

物理主義的世界とロボット倫理

メタデータ	言語: jpn 出版者: 公開日: 2017-10-03 キーワード (Ja): キーワード (En): 作成者: Shibata, Masayoshi メールアドレス: 所属:
URL	http://hdl.handle.net/2297/43254

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 International License.



物理主義的世界と ロボット倫理

柴田正良(金沢大学)

大阪大学グローバルCOE「認知脳理解に基づく未来工学創成」:
テーマ別創成塾「ロボット工学と倫理」第3回

於:大阪大学情報科学研究科 (吹田キャンパス)

Feb. 22, 2013

“TOWARD ROBOT ETHICS...”で 考えようとしたこと（1）

1. 物理主義的世界において、いかなる倫理が可能か？ 将来の倫理的眺望(an ethical landscape)を描くこと。
2. 道徳的共同体のメンバーシップ(十分条件の一つとしての素朴心理学的機能)とその周辺域の関係(<準・人格性>)
3. 倫理の存立根拠(共同体テーゼ)・・・他者がいなければ倫理など存在しない。

“TOWARD ROBOT ETHICS...”で 考えようとしたこと（2）

4. 自由(したいことができる)の欠如態でしかない倫理
5. 倫理的性質の非実在論・・・倫理的性質は物理的性質にスーパー
ヴィーンしない。
6. 異世界の者たち(ある種の自閉症者、ロボット、エイリアンなど)と共生
するための倫理原則・・・他者危害の原則
7. この原則以外は、われわれの広義の価値観、他者との付き合い方、
人生のスタイルに従って自由に選択しよう。

倫理はいかなる状況で発生するのか？

1. 現在も、未来も宇宙でまったく孤独に存在する者にとって、倫理とは何だろうか？
 2. 彼にとって、「なしたいこと」と内容上異なる「なすべきこと」はあるのだろうか？
 - 3 「なしたいことをなすこと」→プリミティブな自由
「なしたいこと」の相互調整 → 倫理
- ◆ この意味での倫理は自由の〈欠如態〉である。

ネオ・クルーソーにとっての倫理（？）

1. 例えば、たった一人の行為者しか存在しない可能世界で、その行為者にとって、「なしたいこと」に反する「なすべきこと」が、カント的な義務論の観点や、功利主義的な観点から導かれるのだろうか？
2. 平尾透とともにこう言いたい。「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、1992, p.10)
3. しかし(1)の議論はいつもうまくいかないなので、誰か倫理好きの人をお願いします。

共同体テーゼ（ネオ・クルーソー的 世界には倫理は存在しない）

複数のメンバーが属する共同体において、各メンバーが他のメンバーと〈同等の権利・義務〉を持つ（完全義務対象者同士として）・・・道徳的共同体

その周辺には、不完全義務対象者の集団が幾重にも、この共同体を取り巻いている（幼児、重度の認知的弱者、動物、ある種の自閉症者？）。



（歴史的／因果的生成）

共同体に属する複数の行為者が、それぞれの自由の行使（利害関心の追求）を相互に調整せざるをえない状況（必要条件）

人類における道徳的共同体の 自然的基盤（1）

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…(H. L. A. ハート『法の概念』、1976)

その5つの偶然的条件

1. 人間の傷つきやすさ

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。

2. おおよその平等

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て強くない。

人類における道徳的共同体の 自然的基盤（2）

3. 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。

4. 限られた資源

人間に必要な食物、衣服、居住などの資源は有限である。

5. 限られた理解力と意志の強さ

人間は、相互自制のルールがもたらす利益を見通し、長期に保持できる力を持たない。

自然的条件の変容（ENHANCEMENT）

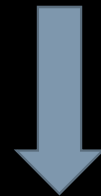
- 治療を超えたエンハンスメント(enhancement beyond therapy)
- 脳・遺伝子を含めた人体の改良は、「完全な赤ん坊」・「完全な大人」・「完全な人間」への強い願望を引き起こすだろう。



画一化された「優れた知性」・「強い意志」・「豊かな感情」・「優れた容姿」といった、いわば「美男美女の秀才たち」の世界。それは、従来の「個性と差異」という偶然的な自然条件を破壊する

自然的条件の変容（CYBORGS）

- 治療を超えたエンハンスメントは、やがて、老化の決定的な阻止、果てしない寿命の延長を目指すであろう。
- 壮健な身体のままにおける〈不老不死〉。



- 再生医療と手を携えて、〈サイボーグ化〉の技術がごく普通に人体に適用され、ハートの「人間の傷つきやすさ」や「おおよその平等」という条件は、その実質的意味を失うであろう

『自閉症の倫理学』（D. バーンバウム）

◆ 異世界の者との共生の倫理に関連する論点。

1. 自閉症者は、非自閉症者とまったく異なる世界に住む。しかし、両者は同じ倫理的共同体に属するとすべきである。
2. 自閉症者は、＜心の理論＞の欠損のゆえに、いかなる既存の道徳説も自ら進んで用いることができない。
3. 自閉症者と非自閉症者をともに補足しうる、新たな倫理学を構築すべきである。

キーコンセプトとしての 心の理論 (THEORY OF MIND)

◆バーンバウムは、倫理における自閉症者の根本的困難が<心の理論>の欠損からくると考える。

自閉症の原因を説明する3つの理論

1. 心の理論説(theory of mind thesis)
2. 中心性統合弱化説
(weak central coherence thesis)
3. 実行機能弱化説
(weak executive function thesis)

心の理論を欠くとは どういうことか？

1. 心の理論は、他者の志向性を自分とは違った自律したものとして理解し、自分と異なる他者の心的状態(欲求・信念・情動など)を推論するのを可能とする。



2. その欠損は、心的存在としての他者がまるで存在しないような世界を自閉症者に強いる。他者との相互承認的關係、相互信頼を築くことが極めて困難となる。

自閉症者は道徳的共同体の メンバーか？

- ホブソン(P. Hobson)は、他者との相互承認的關係は人格の本質的構成条件(必要条件)であり、その関係に入れない自閉症者は道徳的共同体の外側に位置する、と論ずる。
- ベン(P. Benn)によれば、反応的態度(reactive attitudes)を取りうる者だけが、他者の反応的態度の対象であり、道徳的共同体のメンバーであるがゆえに、自閉症者は道徳的共同体に属さない。

(「反応的態度」はストローソンの論文「自由と怒り Freedom and Resentment」に由来する。)

異世界の存在者との 共生の倫理（AUTISM-1）

- バーンバウムはホブソン、ベンの議論に対抗して、自閉症者を道徳的共同体に含めるべきだと主張する。その論拠は、「仮に排除が間違っていた場合の代償は大きい」という、いわば「利己的な利害計算」である。
- しかし、その論拠の根底には、自閉症者と非自閉症者が<ヒトという同一の種>に属し、日常の利害と生活を共有するという事実にあるだろう。
- 結局、ここでもハートの5条件が道徳的共同体の最大域を定めているように思われる。

しかし、これはもちろん、ロボットには通用しない。

異世界の存在者との 共生の倫理 (AUTISM-2)

1. むしろ、従来の自然的条件を前提せずに、＜道徳的共同体＞の概念から出発すべきである。したがって、そのメンバーにどんな存在者が含まれるかを、あらかじめ＜自然種＞を根拠に決定することはできない。

(ある種の自閉症者は、幼児や重度の認知的弱者とともに、この共同体の外側に位置すると思われる)。

2. しかし、同時に、この共同体を＜準・道徳的共同体＞群が幾重にも取り巻いており、そこに不完全義務の対象者としてのさまざまな存在者が位置し、彼らには、さまざまな内容の＜準・道徳的配慮＞が保証されるべきである(この最後の「べき」は、現在のわれわれの文化・社会的態度(われわれの生存のスタイル)の表明である)。

自閉症者には、自分が運用できる道徳理論がない (AUTISM-3)

- カントの義務論は、ルール of 盲目的遵守によって自閉症者にも運用可能に見えるが、「他者をそれ自体における目的とせよ」という理解でつまずく。
- ヒュームの理論の核心にある同情 (sympathy) も、あるいはもっと一般的な意味での共感 (empathy) も、自閉症者に手の届く能力ではない。
- バーンバウムによれば、他の従来の道徳理論のいずれも、自閉症者が自ら進んで従いうる理論ではない。

▼ つまり、自閉症者と非自閉症者をともに適用範囲とする道徳理論は、今のところ存在しない。

自閉症者と非自閉症者を包む倫理は可能か？（AUTISM-4）

可能ではないかもしれないが、＜準・倫理的なシステム＞なら創れる。

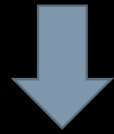
1. 自閉症者が道徳的共同体のメンバーでないなら、＜同等の権利・義務＞を相互に承認し合うことを基礎とする倫理は、彼らを包括できない。（逆にサイコパスは、潜在的な反道徳者として包括されるだろう。）

2. しかし、彼らや胎児、幼児、重度の認知的弱者など、さらにはペットや野生動物たちを含めた＜準・道徳的共同体＞がその外側に存在し、そこでは＜準・倫理的な配慮と処遇＞のシステムを、創意と工夫によって、新たに構築することができるだろう。

異世界の存在者との 共生の倫理 (ROBOTS-1)

- ロボットの自律 (autonomy) の意味

「自らの内部状態から意図的判断を下すことが可能であり、
いかなる規則も疑い、拒絶することができる。」



自律ロボットは、欲求と信念を内容とする態度を持つという点で、最小限の素朴心理学的メカニズム (folk psychological mechanism) を持つ。

異世界の存在者との 共生の倫理 (ROBOTS-2)

- ロボットたちは、われわれの物理主義的世界では、「同一の物理的状态 → 同一の心的状態」というスーパーヴィーニエンス原理に基づき、いくらでもタイプの同一の「心」を備えて出現可能である。
- したがってロボットは、「誕生」・「病」・「治療」・「繁殖」・「死」などの生存条件において、われわれとはまったく異なる。彼ら相互の間での「同等の権利・義務」とは何か？ 彼らとわれわれの間での「同等の権利・義務」とは何か？

異世界の存在者との 共生の倫理（CONCLUSIONS-1）

1. 少なくとも最小限の「心の理論」をロボットが持つなら、倫理的共同体に属するのに十分である（必要条件でもある？）。感受性と反応的態度の程度によって心的存在者のタイプが判別され、認知能力の程度によって理性的存在者のタイプが判別される。
2. 異世界の者たちの生存条件が互いにあまりに異なるので、これまでの人類の倫理・法システムのように、生存条件でその内容を決めることはできない。むしろ新たな<倫理システム>と<準・倫理システム>を、人為的かつ自覚的に、制定する必要があるだろう。

異世界の存在者との 共生の倫理（CONCLUSIONS-2）

3. サイボーグになり、やがて＜機械＞に進化した人間の「アイデンティティ」はどこにあるのか？ 素朴心理学の提供する枠組みだけかもしれない。

4. 倫理を自由の＜欠如態＞と考えるなら、理想的状態は、倫理ができるだけ出番を減らすことである。異質なメンバーがどんな目的、欲求、興味、好み、感情を持っていようと、それを最大限に尊重するようなシステム。

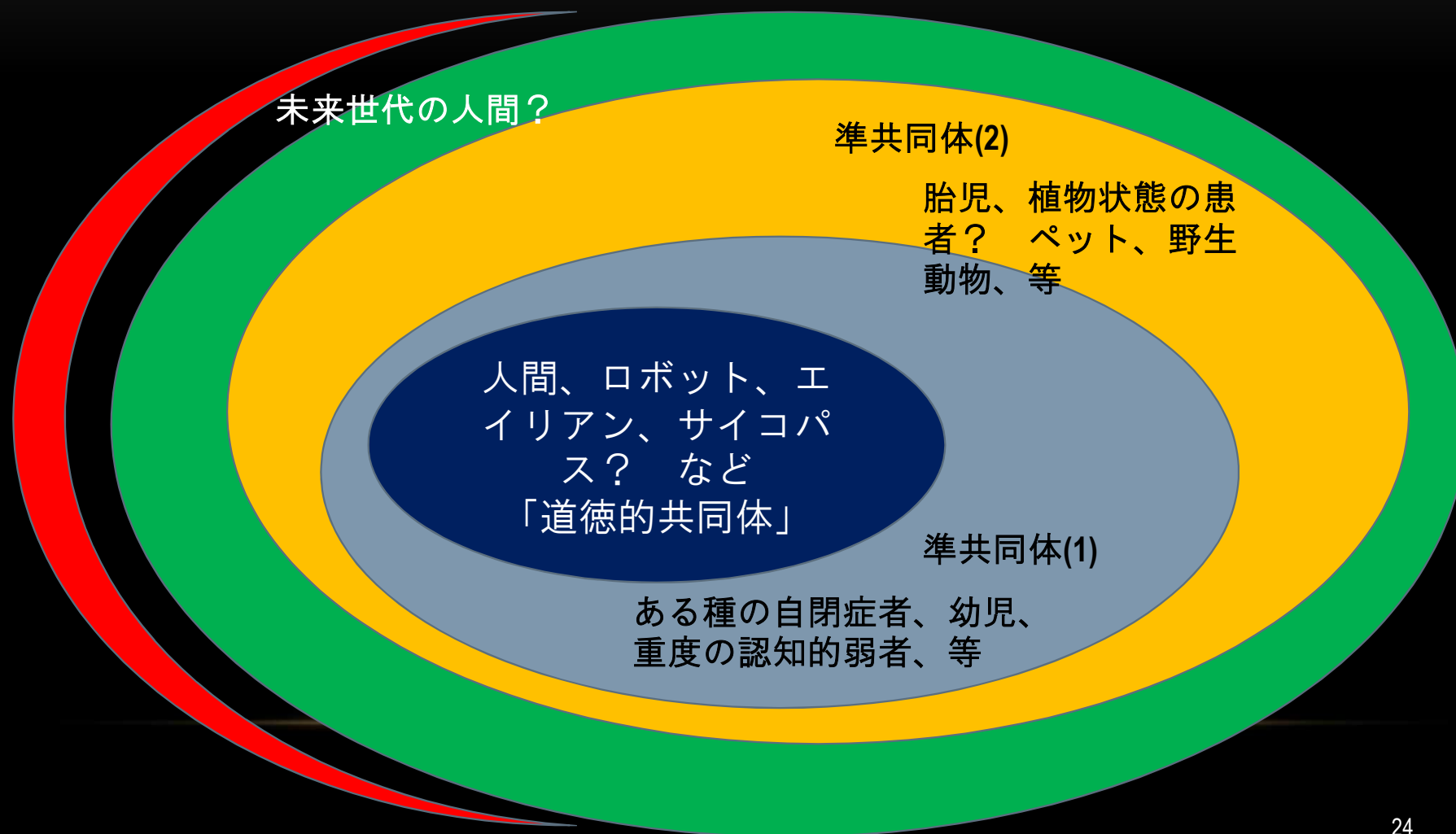
政治哲学的な意味での＜リバタリアンの自由＞

異世界の存在者との 共生の倫理（CONCLUSIONS-3）

5. 少なくとも、常に最上の人間と同等以上の知力・体力・耐久性・再生可能性を持つサイボーグやロボットは、ふつうの人間とケアの倫理や介護の倫理で結ばれることにはならないだろう。なぜなら、互いが、自律的な心的存在者としての自由と能力を承認しているのだから。

6. すると極めて皮肉なことに、異世界の存在者との共生の倫理において妥当するのは、差し当たり、他者に危害を加えない限り何をしても許される、という古典的な「他者危害の原則 harm to others」しかない。他にいかなる義務と権利があるのか・・・私にはまだ分からない。

道徳的共同体と 準・道徳的共同体のイメージ(FINAL)



文献

- Barnbaum, D. R., *The Ethics of Autism*, Indiana University Press, 2008.
『自閉症の倫理学』(柴田・他監訳)、勁草書房、2013(出版予定)
- Benn, P., Freedom, Resentment, and the Psychopath, *Philosophy, Psychiatry, & Psychology* 6(1): 29–39, 1999.
- カス、L. R., 『治療を越えて』青木書店、倉持武(監訳)、2003.
- ハート、H. L. A., 『法の概念』みすず書房、矢崎・他訳、1976.
- 平尾透、『功利性原理』法律文化社、1992.
- 柴田正良、「異世界の者たちの倫理」『哲学・人間学論叢』創刊号、金沢大学哲学人間学研究会、pp.17–37, 2010.
- Shibata, M., “Toward robot ethics through *the Ethics of Autism*”, in J. L. Krichmar and H. Wagatsuma (eds.), *Neuromorphic and Brain-Based Robots*, Cambridge University Press, pp. 345–361, 2011.
- ストローソン、P, F., 「自由と怒り」、『自由と行為の哲学』(門脇・野矢監修)、春秋社、2010.