

「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？

(配布資料)

第1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、「心とは何か」「心を持つとはどういうことか」という問いへの取り組みが不可避。少なくとも平行して問われるべき。
- 「5分でわかる心の哲学・概略史」
 - デカルト的二元論: 「心身因果がある」「他者の心を知りうる」という直観により×。
 - 行動主義: 「振る舞いに表れない心的状態がある」という直観により×。
 - 心脳同一説: 「神経系を持たないものも心を持ちうる」という直観により×。
 - 機能主義: 中国人民全体が心的状態を持ちうることになる?
 - 解釈主義: 心身因果がうまく説明できなくなる? 一人称特権が危機に瀕する?

↓

1つの理論評価基準としての「**直観的理解と理論の合致**」を抽出

- ・ 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。
- ・ 直観は容易に改訂されないが、改訂不可能というわけではない。

第2節. 1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

- 課題を「フルスペックの心」から「痛みを感じられるロボット」へとダウングレード。
- 「痛みとは何か?」から「我々は痛みという語で何を意味しているか?」へシフト。
- 言語学習と「言語実践の階層的発展モデル」
 - ・ 言語は公共的間主観的状况を参照して学ばれる(心的語彙も同様)。
 - ・ 言語習得期には一人称報告も訂正されうる。
 - ・ 主観的感觉に基づく報告(with 一人称権威)は心的語彙の習得後に成立。

↓

階層的発展モデル： 共変関係を利用して使用基準の軸足をシフトさせる

【第1段階】 これこれの間主観的状况のもとで「痛み」という語を用いよ。

【第2段階】 第1段階習得後、内部センサ状態等との対応づけにより、
主観的感觉ベースの報告が可能に⇒使用条件の軸足がシフト。

【第3段階】 一人称報告と神経科学的状態との対応づけ⇒軸足をシフト

- 実装が容易な第一段階の言語実践に定位する「解釈主義的アプローチ」の採用。
⇒具体的にやるべき作業の輪郭を少しずつ絞り込んでいく。

第3節. チートの可能性と「擬人化の壁」

- なお残る3つの問題。(1)具体的な実装機能は？ (2)解釈主義の標準理論との乖離は？ (3)チートの可能性は？ ⇒「現象的スタンス」の導入で問題を縮減。

【解説： 志向的スタンスと現象的スタンス】

志向的スタンス

- ・ 欲求や信念といった心的状態は行動を予測・説明するためのもの
- ・ それらにより行動がうまく説明予測できる⇒「志向的システム／心を持つもの」
例：台所に行き冷蔵庫を開けるといふ父の行動は、「ビールを飲みたい」「冷蔵庫にビールがある」といふ心的状態の帰属によりうまく説明・予測ができる。
- ・ 目に見えない心的状態を公共的に確認できる所に引きずり出せている。

現象的スタンス

- ・ 痛み等の感觉状態は道徳的配慮実践と密接に関係
例：犬の足にトゲ ⇒ 「ワンワン痛いね」「よしよし」
超合金ロボを棒で叩く ⇒ 「壊れちゃうでしょ」
- ・ 道徳的配慮の対象になる⇒痛み等の感觉を持ちうるもの(現象的システム)

- 目標を「痛み」に典型的な input / output」と「道徳的配慮実践惹起機能」に。

第4節 メタ哲学的構図の確認

- 「理論と直観的理解の合致」といふ大枠の理論評価基準は保持
- ただし、「工学研究による直観的判断の創出」と「次世代教育等による直観的判断
そのものの(能動的)改訂」といふ二つの形で、上記構図における直観の層を拡張。
- (註 32)「もはやこれは『哲学』ではないのではないか？」 ⇒ 哲学です！