配慮実践の確立・普及」という正準的事態に訴える必要があると思われる。また、客観性確保のための同様の正準化手続きは、「（現象的シ

ステム認定されたものに対する）個別的状況での痛みの帰属」に関してもなされる必要があるが、このためには「どういうinput/outputパターンであれば配慮実践の生起が《まっとうなもの》と見なされるか」といった分析を通して、「誤った痛み認定／配慮実践」の可能性を確保することが必要と考えられる（この見込みが正しければ、「配慮実践」は、「現象的システム認定」と「個別場面での痛み帰属に必要なinput/outputパターンの同定」という局面で機能する二重の役割を持つことになる）。いずれにせよ、「現象的システム性」および「個別状況での痛みの生起」に関していかに客観性を確保するかについては、紙幅の関係もあり、稿を改めて詳細に論ずることにしたい。

(20) [Craig et al, 2010]など。社会脳研究一般については[開&長谷川、二〇〇九]を参照。

(21) 実際、大阪大学大学院基礎工学研究科・知能ロボット学研究室（石黒研究室）では、これまで本稿で提案してきた構想の多くを共有しながら、プロトタイプを作製しつつ、「道徳的配慮実践を惹起できるロボット」の実現を目指した研究が進められている。

(22) 第一節で見たように、我々の直観との整合性は理論の正当化根拠にも、理論を攻撃するための根拠にもなりうる。そして、ある種のロボットを心あるものと見なす実践が普及した場合、それはこの理論評価基準としての直観を構成するものとなるため、「ある種の実践の創出」はある主張に対する正当化作業の一部を構成しうるのである（このような議論の運びは、少なくとも日本の哲学者にとっては、陳腐とは言わないまでもお馴染みの議論であり、多くの先行研究・論考に見られるものである。ここでは、アクセスのしやすさから以下を挙げておく。[大森、一九八一][山田、二〇〇九][丸田、二〇一〇]）。

(23) 逆に哲学的知見がロボット工学に利用される可能性については本稿で示してきた通りであり、このような互恵的な構図が成立することは大きな学術的前進と言えよう。

(24) もしこういった仕方で抵抗感を弱めることができるならば、比較的高年齢の世代に関してはロボットを配慮対象とすることへの直観的抵抗感をぬぐい去れなかったとしても、次世代教育を充実させつつ時代の進行を待つことで、数十年後には「ロボットに心を認める社会」が形成できるかもしれない。なお、幼児におけるロボットへの心的状態帰属の可否に関する先行研究のレビューとしては[板倉、二〇〇九]も参照せよ。

(25) こういった作業を行うには、我々の概念実践の仕組みを理解し、それに対して介入するための技術や方策を生み出さねばならない。[井頭、二〇一二]では、それぞれの作業に「概念ダイナミクス」「概念エンジニアリング」という呼称が与え

られ、相互フィードバックにより我々の概念実践に関する理解を深めていく、という路線が提示されている。

（26）このような取組みによる正当化は間接的なものに留まるが、たとえば、ある機能主義的立場が要請する機能を持たないにも関わらず配慮実践を惹起するロボットが普及し「少なくともそのロボットは痛みを持つ」という直観が広まった場合、当該機能主義的立場はその直観を阻却するためのコストを負う、という意味で不利になるだろう。

## 文献表


[Anand & Craig, 1996a]：K. J. S., Anand and K. D. Craig, "Editorial: New perspectives on the definition of pain," in *Pain* 67

[Anand & Craig, 1996b]：K. J. S., Anand and K. D. Craig, "Re: Reply to Letters to the Editor from Merskey & Wall," in *Pain* 67

[Brandom, 1997]：R. Brandom, "Study Guide," in［Sellars, 1956/1997］

[Craig et al, 2010]：K. D. Craig et al, "Perceiving *Pain* in Others: Automatic and Controlled Mechanisms," *The Journal of Pain*, Vol. 11, no. 2

[Davidson, 1984]：D. Davidson, *Inquiries into Truth and Interpretation,* Oxford U. P., 1984／D・デイヴィドソン（野本和幸ほか訳）『真理と解釈』、勁草書房、一九九一

[Dennett, 1987]：D. Dennett, *The Intentional Stance,* A Bradford Book, 1987 ／邦訳：ダニエル・C・デネット（若島正・河田学訳）『志向姿勢の哲学』、白揚社、1996

[Knobe & Nichols, 2008]：J. Knobe & S. Nichols (eds.), *Experimental Philosophy,* Oxford U. P.

[Lewis, 1980]：D. Lewis, "Mad pain and Martian pain", in N. Block (eds.), *Readings in the Philosophy of Psychology,* Vol. I., Harvard U. P.

[Merskey, 1996]：H. Merskey, "Response to Editorial: New perspectives on the Definition of *Pain*," in *Pain* 67

[Mölder, 2010]：B. Mölder, *Mind Ascribed: An Elaboration and Defence of Interpretivism,* John Benjamins Pub. Co.

[Robbins & Jack, 2006]：P. Robbins and A. I. Jack, "The Phenomenal Stance," in *Philosophical Studies,* Vol. 127

[Sellars, 1956/1997]：W. Sellars, *Empiricism and the Philosophy of Mind*, Harvard U. P., 1997／W・セラーズ（浜野研三訳）『経験論と心の哲学』、岩波書店、二〇〇六
[Shapiro, 1999]：B. S. Shapiro, "Implications for Our Definitions of *Pain*, " in *Pain Forum* 8(2)
[Wittgenstein, 1953]：L. Wittgenstein, *Philosophical Investigation*, Oxford, Blackwell, 1953／藤本隆志（訳）『ウィトゲンシュタイン全集八　哲学探究』、大修館書店、一九七六
[井頭、二〇一〇]：井頭昌彦『多元論的自然主義の可能性』、新曜社
[井頭、二〇一二]：井頭昌彦「多元論的自然主義は怠惰な形而上学なのか?」、第一回自然主義研究会発表資料（著者から入手可）
[板倉、二〇〇九]：板倉昭二「ロボットに心は宿るか」in［開&長谷川、二〇〇九］
[太田&山口、二〇一〇]：太田紘史・山口尚「反機能主義であるとはどのようなことか」、*Contemporary and Applied Philosophy* 第二巻
[大森、一九八一]：大森荘蔵「ロボットの申し分」『流れとよどみ』、産業図書
[笠木、二〇〇九]：笠木雅史「実験哲学からの挑戦」第一回応用哲学会発表原稿（著者から入手可／〈http://nonameblog.seesaa.net/article/139703851.html〉にて閲覧可）
[柴田、二〇〇八]：柴田正良「感情のクオリアと可能世界」『感情とクオリアの謎』、昭和堂
[丹治、一九八五]：丹治信春「行為の自由と決定論」、大森荘蔵ほか（編）、『新・岩波講座哲学十　行為・他我・自由』、岩波書店
[丹治、一九九六]：丹治信春『言語と認識のダイナミズム』、勁草書房
[丹治、二〇〇九]：丹治信春『クワイン　ホーリズムの哲学』、平凡社
[開&長谷川、二〇〇九]：開一夫・長谷川寿一（編）『ソーシャルブレインズ』、東大出版会
[丸田、二〇一〇]：丸田健「魂に対する態度」、大阪大学大学院人間科学研究科紀要
[山田、二〇〇九]：山田圭一『ウィトゲンシュタイン最後の思考』、勁草書房

（※）本稿の一部は科学研究費助成金（22720007および24720007）の成果である。

（いがしら　まさひこ・一橋大学大学院社会学研究科）