

科学哲学科学史特殊講義（京都大学）

3 古典的AI：哲学的論争

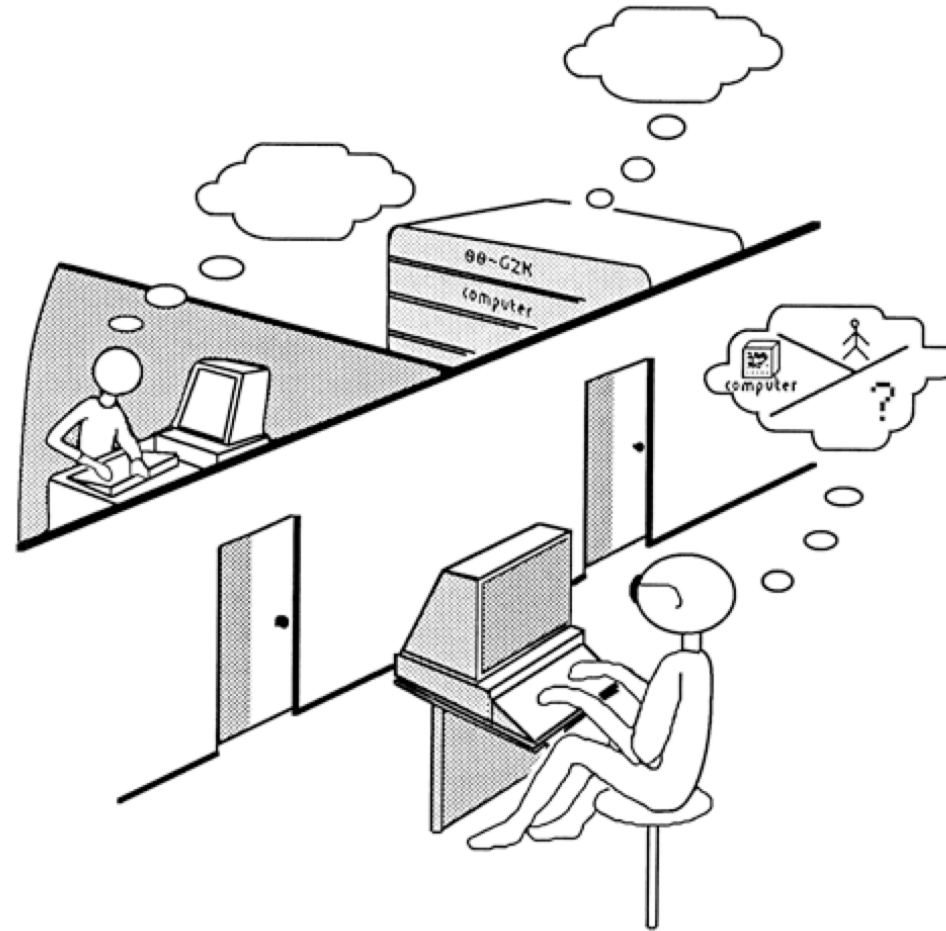
鈴木貴之

（東京大学大学院総合文化研究科）

tkykszk@g.ecc.u-tokyo.ac.jp

1 チューリング・テストと中国語の部屋

チューリング・テスト



チューリング・テストの妥当性：

- チューリング・テストに合格することは、知能の必要条件ではないが、十分条件？
- ただし、話題や時間に関して制限のないテストでなければならない。
- サブ認知的な質問を禁じることの困難さ（French 1990）。
- 人間との比較は必要か？

チューリング・テストとトータル・チューリング・テスト：

- トータル・チューリング・テスト（TTT）：コンピュータは、現実世界で人間ができることすべてを、それと識別不可能な仕方で行えるか？
- TTに合格するためには、TTTに合格できる能力が必要かもしれない（Harnad 1991）。
- そうだとしても、TTの知能のテストとしての妥当性は失われない（Copeland 1993）。

中国語の部屋



サール

「プログラムされたコンピュータは…実際には何も理解していないということを以下において論じたい。コンピュータの理解は（私のドイツ語の理解のように）部分的なもの、あるいは不完全なものですらない。それはゼロなのである。」（『マインズ・アイ』、p.186）

「記号形式の操作はそれだけでは志向性を有していない。その操作はまったく意味を欠いている。それは記号操作でさえない。なぜなら、その記号は何も表していないからである。言語学の専門用語で言えば、それには構文論しかなく、意味論がない。」（『マインズ・アイ』、p. 203）

サールの主張：

- 記号システム仮説が偽であることは、アプリアリに証明可能。
- 記号操作システムは構文論しか扱えない。意味理解には構文論だけでは不十分。

検討：

- 批判1 (Levesque 2009) : サールが想定するような一覽表的なマニュアルは、計算論的に実現不可能。現実世界において思考を可能にするアルゴリズムは、必然的に理解を反映したものとなる。
- Russell and Norvig 2021, Ch.2 : AI研究にとって核心的な挑戦は、巨大な表ではない仕方で合理的行動を生み出すことができるプログラムを見出すこと。

- このようなプログラムが存在するかどうかは、物理的記号システム仮説の是非に関する経験的問題。
- しかし、サールは一覧表的なマニュアルだけを想定しているわけではない。

- 批判2（システム反論）：部屋の中にいるジョンが中国語を理解していないということからは、部屋そのものが中国語を理解していないということは帰結しない。
- サールの応答：ジョンがマニュアルを暗記したとすると、ジョンは適切な応答が可能だが、依然として中国語を理解していないのでは？

- 再批判：このときジョンは本当に中国語を理解していないのか？
 - 類比：筆算の部屋や述語論理の部屋
 - ジョンが中国語を理解していないという直観は何に由来するのか？

「脳のシミュレーターに関して問題なのは、それが脳について誤ったことをシミュレートしていることだ。それがシナプスでの一連のニューロンの興奮の形式的な構造をシミュレートするだけにとどまるかぎり、それは脳に関する重要なこと、すなわちその産出的性質、その志向的状态をつくり出す能力を決してシミュレートしたことにはなるまい。」（『マインズ・アイ』、p. 196）

「私が英語を理解することができ、他のタイプの志向性をもちうるのは、私がコンピュータ・プログラムの具体例だからではない。…そのようなことが可能になるのは、われわれが知る限りでは、私がある種の生物学的（すなわち、化学的、物理的）構造をもったある種の有機体であるからなのであり、この有機体は、ある条件のもとでは知覚、動作、理解、学習、そしてその他の志向的現象を産出することができる。」（『マインズ・アイ』、p. 201）

- サールは、思考あるいは志向性（intentionality）は生物学的な現象だと主張する。しかし、脳の生物学的性質がどのようにして思考や志向性を可能にしているのかについては、一切説明を与えていない。

記号接地問題 (symbol grounding problem, Harnad 1990) :

- 真の意味理解が成立するためには、原初的記号は記号でないものと結びつけられていなければならない。
- この結びつきは、知覚や行為によって可能となる。

2 フレーム問題

知識表現の問題：

- 多くの知識の中から必要な知識をどのように取り出すか？
- 更新が必要な知識をどのように特定するか？

関連性の問題：

- Minskyらの対処法：知識同士の関連性に関するメタ的な知識を関連性公理として与える。
- 問題：
 - 関連性公理を適用しても多くの知識が残る。
 - 関連性公理自体が膨大になる。
 - どのような関連性公理にも例外がありうる。

知識更新の問題：

- 知識ベースの矛盾をどのように解消するか？
- 優先すべき知識を特定するための一般的基準は存在しない。

フレーム問題：

- デネットの例（Dennett 1984）：
 - 与えられた関連性公理をチェックするロボット
 - 可能な関連性をすべてチェックするロボット
 - 可能な関連性のチェック必要性をすべてチェックするロボット

デネット「コグニティヴ・ホイール」

「それゆえわれわれのうちに、あるいは知的ないかなる行為者のうちにも、必要な情報をすべて表示する（ないし貯える）非常に効率的なシステムが存在しなければならない。」（p. 134）

「変化する世界について（たんなる信念ではなく）優れた知識をもっていたとしても、それが有効に利用されるとはかぎらない。フレーム問題というのは、それを有効に利用するためにはどう表示すればよいのか、という問題なのである。」（p. 137）

「ではどのようなシステムが必要であろうか。それは、自分が知っていることの大半を純粹に無視し、いつでも自分の知識の一部をうまく選び出してそれによって作業するようなシステムである。…複雑な行為状況におけるじつに多様な状況のもとで無視すべきことを確実に無視するようなシステムを設計するにはどうすればいいのか。」(p. 140)

松原仁「一般化フレーム問題の提唱」

「この一般化フレーム問題は有限の情報処理能力しか持たない主体（人間やコンピュータはもちろんこれに含まれる）には決して完全解決はできないこと、それにもかかわらず日頃人間はあまりフレーム問題に惑わされていないように見えること、の二点から、人工知能研究においてフレーム問題について考えなくてはいけないのは、人間はあたかもフレーム問題を解決しているかのように見えるのが多いのはなぜであるかという問題（これを疑似解決の問題と名付ける）であること、を主張する。」（p. 179）

知識アクセス問題との類似性：

- 関連性のない要因を除外することが課題。
- 問題は、あることが重要かどうかをどのようにして決定するかではなく、いかにしてほとんどの場合にその決定を回避するか。

知識アクセス問題との違い：

- 知識アクセス問題はタイプに関するもの。フレーム問題は特定の状況に関するもの。
- システムの知識には、永続的な事実に関する知識と一時的な事実に関する知識がある。フレーム問題は後者に関するもの。

フレーム問題がより困難である理由：

- 一時的知識を更新する際には、知識ベースのエントリーがどのように相互依存しているかが問題となる。
- しかし、この相互依存関係自体が状況依存的。

考えられる（不十分な）対応：

- 安上がりなテスト戦略：どのようなタイプの出来事や事実がどのようなタイプの出来事や事実に影響するかを事前に分類しておく。
- 眠る犬戦略：積極的な理由がないかぎり、変化を加えない。

3 考察

第2次ブーム期までの論争からの教訓：

- AIに対するアプリアリな批判は説得的ではない。
 - 「いかなるものもXなしには知能をもちえない。いかなる古典的AIシステムもXをもちえない。それゆえ、いかなる古典的AIシステムも知能をもちえない。」
 - Xの候補：感じ、創造性、個性、学習、自由、直観、道徳性など。
 - Xを具体的に明示化すると説得力は失われる。

- 他方で、実際のAI研究を見ると、このやり方で人間のよう
な知能を実現することは困難であるように思われる。
- 根本的な問題：関連性、状況、意味、常識…

- なぜ人間は関連性を把握できるのか？
 - 哲学者の側にも明確な説明はない。

ドレイファス『コンピュータには何ができないか』

「知能は理解を要求し、理解はコンピュータに常識という背景を与えることを要求するが、その常識を成人した人間がもっているのは、彼が身体をもち、技能を通じて物質世界と相互作用し、ある文化へと教育されるからだ」（邦訳、p. 5）

「伝統哲学の先入見を去って記述的あるいは現象学的な証拠に頼れば、どんな形の知的振舞いにもプログラム不可能な人間の能力が含まれている、という示唆が得られる」（邦訳、p. 487）

「日常世界は、コンテクストを欠いた要素の集合によっては表現＝表象することができない」(Dreyfus and Dreyfus 1988, p.41)

- 問い：

- 人間はどのような方法で関連性の問題に対処しているのか？
- そこにおいて、身体、技能、文化などはどのような役割を果たしているのか？