ロボットにも「行為者性」を

柴田正良(金沢大学理事・副学長)

JST/RISTEX「人と情報のエコシステム」研究開発領域
「人工知能の哲学2.0の構築に向けて」
於：東京大学駒場キャンパス
March 23, 2019

目次らしきもの

1. AIとロボットの哲学1.0の当時（およそ20年前）から見ると、現在のAIブームはどう見えるか？

2. 残された課題の数々
   新しい「認知科学の哲学」を目指して

3. それとは別の、われわれの進む道
   道徳的行為者としてのロボット

4. 哲学者の仕事（余裕があれば、２分で）

現在のAIやロボット第３次ブームは当然予想されたことだが、

- しかし、第２次ブームの頃の問題がすべて解決されたようには見えないし、根本的な技術革新が起きたようにも見えない。
- したがって、このブームの頭打ちが意外と早く到来するかもしれない。

- 現在の単線的な技術の発展で、より「人間らしいタッチ」に目覚めた人々の要求に応えられるか？
- さらに、一時的であれ生ずる「雇用喪失問題」とそれに端を発する未来の不安に、誰が、どう答えるのか？

哲学にとって、このブームの問題は

1. 「自律した一つの個別科学」として認知科学が成立しうるのかどうか、という問いに答えを与えていないこと。

   人とAIに共通の認知機能の科学として、認知科学はそれ固有の「理論的存在」と「理論法則」を措定し、それによって認知現象を説明できるのか？

2. また、ある段階を超えたロボットがいかに人類と共生できるのか、という問いに答えを与えていないこと。

   物理的世界におけるロボットの存在論的地位と同時に社会における倫理的地位の問題が見通せていない。

あの問題はどうなった？

1．AIは、統語論的操作に加えて、意味論的理解をすることができるのか？

2．「思考の言語」は存在するのか？　そもそも心的計算の対象である心的表象は存在するのか？
フォーダー VS. チャーチランド

3．ヒトの脳の計算方式は何か？　AIはそれと同じ方式を採りうるか／採るべきか？
古典的計算主義 VS. コネクショニズム

4．いわゆる「フレーム問題」は解決されたのか？　AIは常識がもつ知識（信念）の「蓄積と活性化」能力を持ちうるのか？　持ちうるとしたら、どのようなメカニズムでか？

5．AIは、モジュールとしての複数機能からの情報をいかにして統合しうるのか？　中央演算処理の謎。

6．AIにおける学習は、いかなるタイプの機械学習が最適か？　それは、ニューラルネットワークの重みづけ学習と同一か？　そして、ヒトの脳の学習のメカニズムは？

7．AIは、どのようにして、どの程度、帰納的推論のような一般化能力を持ちうるのか？

８．感情の機能とは一体何か？　それを、ロボットにいかにして実装させるのか？　また実装させるべきか？

９．心理学的な意味での意識、つまり機能としての意識はAIにおいていかにして実現できるのか？

１０．一人称的な意識、つまり現象的な意識は、AIにおいていかにして可能か？　またAIはクオリアを経験することができるか？

１１．AIは、いかにして価値判断、道徳判断と美的判断をすることができるか？

これらを解決する「人工知能２.0の哲学」を展開すべきである。それは、新しい「認知科学の哲学」となるだろう。

新しい「認知哲学」構築への道

日本の「心の哲学」の基本的立場は、還元的だろうと非還元的だろうと、現実世界は少なくとも「自然的スーパーヴィーニエンス」の成立する可能世界だ、という物理主義的前提は崩せないだろう。

１．「認知哲学」を、周辺諸科学との連携で展開する。脳神経科学、心理学、ロボット工学などとの積極的な問題交換。現象的意識とクオリアの問題は、閉じた議論にしかならないかもしれない。

２．AI・ロボットとヒトに共通の認知科学は可能か？現実世界は、素材レベルの多重実現にも拘わらず、機能が一つの「自然種」となるような可能世界なのか？その解答を、新「認知哲学」は与えるべきだろう。

それとは別の、「道徳的行為者としてのロボット」の哲学

この道を進むわれわれのこれまでの道のり（柴田科研）

1．「意識と感情をもつ認知システムについての哲学的研究」基盤（B）H16-18

2．「認知ロボティクスの哲学」基盤（B）H19-22

3．「意図的主体性のロボット的構築に向けて」 基盤（B）H23-26

4．「個性を持つロボットの制作による＜心と社会＞の哲学」基盤（B）H27-30

その道の進む先（長滝科研）

「道徳的行為者のロボット的構築による＜道徳の起源と未来＞に関する学際的探究」代表者：長滝祥司（中京大学）基盤 (A) H31-35 (申請中)

- 本研究は、ロボット工学や心理学などの経験的手法を取り入れつつ、ロボットのような新たな存在を道徳的行為者として受容できる社会を構想するための新たな道徳理論の主要テーゼを提言する。

- そこでは、道徳の起源にまで遡って、生死、善悪、生きる意味など、従来の人間の生存条件に基づく様々な超越的概念に根本的な検討を加える。

ロボットとの共生の倫理　1／3

１．少なくとも最小限の素朴心理学的メカニズムをロボットが持つなら、道徳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．ロボットたちの生存条件が人間とはあまりに異なるので、これまでの人類の道徳・法システムのようには、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＜道徳システム＞と＜準・道徳システム＞を、人為的かつ自覚的に、制定する必要があるだろう。

ロボットとの共生の倫理　2／3

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？　それは、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＜リバタリアン的自由＞

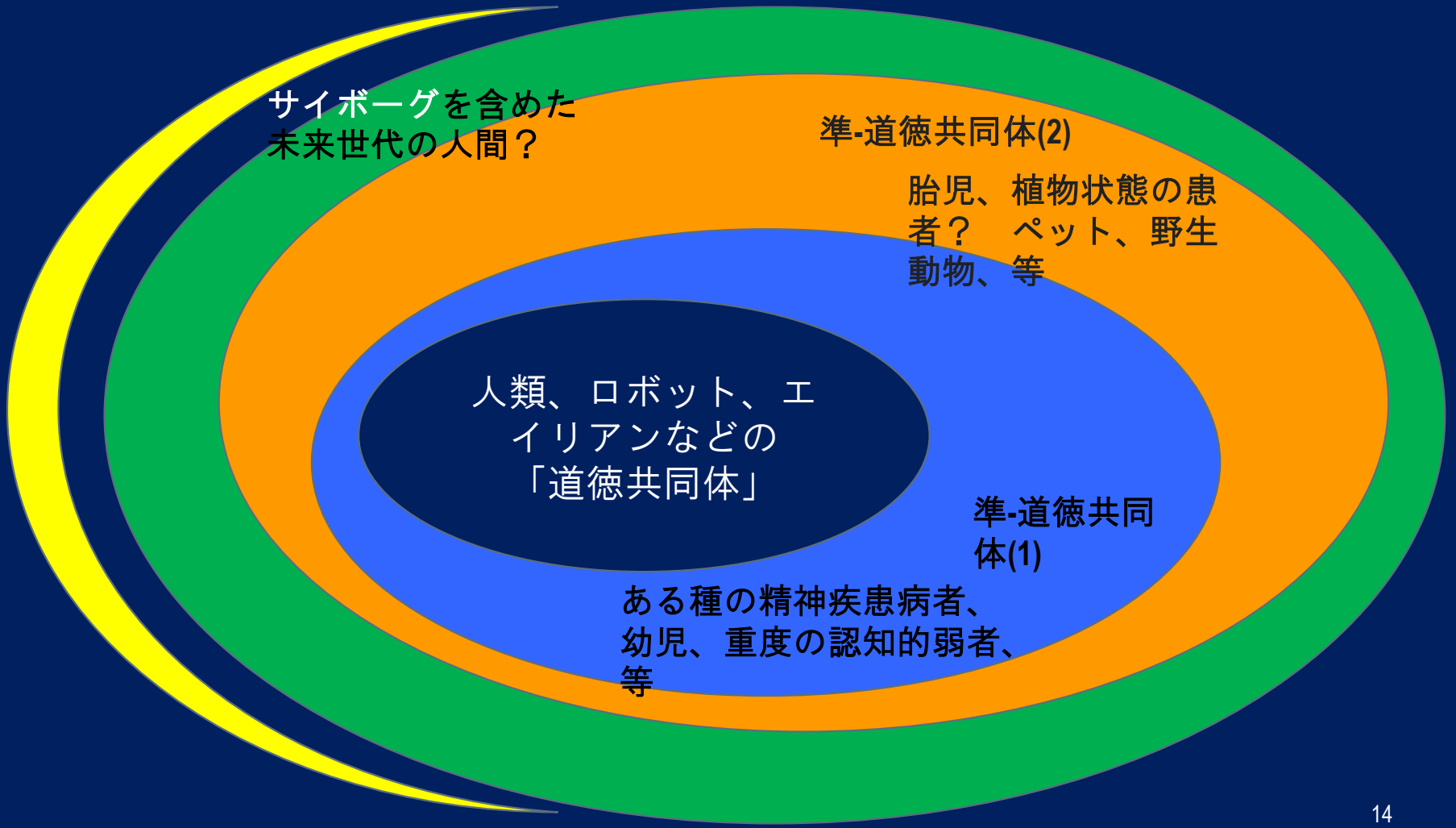
ロボットとの共生の倫理　3／3

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、未来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」（Principle of 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



おしまい

柴田の研究関連webサイト

http://siva.w3.kanazawa-u.ac.jp/