

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

メタデータ	言語: Japanese 出版者: 公開日: 2017-10-03 キーワード: 作成者: 柴田, 正良, Shibata, Masayoshi メールアドレス: 所属:
URL	http://hdl.handle.net/2297/24268

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——

柴田 正良

1. 倫理はなぜ必要なのだろうか？

いま、ある行為者が宇宙の閉じた領域で、たった一人で存在しているとしよう。彼（彼女でもかまわないが）はわれわれ人間ほどには知的であるが、生存のために仲間を必要とせず、また生存する上でとくに何の使命も帯びていない人間以外の存在者（行為者）だとしてみよう。また、彼は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、それらが崩壊して長らく一人でずっと存在してきたし、これからもそのような状態で長く存在していくのだとしてみよう。つまり彼は、現在も未来も、それどころか現実的であれ可能的であれ、いかなる共同体のメンバーでもない。ただ、彼は過去に何らかの共同体のメンバーであった、という事実とその記憶があるにすぎない。この想定は、ロビンソン・クルーソーが人間社会から切り離されたままにいる状況に似ていると思われるかもしれない。しかし、少なくとも決定的に違う点の一つが存在する。ロビンソン・クルーソーは、空間的には彼の属する共同体から切り離されているけれども、依然としてその共同体のメンバーであり、ただ共同体からすれば彼がたまたま「行方不明」になっているというだけのことである。事実、彼のいる島に他の誰かが新たに漂着してくることは可能であるし、彼が救出されて元の共同体に復帰することも可能であろう。

さて、われわれの想定する新クルーソーの方に戻ろう。私には彼のような人生が生きるに値する面白みのあるものだとは思えないが、彼は気ままにその日その日を暮らしていくことができる。もしかすると、彼は神のごとき「創造」を楽しんでいるかもしれない。彼を妨げるような仲間や敵がいるわけではないし、彼が妨げるような仲間や敵がいるわけでもない。ある意味で徹底して孤独なこの新クルーソーにとって、倫理とは何だろうか？ あるいは、彼にとってどのような倫理が必要なのだろうか？ いささか唐突な問いかもしれないが、これは倫理がそもそも人間存在のどこに根ざすかに答えるための問いである。典

型的な二つの規範的倫理、カント的義務論と功利主義を考えてみよう。

例えば、カント倫理学における義務の一つの定式化『実践理性批判』では、「人のなすべきこと」は次のように説かれる。

「君の意志の格率が、常に同時に普遍的立法の原理として妥当しうるように行為せよ」(Kant (a), p.140, 邦訳 p.50)。ここで意志の格率とは、行為者の「主観的実践の原理」のことである。すると、これは要するに、自分が何かをなそうとする際の行為方針が万人の行為方針として普遍化しうる、つまり普遍化しても法則として自己挫折(矛盾)しない、ということが各自の行為方針に求められているということである。したがって、法則化できない格率は、その内容のゆえにはなく、その形式的な特徴のゆえに行為者のなすべきことたりえないことになる。例えば、法則化できない格率の例としてよく引き合いに出されるものに、「約束は、自分の都合によりいつでも勝手に破れ」がある。これはなぜ法則化できないとされるのだろうか。もし約束がいつでも当事者の勝手な都合で破られるものであるなら、約束が結ばれても一般にはそれを信頼することができないだろう。とくに、当事者の一方にとって約束を果たすことが不利益であったとしても「約束だから」それが果たされる、ということを他方が期待することはほぼ絶望的となる。したがって、約束が最も必要とされるまさにその場合に約束が効力を発揮しない、ということこそ意味するだろう。つまり、先の格率は、もし法則化されるなら自己挫折的に約束という行為自体の存立根拠を掘り崩してしまうのである。これは、約束の成立条件(構成的条件)の一つに、「約束は、約束するその当事者を遵守へと拘束する」というものがあるからだと考えられる。つまり、「勝手な約束破り」は、約束を可能とする条件の一つと矛盾するのである。

カントが、格率の法則化可能性を「なすべきこと」の必要条件としてだけでなく、十分条件としても考えていたのかどうかはともかく¹、ここで問題にしたいのは、新クルーソーにとって法則化できないような格率は存在するのだろうか、ということである。新クルーソーは、現実的にも可能的にも、いかなる共同体のメンバーでもない、ということを目指そう。彼の行為の中には、他者との相互性を構成要因とするような対他的行為は存在しない。協力も、同意も、裏切りも、否認も、共同体の他のメンバーがそもそも存在しないのであるから、彼には原理的になしえない行為タイプである。言いかえれば、他者との相互性を構成要因としない行為しか彼には行うことができない。すなわち、因果的実効性を及ぼすものとして彼が行う行為は、自分自身か、あるいは自分以外の自然の他の部分に働きかける行為のみである。それらの行為に関する方針が法則化できない、という

理由が何かあるだろうか？ 万人が（といってもこの場合、適用対象は実際には彼一人だけ）行うように普遍化されたときに行為自身が自己挫折するような格率は、この場合あるようには思えない。なぜなら、ある行為タイプ a の構成条件を掘り崩すような格率に従う行為タイプ b を行いながら、なおその行為が a タイプであるための観点が存在しないからである。というのも、この場合、彼は行為タイプ a をなすのを止めて、たんに行為タイプ b を行うように心変わりをしたにすぎないからである。彼の行為が挫折した a タイプであり続けるためには、それにとってその行為がなお a タイプであるような、他者の存在が必要である。

したがって、新クルーソーのどの行為格率も法則化の点で区別がつかないように思われる²。すなわち、彼の占める閉じられた宇宙空間において、彼の行為を評価すべき倫理的観点というものは存在しない。言いかえれば、彼のいかなる行為も、倫理的に善でも悪でもないのである。

ここで唯一、考察すべき例外と思われるものは、新クルーソーの行為が自然の他の部分に向かうのではなく、自分自身に向かった場合の極端な行為、すなわち自殺である。自殺もまた、カントの言う意味での法則化が不可能な格率であると言われる。その理由は、もし万人が「望むなら自殺せよ」という格率に従い、万人がそれを望む場合にはいかなる行為者も存在しなくなるからだ、というものである。つまり、この格率を採用する行為者がまったく存在しなくなる可能性がある以上、この格率の法則化は自己挫折的だということである。しかし、この自己挫折的事態の根拠は、格率内部の形式性ではなく、格率が実現される世界の実質性にある。したがって、それは法則化不可能ではないであろう。これはちょうど、宇宙の歴史中である期間、ある化学物質群が一定の化学法則に従って振る舞っていたのだが、それ以後その化学物質群が存在しなくなった場合に似ているかもしれない。この場合、それらの化学物質群を支配していた化学法則も存在しなくなったと解釈する必要はない。その化学法則は適用対象をもたないまま、空虚に成立していると考えることができる。同様に、「望むなら自殺せよ」という格率の実現がたまたまその世界で行為者の完全な消滅を導いたとしても、この格率は採用者がいないままそこで空虚に法則化されていると言うことができる。したがって、新クルーソーの行為格率において、自殺が特別な例外をなすわけではない。そうであるなら、カントの法則化可能性という枠組みからするなら、新クルーソーの世界においては、倫理なるものは必要ではないし、そもそも存在しないように思われる。つまり、彼にとって「なしたいこと」とは別の「なすべきこと」

は存在しないように思われるのだ。

さて、次に功利主義的な枠組みにおいて、新クルーソーの「なすべきこと」がどうなるかを見てみよう。ベンサムの「最大多数の最大幸福は道徳と立法の基礎である」という功利主義思想は、ミルの功利性原理によって次のように表現されている。

「功利性、あるいは最大幸福原理を道徳の基礎として承認する考え方は、行為は幸福を増す傾向に比例して正しく、幸福の逆を生み出す傾向に比例して不正である、と主張する。」(Mill, p.210)

われわれの新クルーソーが閉じた宇宙空間において、現実的にも可能的にも孤独であるということを読み起こそう。功利性原理の言う「最大多数」は、この場合、彼「一人」のみしか意味しない。したがって、多くを論ずるまでもなく、彼がいかなる行為をなそうとも、彼が自分自身の幸福を減じるような行為を行わない限り、「なしたいこと」に反して彼に課される「なすべきこと」は存在しないように思われる。そして、彼が自分の幸福を損なうような行為を自覚的になそうとすることは一般にはありえない。例えば、認識の誤りや行為の不手際によって、新クルーソーが不幸に見舞われるということはあるだろう。例えば、彼は傍らの緑の光る物体に触れたいという強い欲求に駆られるかもしれないし、急峻な岩山を登りたいという気持ちを抑えられなくなるかもしれない。そしてそれらの行為が、自分を傷つけ、自分の幸福を損なうという結果を彼にもたらす場合もあるだろう。しかし、だからといって、彼はそれらの行為を「なすべきではなかった」のだろうか。もし何らかの外的原因によってそれらの行為をなすことが妨げられたなら、彼の欲求はくじかれ、彼の幸福は大きく減じられたであろう。したがって、彼は自らの幸福の増大のために、結果を知りえない行為遂行の時点ではそれをなすべきであったであろう。それでも彼が過去の不幸な結末を振り返って「あの時、ああすべきではなかった」と嘆くような場合、彼の後悔は倫理的な観点からなされているのだろうか。共同体の他者がまったく関与しない行為によって、人が勝手に不幸な結末を自分だけに招いた場合、「彼は倫理的に不正な行為をはたらいた」と言うのであれば、その評価には何か異様なものが含まれているように思われる。それは、そもそも功利性原理が他者との関わりにおいて初めて必要になってくるものだからである。功利性原理の前提は、「人は自らの幸福を最大にするように行う生き物だ」ということだ。したがって、一人で放っておかれれば幸福を自然に追求するそのような行為者にとって、功利性原理はまだ実質的な意味を持っていない。ある意味では、そのような行為者はすべての行為において功利性原理を空虚に満たしている、ということ

ができる。というのも、このような行為者が複数存在し、互いに利害を関与させあうことが不可避となって初めて、「自らの幸福を最大にする」ために各人の行為が調整されなければならないからである。功利性原理はその調整の指針である。それゆえ、概念的な「以前」という意味において、共同体の他者が存在する「以前」の個人の行為を倫理的に評価するのは、本末転倒である。平尾透がミルを解釈して言うように、「・・・道徳とは究極的に社会的又は対他的なもの・・・」なのである（平尾、p.10）³。

したがって、功利主義的な枠組みにおいても、新クルーソーのいかなる行為においても倫理なるものは必要ないように思われる。念のために、それをもう二つの点で確かめておこう。一つは自殺であり、もう一つは、彼の閉じた領域において生息するかもしれない彼以外の感受的存在者である。自殺もまた、一見すると功利性原理に反するように思われる。というのも、自殺は行為者のその後の幸福を得る機会を永遠に奪ってしまうからだ。しかし、行為者のその後の可能的な幸福はその後の可能的な不幸と比較されるべきであり、その比較結果は確定することが不可能である。したがって、新クルーソーのような絶対的に孤独な行為者においては自殺が影響を及ぼすような他者が存在しないがゆえに、自殺の幸福に対する関係は、彼本人のその時点での〈実際の幸福〉と〈実際の不幸〉の関係のみに局限されるはずである（「実際の幸福／不幸」には、自分の未来の状態に関する予測結果が含まれてもよい）。すると、彼の自殺の理由が何であれ、彼はその時点で「自らの幸福を最大にする」ために自殺を選んだということになる。したがって、自殺という極端な行為においても、彼にとって「なしたいこと」とは別の「なすべきこと」は存在しないことになる。

最後の問題は、「・・・功利主義の基準は、行為者自身の最大幸福ではなく、幸福の総計の最大量だからである」（Mill, p.213）と言われる際の「幸福の総計」に、快苦を感じず、人間以外の存在が何か含まれていないかどうかである。つまり、もし新クルーソーの世界に、彼と共同体と一緒に形成するメンバーとはなりえなくとも、傷つけられたり踏みつけられたりすれば苦痛を感じずような感受的な生き物が存在したとすれば、彼はそれらの「幸福」に対して何らかの義務をもたないのであろうか。しかし、功利主義の定式化においてよく問題となる「考慮されるべき善の総量には、すべての感受的存在者のすべての快苦が含まれる」という考えには、どこか不明瞭なものがある。そもそも、地球のような環境の中で進化した行為者は、基本的に自分を動かすエネルギーを得るために摂食・代謝の方法に依っている。したがって、行為者は食物連鎖の中に位置づけられ、食物連鎖の中には快

苦を感じず生物体がふつつ含まれている。それは、彼の幸福の実現のために他の生物体の不幸の実現が必要であることを意味する。このとき、行為者の捕食によって実現される彼の幸福と、捕食される者の不幸をいかにして比較し、幸福の総量を計算することができるのだろうか。その計算の結果、行為者があまりに多数の生物体を捕食するがゆえに存在を止めることが彼の「なすべきこと」である、というような結論が生じるのだろうか。功利主義はそのような「生来の呪われた存在」という可能性を含意しているのだろうか。しかし、功利原理がそもそも各人の幸福追求の調整原理であることを思い起こそう。つまり、功利原理はそれに服する各行為者が調整に応じる立場にあり、かつまたそのための基礎的能力を有することを前提としている。逆に、例えば、地球における人類の登場以前の食物連鎖の世界を考えてみよう。その世界に生きる快苦を感じず生物体に功利原理を適用することは、およそ彼らが理性能力も言語も持たず、相互の複雑なコミュニケーションも不可能である以上、まったく的的外れであるように思われる。つまり、そもそも自ら功利原理を採用する能力を持たない存在者に功利原理を適用するのは、本末転倒だということだ。もちろん、これは現在の人間にとって、いくらでも動物たちを虐待したり殺したりすることが許される、ということではない。しかし、この「許されない」は相手に対応する権利を生じさせるという意味での「なすべきでない」(完全義務)ではなく、そうしないことが推奨されるという意味での「なすべきでない」(不完全義務)である。人間は後者の意味で、他の種の生物体を不幸にすべきではない。人種や性別などによる権利の制限が撤廃されて、同一の権利を有する人間の範囲が拡大してきたという歴史的事実は、相互に完全義務を負う人間の範囲がたまたま拡大してきたということであって、人間が道徳的に配慮すべき対象者の拡大は常に完全義務対象者の拡大である、ということではない。ある種の人間非中心主義的な環境倫理が目標とする「あらゆる存在者への同一の権利の拡大」は、この二つの義務のあり方を(ある場合には故意に)混同していると思われる。ともあれ、以上のことは、功利主義的な枠組みにおいても、新クルーソーにとって、「なしたいこと」とは別の「なすべきこと」は存在しないということである。

さて、これまでのいささか奇妙な議論がさまざまな点で穴だらけであることは率直に認めたいと思うが⁴、これによって私が示したかったことは、ただ一つ、新クルーソーのような存在にとって、そもそも倫理なるものは存在しないということである。彼がいかに振る舞おうとも、それは倫理的に善でも悪でもない。彼の「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、このことはさらに、行為者に倫理が発生するとす

れば、それは、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、ということを示唆している。これを、ここでは「共同体テーゼ」と呼んでおくことにする。共同体テーゼは、同一の共同体への複数のメンバーの所属が倫理の発生の十分条件であるとまでは主張しない。それは必要条件である。共同体が存在しながらも、新クルーソーのように、すべてのメンバーにとって「なしたいこと」とは異なる「なすべきこと」が事実において長期にわたって存在しないとすれば、それこそは、彼らにとって倫理的にもっとも好ましい世界だと私には思われる。したがって、「なしたいことをなすこと」が最もプリミティブな意味での自由だとすれば、倫理は、互いの自由の衝突を避けるために何らかの仕方で相互の「なしたいこと」を調整せざるをえないがゆえに必要とされる、ということができる。自由の<欠如態>としての倫理という観点からすれば、倫理的価値において、最も尊重されるべきはこのリバタリアンの自由である。

2. ロボット倫理の前提・・・物理主義的世界

ロボットの倫理は、ロボットをどのような存在だと考えるかによって内容が定まってくる。例えば、有名なアシモフの「ロボット三原則」を見てみよう。

- 第一条 ロボットは人間に危害を加えてはならない。また、その危険を看過することによって、人間に危害を及ぼしてはならない。
- 第二条 ロボットは人間にあたえられた命令に服従しなければならない。ただし、あたえられた命令が、第一条に反する場合は、この限りでない。
- 第三条 ロボットは、前掲第一条および第二条に反するおそれのないかぎり、自己をまもらなければならない (アシモフ、p.5)。

しかし、正確にはこれは倫理ではない。とくに、ロボットたち同士が同一の権利・義務を互いに有することを前提とする倫理でもないし、人間たちとロボットたちが同一の権利・義務を互いに有することを前提とする倫理でもない。これはせいぜいのところ、人間のよりよき道具となるための「仕様書」にすぎない。よく指摘されるように、ロボット三原則は、少し内容を修正すれば家電製品が守るべき三原則に早変わりする。例えば、それは、(一) 安全 (人間を危険にさらさない)、(二) 従順 (人間の意図通りに動く)、(三) 堅

牢（簡単には壊れない）のように、対応する「家電三原則」に書き換えることができる。しかし、われわれがいま求めたいのは、人間とロボットが同一の権利・義務をもって共生するための倫理である。誕生（製造）、死（破壊）、認知能力、身体能力、外観、生殖（再生産）といった多くの基礎的な存在条件が大きく異なる2つのグループが同一の共同体を形成するという事は、いかなる倫理が妥当とされることなのであろうか。しかし、その議論を始める前に、人間に使われる道具ではなく、人間と共同体を形成するロボットを作成しようということの存在論的な前提を明らかにしておこう。それは、一言でいえば、そうしたロボットの制作を可能とするわれわれの世界が物理主義的世界だということである。

物理主義それ自体の正確な定式化はここでは手に余るので、簡単に、それは以下の2つのテーゼから成るということだけを確認しておこう。

- (i) いかなる個体も物理的個体である（つまり、非物理的個体としての靈魂や精神は存在しない）。
- (ii) いかなる性質も物理的性質にスーパーヴィーンする（つまり、基盤性質である物理的性質が同一なら、対応する心的性質も必ず同一である）。

つまり、われわれの世界は、物理的な事実が決まれば、それ以外の仕方の特徴づけられるすべての事実が決まるような世界である。例えば、典型的に同一の脳状態には典型的に同一の心理状態が対応する。また、典型的に同一の環境世界の中での典型的に同一の身体行動には、典型的に同一の行為が対応する。したがって、ある心理状態「空港に行かなければならない」が実現するためには、それに典型的に対応する脳状態が実現すれば十分である。逆に、典型的に同一の脳状態が二つ実現しているにもかかわらず、一方の脳が心理状態「オムレツを作りたい」を実現し、もう一つの脳が「空港に行かなければならない」を実現する、というようなことはスーパーヴィーンニエンスの原理からして不可能である⁵。

しかし、ロボットは工学的に作られる限り、人間と典型的に同じ脳状態をもつことは不可能である。すると、人間と同じ心理状態をもつようなロボットは作ることが不可能なのだろうか。幸いなことに、スーパーヴィーンニエンス関係は多重実現（multiple realization）を許す。つまり、典型的に同一の心理状態を、典型的に異なる複数の物理状態によって実現できるのである。実はこれこそが、人工知能やロボットを作成するための前提である。というのも、いかにぼんやりしたものとはいえ、われわれと「同じ」ような働きをする存在者を人工的に作り出すことができる、という確信がなかったなら、

今日のロボット作成へと続く努力はそもそも生まれなかつただろうからである。人間のさまざまな機能の人工的模倣は、新しい領域へと拡張されるその度に、この多重実現から「存在論的支援」を受けていたとすることができる。

しかし、この多重実現は、因果的な手段によって実現することのできる機能については存在論的に強力な論証を与えることができるが、〈感じられる限りでの感覚〉、いわゆるクオリアや現象的意識については論争の余地がかなりある⁶。機能的に人間と「同じ」ロボットが作られても、彼らが本当は快樂も苦痛も感じていないということが真実なら、例えば、彼らを功利原理の適用対象者として扱うのにどのような根拠があるのだろうか。するとここで、理性なき感受的存在者（下等動物など）とは別の問題を、感受なき理性的存在者（ロボット）は提起するように思われるかもしれない。しかし、現実世界が、機能的に人間と同型のロボットを作ることを許す世界であるなら、ロボットがクオリアなきゾンビ・ロボットでしかありえないかどうかは、物理主義を受け入れる前の状態における、いわゆる哲学的な他我認識問題と同じ構造をしているとすることができる。したがって、その問いは、「純粹に認識論的な問題」としての他我認識問題が解決不能であるように、解決不能である。しかし、われわれは、他我認識問題の困難さをまったく意に介さずに、お互いに同一の権利・義務を有する共同体を形成してきた。つまり、われわれが他人を同じ共同体のメンバーとなしてきた根拠は、他者の心的状態に関する認識論的な確認ではなく、実践的に動機づけられた〈存在論的な決断〉なのである。

それゆえ、人間と機能的に同型のロボットが人間とく同じクオリアや現象的意識をもっているのかどうかは、なお哲学的に重要な議論のテーマであり続けるとしても、ロボットが人間と一緒に共同体を作れるかどうかには決定的な重要性をもたない。むしろ問題は、ロボットが人間と機能的に同型でなく、それを超える機能を持ったときに人間と一緒に共同体をどのように作れるか、ということの方であろう。

そこで先へ進む前に、このような機能的ロボットを可能にする物理主義の世界で人間たちにかなることが生じうるかを少しだけイメージしてみよう。現在、医療や倫理の領域で問題になっているエンハンスメント（enhancement）の目的は、さまざまな手段とレベルで人間の機能を強化し、最終的な理想としては、若いときの健康と強さを保ったままいつまでも長生きすること（不老不死）である。人間の生物学的な構造の完全な解明はもちろん遠い将来のことであろうが、物理主義の世界では原理的に、生物学的なレベルの現象はすべて物理的なレベルの現象によって決定される。したがって、原理的には、物理的に操

作可能であればいかなる現象も実現可能であろう。しかし、もちろん、それぞれのレベルの現象はそのレベルの法則によって支配されている。したがって、生物体としての人間の（エンハンスメントとしての）変容可能性には生物学的な限界があるのはもちろんである。そのことを念頭に置きつつ、いま考えられているエンハンスメントの例を挙げてみよう。まず、エンハンスメントが「治療を越えて」進むとき、治療されるべき状態をできるだけ＜根本的に＞改善しようとする方向に向かうだろう。認知症の改善に治療薬が用いられたあとは、脳内の神経細胞の再生・増加が試みられるだろう。また、自閉症を含む発達障害や精神障害に対しては、現在、話題になっているオキシトシンなどの薬物治療も行われるだろうが、脳内の部位に関する神経組織上の原因が特定されればそれを脳外科的に治療するようになると思われる。脳神経学的な治療は、当然、出生前の遺伝子診断にまで拡大され、そもそもダウン症候群などの先天性障害をもった子供が生まれないように、遺伝子や染色体に対する遺伝工学的な技術が用いられることだろう。ここから「治療を超えて」われわれが行き着く先は、より望ましい形質、例えば、人並み以上の知性や運動能力、人並み以上の体格や容姿、強い意志と豊かな感情、誠実さと明るい性格などを持ったくより望ましい赤ん坊、＜完全な子供＞をどの親もが持つことである。すでに存在している「不完全な」大人たちも、この点では例外ではない。筋肉の衰えを防ぐための生物工学的な処置や、免疫システムの機能維持、成人病の克服、体型を含めた容姿の改善がふつうに求められ、誰もが＜完全な大人＞になろうとするだろう。そして、最終的には、遺伝工学的な手段を駆使した老化の決定的な阻止、果てしない寿命の延長が目指されるに違いない。この結果は、われわれ人類がこれまで経験したことのない「自然状態」、「生存条件」の出現である。認知能力を例に取れば、誰もが秀才もしくは天才となる可能性が極めて高い。確かに、その高い認知能力はそれほど画一的なものではないかもしれない、それでも、バラツキはきわめて小さな範囲に収まるだろう。さらにそれは、体格や容姿に関しても同様である。その結果、個性についてのわれわれの理解は現在とは大きく異なってくるはずだ。戯画化していえば、世の中、美男美女の天才ばかりになるかもしれない。＜完全な人間＞は必ずしもまったく同一の性質の集合体を意味するわけではないが、かなり似たようなタイプの人間だらけになるのはほぼ間違いないであろう。ロシアのある小説家の言葉のように、人々が不幸である理由は千差万別だが、＜幸福のかたち＞は誰にとっても同一であるからだ。

H.L.A. ハートは、その著『法の概念』のなかで、人間が生存という目的の下にあると

した場合に、法と道徳がなぜ特定の内容を含まなければならないのかについて述べている。法と道徳がもつ最低限の内容は、人間に課されている偶然的な自然条件に由来する。つまりわれわれの自然的な条件が、人間が生存するために必要な特定のルールを要求し、それなくしては「生存という最小限の目的をも推し進めることはできない」（ハート、p.211）最低限の法と道徳になるわけである。ハートは、その自然的条件を以下のように挙げている（Cf. ハート、pp.212ff）。

(a) 人間の傷つきやすさ

法と道徳が行為の積極的な要求ではなく行為の禁止という消極的な命令を基本とするのは、人間はときに他人に身体的危害を加えることがあり、また攻撃されれば傷つきやすいという事実があるからである。

(b) おおよその平等

人間の肉体的な強さ、機敏さなどはさまざまに異なるが、誰も他人の協力なしに長期間、誰かを服従させるほど抜きんで強くはない。

(c) かぎられた利他主義

人間は悪魔でもないし天使でもない中間的な存在であるから、相互自製のルールが必要でもあり可能でもある。

(d) かぎられた資源

人間が生存するために必要な食物、衣服、居住などの資源は無尽蔵にあるわけではないがゆえに、最小限の所有制度およびそれを尊重するための特別なルールが不可欠となる。

(e) かぎられた理解力と意志の強さ

相互自製のルールがもたらす利益は長期的には誰にも明白であるが、それらを理解しそれらに持続的に従うための人間の力は万全ではない。それがために理性は、ルールへの自発的な服従を保証するための強制的な体系を要求する。

これらは、あくまでもわれわれ人間がこれまで置かれていた偶然的な事情であって、人間の必然的な条件ではない。したがって、先ほど見たように、物理主義的な世界においては、今後、これらの条件が大きく変わる可能性がある。少なくともロボットたちは、易々とこれらの自然的な制約条件を踏み越えて行くだろう。そのとき、どのような倫理が必要とされるのであろうか。そのことを多少なりとも明確にするために、次の節では、逆に、

われわれにおいて倫理を現にあるような仕方でも可能としている条件の一つに触れておこう。ハートの五つの条件は、われわれがなぜ現にあるような倫理を必要とするのかについての示唆を与えるものであった。次節の自閉症者の倫理において示されるのは、現にあるわれわれの倫理を可能とする条件としての「他者の心の理解」である。

3. 自閉症者の倫理

自閉症者は現在、それに関連する遺伝学的な探索も含めてさまざまな原因究明がなされているが、決定的なことはまだ分かっていない。また、自閉症が、単一の症状を呈するのではなく、言語発達や知能の遅れがあるタイプからノーベル賞を取るような才能に恵まれたアスペルガーのタイプまで、幅広い症状の連続性（スペクトラム）を形成するのも大きな特徴である。本節では主に、デボラ・バーンバウムによって最近出版された注目すべき著書『自閉症の倫理』（Barnbaum, 2008）に沿って議論を進めたいと思う⁷。自閉症者の特徴の中でわれわれにとって一番問題となるのは、自閉症者が他人の志向的状态（intentional states）を自分自身の志向的状态と異なるものとして認識できない、という点である。例えば自閉症者は、サイコパス（精神病質者）と異なり、一旦気づけば他人の苦境にまったく無関心であるということはないものの、そもそも他人の気持ちを察することができないと言われる。これをいわゆる「心の理論（theory of mind）」の機能不全と捉えるのは自然であろう。その場合、自閉症者は自分の「心の理論」がうまく機能しないので、他人にそもそも志向的状态を帰属させることが困難になる。あるいは別の見方では、この欠陥のせいで、自閉症者は自分の志向的状态を誤って他のすべての行為者に帰属させてしまう。いずれにせよ自閉症者の多くは、すでにいろいろな点で有名なあの誤信念課題をクリアすることができない⁸。もちろん、問題は信念だけではない。他人が心を持つと認識することは、他人が独立した心的状態、例えば、その人自身の欲求、好み、感情などのあらゆる種類の志向的态度をもつ、自分とは別的人格だと認識することである。

自閉症のこの独特の特徴を説明すると称するのは「心の理論」説だけではない。バーンバウムによれば、他に有望なものとして、「中心的整合性の劣化 Weak Central Coherence」説と「実行機能の劣化 Weak Executive Function」説がある。それらの詳細にはここでは触れないが、「心の理論」説に比べて、前者は、なぜ自閉症者がしばしば全体よりもあまり重要な意味をもたない部分に固執してしまうのかということをもよりよく説明し、後者は、なぜ自閉症者がしばしば常同的で反復的な行動にこだわってしまうのかと

いうことをよりよく説明するように思われる。しかし、バーンバウムが言うように、これらの3つの説明は互いに矛盾し合うわけではない。というのも、それらのどれもが、自閉症全体の首尾一貫した原因説明を提供しているというよりも、自閉症の特徴を仮定上の認知機能によって「再記述」している段階だからである。したがって、それらは各々がそれなりに自閉症のある特徴群とうまくフィットするが、すべての症状をその発生メカニズムから説明しているわけではない。そのように見るならば、自閉症者の最も本質的な特徴を最もよく「記述」しているのは、やはり「心の理論」説だということができよう。

さて、「他人にその人の心的生活があること、そしてそれが何であるかを本当の意味では認識できない」という人々にとって、倫理とは何だろうか。この種の問題は、どのような特徴を備えた存在が人格 (persons)、もしくは道徳的共同体のメンバーたりうるか、ということとして議論される。自閉症者は非自閉症者と同じ道徳的共同体のメンバーとなりうるのか？ バーンバウムはヌスバウム (Nussbaum)、スキャンロン (Scanlon)、パーフィット (Parfit) らの議論を検討した上で、もっとも極端なベン (Benn) の見解を退ける。道徳的共同体のメンバーシップの議論は、常に、メンバーシップの必要条件を定義したとたん、それを満たさない個人やグループを道徳的共同体から除外するという論理的含意をもつ。ベンが提示したのは、まさに、自閉症者を非自閉症者の道徳的共同体から追放する議論であった。バーンバウムよれば、ベンの議論は、よき人生、人の幸福とは人が他人と結ぶ関係のうちにあるというヌスバウムらの論点を、道徳的共同体のメンバーシップを構成するための必要条件としたものである (Barnbaum, pp. 92f)。それによれば、自閉症者は、「心の理論」を欠くがゆえに他人に対する志向的態度を取りえず、道徳的共同体のメンバーにはなれない。彼らは、生物学的には人間であっても、道徳的な意味での人格ではなく、道徳的共同体の外部に位置する。ベンの言い回しでは、反応的態度 (reactive attitudes) を取りうるものだけが他者の反応的態度の対象たりうるが、自閉症者は反応的態度を取ることができない。反応的態度とは怒りや不満や誇りのような情動であり、それらは怒りや不満や誇りを持ちうる者以外に向けられても意味がない。ところが、自閉症者は、「心の理論」(他者の心の理解) を欠いているので情動的状況においても適切な反応をすることができない(社会的もしくは情動的な相互性の欠如)。しかしベンによれば、反応的態度を持つことができ、それゆえ他者の反応的態度の対象たりうる者だけが、道徳的共同体のメンバーなのである。したがって、自閉症者は道徳的共同体のメンバーではない。

バーンバウムがこのベンの議論を退ける根拠は何だろうか。その根拠は基本的に帰結主

義的なものであるように思われる。ベンのように自閉症者を非自閉症者から切り離すことは、逆に、非自閉症者を自閉症者から切り離すことである。その結果は、第一に「道徳的に不正な行為」が出現し、第二に「自閉症者と非自閉症者の両方の道徳的地位が墮落する」(Barnbaum, p.101)。ナチや旧日本軍の大量虐殺が示しているように、ある人間たちが他の人間たちを自分たちの道徳的共同体から締め出すとき、恐るべき結果が生じうる。そして、「誰にしても、その道徳的な地位は、他の人間が道徳的考慮に値しないということを是認したときにはいつでも損なわれる」(ibid., p.102)。それゆえ、「もしわれわれが間違っていた場合に支払うべきコストはきわめて高いがゆえに、誰が道徳的共同体のメンバーであるべきかを決定する際には、できるだけ範囲を広げるべきである」(ibid)。だがそれはまさしく、倫理的顧慮とは相互性 (reciprocity) を本質とするがために自分たちにも及ぶ悪しき結果を避けたい、という「利己的」な理由からではないのか？ バーンバウムはそれでよいと答える。その議論は興味深いのだが、結局、その「利己的」な理由を実際に効果あらしめるのは、あくまで自閉症者と非自閉症者がともに人間という同一の種に属し、日常で多くの利害と生活を共有しているという自然的な条件に他ならないように見える。それが証拠に、バーンバウムは道徳的共同体のメンバーに鳥や魚や獣といった動物を含めようなどとまったく考えていないと思われるからである。したがって、もちろんロボットも彼女の念頭にはないであろう。つまり、道徳的共同体の範囲の最大域を決定しているのは、おおよそ、先にハートが指摘した5つの偶然的な自然条件である。このことは、自閉症者と非自閉症者を分かつ深い谷を埋めるような、両者に共通して適用可能な倫理というものがそう容易くは見つからないことに対する彼女の困惑となって現れている。彼女の見たところ、同一の自然条件の下にあるにもかかわらず、両者の世界はきわめて異質なのである。とすれば、自閉症者の倫理を超えて、確実にハートの自然条件群を逸脱するロボットと人間をともに補足するような倫理を見出す、あるいは案出しようとするなら、それはなおさらに難しいこととなるだろう。

自閉症者と非自閉症をともに包括するような倫理を探すために、バーンバウムは「どの倫理学説が正しいのか」と問うのではなく、その手前で「どの倫理学説が両者ともに適応可能なのか」と問う (ibid., pp.111f)。というのも、たとえ真なる倫理学説があったとしても、例えば認知能力不足のせいでそれが要求することを誰も知ることができないとすれば、それは「<すべし>は<できる>を含意する」という道徳的原理を踏み外していることになるからである。その場合、その倫理学説が示す正しい行為と不正な行為の区別を道徳的

共同体の誰も知りえないのだから、その倫理学説の教えは彼らにとって実行不可能である。ところで、これまでのさまざまな観察から、自閉症者は、そもそも道徳的センスを持っていないのではないか、道徳的ジレンマを理解できないのではないか、道徳的問いとほかの種類の問いを区別できないのではないか、つまり道徳不感症なのではないか、と疑われているのである。つまり、自閉症者にとって実行可能な倫理学説はそもそもあるのか、という疑いがかけられているのである。もしその答えが自閉症者にとって不幸なものであったなら、彼らが進んで従う倫理的導きは存在しないということになるかもしれない。

バーンバウムの結論から言えば、ヒュームやカントの理論、あるいは J. ダンシーの個別主義 (particularism) や W. D. ロスの一応の義務 (prima facie duties) のいずれもが、自閉症者の特異性に阻まれて彼らを適用対象とすることができない。つまり、自閉症者と非自閉症者にまたがって両者に共有される倫理学説の可能性は見込みがないのである。では、自閉症者のどのような特異性が、それらの倫理学説の適用に困難をもたらすのか。ヒュームとカントに関する彼女の議論の一部だけを以下、簡単に見てみよう (ibid., pp.114ff)。

ヒュームによれば、道徳性は判断されるものと言うよりは感じられるものである。道徳は感性 (sentiment) によって決定される。とくに一つの情動、同情 (sympathy) がヒュームの道徳理論の核心である。しかし、同情もしくは共感 (empathy) は他人が志向的状态をもつこと、しかもそれが自分自身の志向的状态と違ったものでありうることの認識を必要とする。これは、単に他者についての信念を形成することではなく、他人がもっている信念を認識することを意味する。しかし、これこそ「心の理論」の欠落のゆえに、自閉症者にとって乗り越えがたい壁となる。他人の心的状態に接近するわれわれの方法の説明としては、「心の理論」説とシミュレーション説の2つがあり、それらの違いは問題の状況にある他人の心的状態を「心の理論」を用いて予測するか、あるいはその状況にあったとしたら自分もつであらう心的状態をシミュレーションによって生成しそれを他人に移入するかの違いである。われわれがどちらのメカニズムによって他人の心を理解しているのかはまだ決着がついていないが、自閉症者は、どちらのメカニズムも用いることができない。とくに自閉症者は、共感を自ら進んでもつことができない。この欠落は道徳的な無関心を結果する。したがって、同情もしくは共感なくしては、行為者はヒュームの道徳理論に自ら従って (正しい行為であれ不正な行為であれ) 行為することができないのである。

他方、カントの倫理学説は道徳的に正しい行為をするために共感などの情動をむしろ退

けるのであるから、一見して、自閉症者に適応可能な理論となっているように思われるであろう。実際、カントは、行為が道徳的価値を持つのは義務の認識からそれがなされる場合のみであり、同情、哀れみ、愛などの他の動機は道徳的価値のためには必要でないばかりか、助けにさえならないと考えている。したがって、他人への共感（熱い方法論）に問題がある自閉症者は、行為選択においてルール支配の側面が強いカントの倫理学説（冷たい方法論）を用いるなら、ある状況における義務の認識に導かれて正しい行為を行うことができるであろう。しかし、バーンバウムによれば、実は事はそう簡単ではない。例えば、カントの定言命法（の少なくとも幾つかのヴァージョン）は、心の理論を欠く自閉症者には効き目が無い。カントの定言命法の一つは、「あなた自身の人格ならびに他のすべての人の人格に例外なく存するところの人間性を、いつでもまたいかなる場合にも同時に目的として使用し、決して単なる手段として使用してはならない」（Kant (b), p.61, 邦訳, p.103）というものである。カントは、道徳的共同体のメンバーは誰か（の目的）にとつての主観的（手段的）価値として存在しているのではなく、他の存在物がそのために存在する、それ自身における客観的価値として存在すると考えていた。したがって、メンバーはみなそれ自身において目的なのであって、決して他の存在のための手段ではない。しかし、自閉症者は、他人を何かの目的に対する手段としてではなく、それ自身における目的として扱うことができるのだろうか。私自身の観点から離れて他人がそれ自身において目的であるということを認識することは、その他人に彼自身の自律した観点を想定することであり、それは、他人が私とは別個の〈目的-手段〉関連、つまり世界に対する志向的態度をもっているということを認識することだ。しかし、すでに何度も見たように、自閉症者は、他人に対して自分とは異なる志向的態度を帰属させることが極めて困難であり、その結果、他人が彼固有の志向的態度の起点であること、つまり〈それ自身における目的〉であることを認識することができないのであった。かくして、カントの道徳理論もまた、自閉症者の特異性とうまく折り合いをつけることができない。バーンバウムは、カント的な理論が自閉症者にうまく適用できないということは、自閉症者を道徳的共同体から締め出すのではなく、むしろカント的理論を拒絶する理由になるかもしれない、と述べている (ibid. p.130)。

4. 結語

自閉症者の倫理およびロボットの倫理がどのようなものとなるのかに関して、これまでの議論からは残念ながら決定的なことは何も言えない。しかし、それらの構築に向けて幾

つかの論点を確認しておこう。

まず、われわれの新クルーソーの状況が示唆しているように、倫理は他者たちと同等の権利をもつ共同体が存在しなければ意味がない（共同体テーゼ）。そのうえで、「なしたいこと」とは別の内容の「なすべきこと」が発生するのは、その共同体の内部での行為の調整の結果である。したがって、＜自由の欠如態としての倫理＞という側面からすれば、倫理はできるだけ共同体のメンバーの「なしたいこと」、つまり自由の実現を尊重するものが望ましい。

第二に、われわれの倫理が現にあるような内容をもつのは、ハートの指摘したような幾つかの偶然的な自然条件のおかげである。しかし、この自然条件は、今後の自律ロボット出現の前提として考えるなら、物理主義的世界における改変可能な条件として捉え直さなければならない。そして物理主義的世界(現実世界)が可能とする新たな「自然」条件を利用すれば、人間たちにとって重要な価値の源泉であった条件、すなわち「偶然おける差異」、「個性」はますます平均的で画一的なものに「改良」されていくように思われる。ここにおいては、人間たちにおける「おおよその平等」や「かぎられた資源」などは、倫理や道徳の内容を規定する重要な意味をもちや失うかのように思われる。一見した「まったくの平等」や「無際限の資源」などは、人間にとっての生存の意味を変え、そのことによって倫理の意味を変えるかもしれない。しかし私には、その先を見通すことはまだできない。他方、ロボットたちにとっては、ハートの指摘した「自然条件」は大した制約条件にはならない。エンハンスメントの結果として持ちえた人間の最高の心的能力と同等以上の心的能力を持ちながら、人間より「傷つきにくく」耐久性があり、半永久的な「命」を持つロボットと人間は、いかなる道徳的共同体を一緒に形成することができるのだろうか。ここでは、理想を追い求めて脳を含めた身体を改造する人間たちとは別な意味で、「おおよその平等」は倫理の内容を規定する意味を失ってしまうように見える。能力や強さの点でバラバラでありうるロボットたちと人間たちもつ同等の権利・義務とは一体何なのであろうか。

他方、自閉症者の特異性が示すところによれば、道徳的共同体のメンバーとなるためにロボットたちもつべき能力の第一のものは、「心の理論」が与える能力である。これはもう少し一般的にいえば、人間たちもつ素朴心理学的能力である。「心の理論」が他人の心的状態の認識に果たしている役割、およびそれによって可能となる道徳的態度を形成する上での役割は、バーンバウムの議論においては少し誇張されているように思われる。しか

し、「心の理論」を中核とした素朴心理学のメカニズムをもしロボットがまったく持っていないとすれば、そのロボットは、他のロボットおよび人間たちと真に相互的な関係に立つことはできないだろうし、それゆえ真の道徳的共同体を形成することもできないだろう。なぜなら、そのロボットにとって、道徳的に顧慮すべき<他人の心>と言うべきものが存在しないからである。ロボットたち同士、およびロボットと人間たちの生きる条件が極めて異質であるという点は、自閉症者と非自閉症者の世界が互いに極めて異なっているということにおいて先駆的に示されている。しかし、その二つの世界にまたがった倫理がいかなる内容を持つのかに関して、バーンバウムもまだ結論を出していない。彼女が示したのは、二つの世界をつなぐ倫理がいかに困難であろうとも、自閉症者を道徳的共同体から除外してはならないということである。しかし、その結論は直観的には妥当だとしても、その最終的な理由が、「<同じ自然的条件>にある存在者を道徳的共同体から除外することは結果として除外した存在者の道徳性を墮落させる」ということである限り、その議論をロボットたちに適用することはできない。なぜなら、ロボットたちはまさに人間たちと<同じ自然条件>にあるとは限らないからである。

それゆえ、われわれがここで辿りつくことのできる最後の論点としては、ハートが述べたような自然的条件には基づかない倫理、つまり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を改めて制定することであろう。それは、ロボットたちも自閉症者も非自閉症者も、さらにそれ以外の特異な存在者のグループをも、彼らが最低限の素朴心理学的な他者理解を示す限りにおいて、同一の道徳的共同体に含める倫理である。その際、これでのところで私が掛け値なしに主張できることは、これらの異質なメンバーたちがどのような生存の目的、欲求、興味、好み、感情を持っていようと(あるいは持っていないとも)、他者と関係しながらともに生存する限りにおいて、最大限に他者の自由を尊重しなければならない、という古くからの自由の原則だけである。

(金沢大学人間社会学域人文学類教授、独立行政法人科学技術振興機構：RISTEX)

注

1 もし法則化可能性が、格率が「なすべきこと」となるための十分条件なら、例えば「手袋をはめるときは右手からはめろ」というような、およそ倫理的とは思われない無数の格率も「なすべきこと」となるであろうし、それと実質的に矛盾する「手袋をはめるときは左手からはめろ」を誰かが採用したとしても、それは法則化可能性のテストをパスするであろう。もっとも、法則化可能性は「なすべきこと」の必要条件にすぎないかもしれない。しかし、その場合でも、問題の十分条件を構成するため

- の追加条件は、カントの精神に沿う限り、格率の内容に関わる実質的条件ではなく、形式的条件であるだろう。
- 2 「なすべきこと」を構成するために法則化可能性に追加されるべき条件が何であれ、それが形式的条件である以上、新クルーザーが採用するいかなる格率も、「a かつ非 a をなせ」のような、そもそも行為を開始しえない矛盾したもの以外は、「なすべきこと」となる資格において区別をつけることはできないであろう。
 - 3 平尾は続けてこう述べる。「それ〔道徳〕はそもそも他人との関係においてのみ成立する。従って、自分自身にのみ関係する事柄は本来的には道徳的判断の対象たりえない」（平尾 p.10）強調は、本文の引用ともに原著者。
 - 4 例えば自殺は、「目的自体としての人間性という理念」と両立しないがゆえに定言命法に反する、というカントの主張はここでは検討していない。（Cf., Kant (b), p.61. 邦訳, p.104）
 - 5 厳密に言えば、このように素朴心理学的に概念化された心理状態は、当人の脳状態にローカルにはスーパーヴィーンしない。もっともそれは、解釈者の脳状態などを込みにした十分大きな物理的な時空領域にグローバルにスーパーヴィーンする、と言うことはできる。
 - 6 例えば、クオリアの多重実現と、機能に導かれた基盤性質（もしくは機能的性質）へのスーパーヴィーニエンスの擁護については、拙論（柴田 2008）を参照されたい。
 - 7 以下は、バーンバウムのまとめによるアメリカ精神医学協会（APS）の自閉症診断基準第4版（DSM-IV）であり、自閉症との診断が下るには、少なくとも（A）から2つ、（B）および（C）それぞれから1つ、全体で少なくとも6つの項目に当てはまる必要がある（Barnbaum, pp19f）。

(A) 少なくとも以下のうちの2つの項目によって示されるような、他人との社会的相互行為における質的な障害

- i. 目と目で見つめ合うことや顔の表情のような、多様な非言語的行動における障害
- ii. 発達水準にふさわしい仲間関係を作ることができない
- iii. 他人と興味や楽しみを共有しようと自発的に求めることがない
- iv. 社会的もしくは情動的な相互性の欠如

(B) コミュニケーションにおける質的な障害

- i. 話し言葉の発達の遅れ、もしくは欠如
- ii. 十分な会話ができるものの、会話を始めたり継続したりすることができない
- iii. 常同的で反復的な言葉の使用、もしくは本人だけの特異な言葉の使用
- iv. 発達水準にふさわしい、変化に富んだ「物まね遊び」や社会性のある「ごっこ遊び」を自発的にできない

(C) 行動、興味および活動の幅が制限され、反復的で常同的なパターンとなっている

- i. 異常な強さもしくは熱心さをもって、一つもしくは複数の興味パターンに没頭する
 - ii. 役に立たない習慣や儀式に強迫的にこだわる
 - iii. 常同的で反復的な特異な身体運動
 - iv. 物体の一部への執拗な没頭
- 8 誤信念課題には幾つかのヴァージョンがあるが、もっとも単純なものは以下のようなストーリーである。被験者である幼児は、サリーとアンがおはじきで遊んでいるところを見せられる。しばらく遊んでから、二人はおはじきをバスケットの中に入れてしまう。その後、サリーが部屋を出て行ってから、アンはそのおはじきを箱に入れ替える。そこで被験者はこう尋ねられる。「サリーが戻ってきて、おはじきがましくなったら、どこを探すでしょう?」「心の理論」を健全に発達させている子供は、サリーが最後におはじきを入れたところ、つまりバスケットを探すだろうと答える。しかし、自閉症の子供の多くは、自分が知っている正しい答え、つまり箱を探すだろうと答える。つまり「心の理論」は、他者の信念が偽であることの認識を可能とするのである。

なお翻訳がすでにある場合には基本的にそれに従ったが、一部、表現を変えたところもある。この場を借りて、翻訳を利用させていただいた邦訳者に感謝の意を表したいと思う。

文献

- アシモフ、I. 『われはロボット』早川書房、小尾芙佐訳
- Barnbaum, D. R., *The Ethics of Autism*, Indiana University Press, 2008.
- Baron-Cohen, S., *Mindblindness*, The MIT Press, 1995.
- Benn, P., Freedom, Resentment, and the Psychopath, *Philosophy, Psychiatry, & Psychology* 6(1): 29-39, 1999.
- Goldman, A. I., *Simulating Minds*, Oxford University Press, 2006.
- ハート、H. L. A., 『法の概念』みすず書房、矢崎・ほか訳、1976
- 平尾透、『功利性原理』法律文化社、1992.
- Kant, I.(a), *Kritik der Praktischen Vernunft*, Kant Werke, Band 6, Darmstadt, 1975.(カント『実践理性批判』岩波文庫、波多野・宮本訳、1927.)
- (b), *Grundlegung zur Metaphysik der Sitten*, Kant Werke, Band 6, Darmstadt, 1975.(カント『道徳形而上学原論』岩波文庫、篠田英雄訳、1960)
- カス、L. R., 『治療を越えて』青木書店、倉持武(監訳)、2003
- Mill, J. S., *Utilitarianism*, Collected Works of John Stuart Mill, Vol. X, University of Toronto Press, 1969.
- 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』所収、昭和堂、長滝・柴田・美濃(共編)、2008

なお、本論文は、科学研究費「認知ロボティクスの哲学」(代表：金沢大学・柴田正良)、および RISTEX 「自閉症にやさしい社会：共生と治療の調和の模索」(代表：金沢大学・大井学)の研究成果の一部を含んでいる。