ポール・ワイスの思考実験によるエージェントに対するメンタルイメージの把握

Understanding Humans' Mental Image of Agents by Means of "Paul A. Weiss' Thought Experiment"

小松 孝徳 (tkomat@meiji.ac.jp)

Takanori Komatsu

明治大学総合数理学部
School of Interdisciplinary Mathematical Sciences, Meiji University

Abstract: これまで著者は，「トロッコ問題」「テセウスの船」といった思考実験課題にエージェントを登場させることで，ユーザがそのエージェントに対してどのようなメンタルイメージベースの認識を抱いていたのかを把握する手法を提案してきた．そこで本稿では「ある対象を完全に破壊する前と後とで，いったい何が失われたと感じるのか」を問う「ポール・ワイスの思考実験」に着目し，ロボットや AI システムといった新しい技術がユーザにとってどのように認識されているのか，その把握を目指した二つの調査について報告する．

はじめに

ロボットや AI システムなどのエージェントとユーザとの間に円滑なコミュニケーションを構築するためには，ユーザがそのエージェントをどのように認識しているのかを理解することが非常に重要である．なぜなら，ユーザのエージェントへの認識は，その二者間のインタラクションの在り方自体を大きく左右するものだからである．なお本研究では，あるエージェントがユーザの前に実際に存在しているという状況での知覚ベースの認識ではなく，ユーザがそれらのエージェントに対してどのようなメンタルイメージを抱いているのかというメンタルイメージベースの認識に着目する．その理由は，あるエージェントがユーザの前に実際に現れた場合であっても，ユーザはそのエージェントに対してすでに抱いているメンタルイメージをもとにそのエージェントに対するインタラクションを開始すると考えられるからである．

これまで HRI (Human-Robot Interaction)や HAI (Human-Agent Interaction)といった研究分野にて，ユーザがエージェントをどのよう認識しているのかを把握する研究が数多く行われてきており [8, 10-12, 15, 33, 37, 38, 41, 42]，それらの研究は，「ユーザのエージェントに対する認識をユーザ自身に明示的に尋ねるもの」と「ユーザのエージェントに対する認識を暗黙的に抽出しようとするもの」，の二種類の研究アプローチに大別することができる．

前者のアプローチに相当する研究としては，アンケートなどの質問紙調査 [4, 6, 19, 27, 31]，半構造化インタビュー [1-3, 7]などを用いて，参加者のエージェントに対する認識を直接的かつ明示的に尋ねるという研究が挙げられる．しかしながらこれらの手法には，参加者がすでに意識している認識しか抽出することができない，さらには種々の回答バイアスの影響を受けやすいという欠点がある [23, 25, 28-30]．

一方，後者のアプローチに相当する研究としては，脳活動や皮膚電位といった生体信号 [21, 22, 39, 40] やユーザが不随意的に表出する非言語情報 [26, 36] を測定することでユーザのエージェントに対する認識を把握するという研究が挙げられる．しかしながらこれらのアプローチでは，大がかりな測定設備および統制された実験環境が必要不可欠となる．それ以外の代表的アプローチとしては，対象物と統制物との間の無意識的な関係性を測定する心理学的手法である潜在連合テスト(Implicit Association Test) [13, 14]が挙げられる．しかしながらこのテストは対象物と統制物との相対的な差異を比較することしかできないという限界がある．

このような中，後者のアプローチの一つとして，思考実験課題にエージェントを登場させるという研究が注目されている [16-18, 24]．例えば，Malle et al.

[24] は「トロッコ問題」にて「トロッコに乗った 5 人を見殺しにするか，引き込み線の先にいる 1 人を犠牲にするのか」という選択を迫られる対象としてロボットと人間をこの思考実験に登場させた．そして，これらの対象はこの状況でどのような行動をするべきかを尋ねる調査を実施した．その結果，選択を迫られているのが人間の場合はいずれの選択肢を選んでも非難されない一方，ロボットの場合は「5 人を見殺しにする」ことを選択した場合，非常に強く非難されるということが明らかとなった．このことから，私たち人間はロボットに対して人間に求めるのと異なる道徳的規範を無意識的に適用していると Malle et al らは結論づけた．このように思考実験課題にエージェントを登場させることで，エージェントに対する無意識的な認識を抽出するという手法は非常に有効な手法であると考えられる．なぜなら，トロッコ問題のような回答に逡巡するような思考実験課題に取り組んでいる際，そのユーザの認知資源のほとんどはこの課題における意思決定に費やされ，そこに登場するエージェントへの認識には費やされていないために，結果としてユーザのエージェントに対するある種「正直な」認識の抽出が可能となると考えられるからである（図 1）．

図 1：思考実験課題とユーザの認知資源の関係

そこで，Komatsu and Shirai [18] は「ポール・ワイスの思考実験」という思考実験課題に注目して，エージェントに対するメンタルイメージベースの認識を抽出する手法の検討を行った．「ポール・ワイスの思考実験」とは，オーストリア生まれの細胞生物学者 Paul A. Weiss によって提案された以下のような思考実験課題である [43]：「一つの試験管にはヒヨコ胚，もう一つの試験管にはヒヨコ胚を完全にすりつぶした液体が入っている．両者の中身は分子レベルでは完全に同一であるが，生物学的なシステムは同一といえるだろうか？」つまりこの思考実験課題は，「対象を完全に均一化（もしくは粉砕）することで何が失われ，何が失われなかったか」を検討するものである．Komatsu and Shirai [18] は，この思考実験課題にコンピュータ，ロボット，人間という三つの対象を登場させて調査参加者がこれらの対象をどのように認識しているのかを把握する調査を実施した．具体的には，以下のようなシナリオを調査参加者に提示した：「あるコンピュータ/ロボット/人間を巨大ミキサーにかけて完全に粉砕すると，そのコンピュータ/ロボット/人間由来の粉末/液体を得ることができた．このコンピュータ/ロボット/人間が粉砕されたことで，「失われたもの」は何でしょうか？思いついたものをすべて回答してください．」この問題を提示された参加者は，そのエージェントを構成する要素の中でも，「それをそれたらしめている最も重要な要素」に注目して回答すると予想されるため，参加者がエージェントをどのように認識しているのかを効果的に把握できると考えた．そして参加者の回答に対してテキストマイニングの一種であるコレスポンデンス分析を実施した結果，参加者はロボットを「コンピュータでも人間でもない存在」として認識していたことが明らかとなった．

そこで本稿では「ポール・ワイスの思考実験」によって，さらに多様なエージェントに対する認識の抽出を試みた二つの事例について報告する．まずは，Komatsu and Shirai [18] が注目した，コンピュータ，ロボット，人間という三つのエージェントに加えて AI システムというエージェントを追加した四種類のエージェントに対する調査を実施した（調査 1）．近年私たちの日常生活に急速に普及してきた AI システムを，私たちはどのように認識しているのだろうか．すでに我々にとってなじみ深いコンピュータ，まだ完全に普及しているとは言い難いもののすでに存在としては広く知られているロボットと，AI システムを比較することで，AI システムという非常に新しい技術がどのように人間に認識されているのかの把握を目指した．続いて，調査 1 で対象としたコンピュータ，ロボット，AI システムというエージェントは，非常に曖昧かつ幅広い概念であると考えられるため，参加者によってはこれらのエージェントを具体的にイメージすることが難しいとも想像できる．そこで，調査対象となるエージェントを，より具体的な対象に置き換えて調査を行った（調査 2）．具体的には，コンピュータは「スマートフォン」「ノート PC」，ロボットは「掃除ロボット」「ヒューマノイド」，AI システムは「生成 AI」「対話型 AI」とより日常的に使用されているエージェント名に置き換えたうえで，同様の調査を行った．

したがって本稿では，調査 1 および調査 2 それぞれに対して，以下二つの研究課題（RQ）を設定したといえる．

- RQ1: ポール・ワイスの思考実験に，コンピュータ，ロボット，人間，AI システムという四種類のエージェント登場させた場合，参加者はこれらのエージェント，特にロボットおよび AI システムをどう認識しているのか？
- RQ2: 調査 1 よりもより具体的なエージェントを登場させた場合（例．コンピュータではなくノート PC およびスマートフォン），参加者はこれらのエージェント，特にロボットおよび AI システムをどう認識しているのか？

これら二つの RQ は本稿を読み進めていく上での指針となるであろう．

調査 1

調査概要

人間，コンピュータ，ロボット， AI システムの四種類のエージェントをポール・ワイスの思考実験に登場させて，エージェントが完全に粉砕されると何が失われたのかを回答させる調査を実施した．具体的には，以下の実験シナリオのいずれかを調査参加者に提示した．なお，シナリオはテキスト情報のみで構成されており，エージェントの写真やイラストなどは付与されていない．

- **人間**：ある人を巨大ミキサーにかけて完全に粉砕すると，その人由来の液体を得ることができた．
- **コンピュータ**：あるコンピュータを巨大ミキサーにかけて完全に粉砕すると，そのコンピュータ由来の粉末を得ることができた．
- **ロボット**：あるロボットを巨大ミキサーにかけて完全に粉砕すると，そのロボット由来の粉末を得ることができた．
- **AI システム**：ある AI（人工知能）システムを巨大ミキサーにかけて完全に粉砕すると，その AI システム由来の粉末を得ることができた．

これらの実験シナリオに続いて，以下のような質問を参加者に提示し，具体的な回答を求めた（XXX にはエージェント名が入る）．

- この[XXX]が粉砕されたことで、「失われたもの」は何でしょうか？思いついたものをすべて回答してください．

四種類のエージェントに対する参加者の回答は，テキスト分析ソフトウェア「KH-Coder」[1]のコレスポンデンス分析（対応分析）によって解析された．本分析のパラメータは以下の通りであった：最小出現数は 2，最小文書数は 1，最大出現数および最大文書数は不定，除外対象は名詞 B，動詞 B，形容詞 B，副詞 B，否定助動詞，非依存形容詞，分析対象語数は上位 60 語．

コレスポンデンス分析とは，対象物（エージェント）ごとに回答で使用された単語の違いや共通点を可視化する手法である [20]．具体的には，クロス集計表を作成し，行要素としてすべての回答から抽出された単語を，列要素として四種類の外部変数（四種類のエージェント）を設定することで，単語と外部変数との関係を明確にすることが可能となり，対象物に対する参加者の認識構造を明らかにすることができるものである．

参加者

「Yahoo！Japan クラウドソーシング」[34]にて募集された 409 人が調査に参加したが，そのうち 7 名が無意味な回答のために除外され，残り 402 人（男性 288 人，女性 106 人，無回答 8 人；20～78 歳，M = 50.80 歳 (SD. = 10.34)）の回答を分析対象とした．報酬として一人 50 PayPay ポイントが付与された．

402 人の参加者のうち，100 人がエージェントが人間のシナリオ，99 人がコンピュータ，98 人がロボット，105 人が AI システムのシナリオに割り当てられた（参加者間計画）．

結果

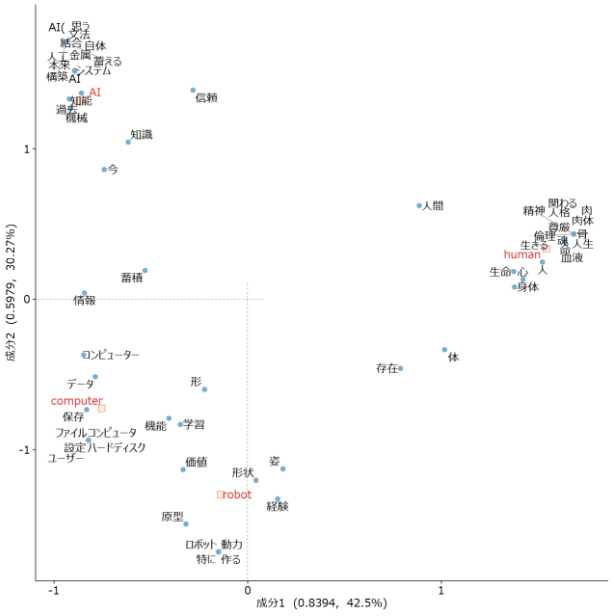

図 2：四種類のエージェントへの回答に対するコレスポンデンス分析結果

図 2 に参加者の回答に対するコレスポンデンス分析の結果を示す．この図より，四つの赤字で示された

[1] https//khcoder.net/

外部変数のうち，「人間」「AI システム」「コンピュータ/ロボット」が原点を中心とする三角形の頂点付近に配置されていることが分かった．このコレスポンデンス分析では，ある外部変数への回答で多く用いられる単語が他の外部変数への回答で用いられている単語と似ている場合には，これらの外部変数は近い位置に配置され，用いられている単語が異なる外部変数同士は，遠い位置に配置される．このことから，本調査で注目した四種類のエージェント（外部変数）は，「人間」「AI システム」「コンピュータ/ロボット」という三つのグループに分類されていることが分かった．

また，コレスポンデンス分析では，ある外部変数に特徴的な単語はその外部変数付近に配置され，複数の外部変数間で共通に使用される単語はそれらの間に配置される．この観点から各外部変数を観察してみると，外部変数「人間」の周囲には，「肉体」「血液」といった人間を構成する物理的要素に関する単語，「精神」「生命」といった人間の内面的要素，また「倫理」「尊厳」といった人間の権利に関する単語が配置されていることが分かった．外部変数「コンピュータ」の周囲には「保存」「設定」といったコンピュータの機能に関する単語や「ファイル」「ハードディスク」といったコンピュータを構成要素に関する単語が配置され，「ロボット」の周囲には「形状」「原型」といったロボットの外見に関する単語や「動力」「経験」といったロボットの機能に関する単語が配置されていた．さらに「AI システム」の周囲には，「知能」「知識」「文法」といった AI の持つ高度な能力に関する単語が配置されていることがわかった．

外部変数の「人間」と「ロボット」の間には，「存在」「体」という物理的身体に関する単語が，「コンピュータ」「AI」の間には「情報」「蓄積」といった情報処理に関する単語が配置されていることも明らかとなった．さらに，「ロボット」と「コンピュータ」の間に「学習」「価値」といった単語が配置されているが，非常に興味深い使われ方をしていたのは「学習」という単語であった．具体的に「学習」という単語は，『ロボットが固有に学習した記憶・データ・経験・価値観など』『コンピュータがインターネット上で学習して蓄えてきた事柄』といった使われ方をしていることが明らかとなった．このことから，コンピュータは「学習」，さらにロボットは「学習」と「経験」によって動的に知識やデータを獲得しているエージェントであると認識されていたことが示唆された．

コンピュータと似たような情報処理をしていると認識されている AI システムであるが，その中でも興味深い使われ方をしていたのは「過去」という単語であった．具体的に「過去」という単語は，『蓄積された過去の情報』『過去から蓄積されてきた様々な人間の叡智や知識』といった使われ方をしており，AI システムは過去から蓄積された静的なデータや知識を用いて知的な活動をしているエージェントであると認識されていたことが示唆された．能動的に情報を学習する「コンピュータ」「ロボット」，過去からの静的な情報を利用する「AI システム」という認識の違いが，図 2 におけるこれらのエージェント（外部変数）の配置に反映されていると考えられた．

これら四種類のエージェントに対する回答の解析から，「AI システム」は人間でもコンピュータでもロボットでもなく，過去から蓄積された静的な知識を操る知的な存在，さらには「ロボット」は人間や AI システムでもなく，コンピュータと同じく学習によって，さらには自身の経験によって情報を動的に獲得していく物理的な身体をもつ存在として認識されていたことが示唆された．

調査 2

調査概要

調査 1 で着目したエージェントのうち，コンピュータ，ロボット，AI システムをより具体的な対象名に置き換えて同様の調査を実施した．具体的には，コンピュータは，「スマートフォン」「ノート PC」，ロボットは，「掃除ロボット」「ヒューマノイド」，AI システムは，「生成 AI」「対話型 AI」とより日常的に使用されているエージェント名に置き換え，人間を含めた合計七種類のエージェントによって調査を実施した．

ノート PC，スマートフォン，掃除ロボット，ヒューマノイドが登場する実験シナリオは，調査 1 におけるコンピュータもしくはロボットのシナリオのうちエージェント名を単純に置き換えたものを使用した．生成 AI および対話型 AI の実験シナリオは，エージェント名を置き換えただけではなく，実験シナリオの最後にそれぞれ「※生成 AI：画像・文章・音声などを新たに生成する AI（例：ChatGPT、Google Gemini など）」「※対話型 AI：人間との自然な会話ができる AI（例：iPhone の Siri、Google アシスタントなど）」という具体的な AI システムの動作および製品名についての説明を付与した．実験シナリオの提示後，参加者に提示される質問は調査 1 と同様であった．

参加者

「Yahoo！Japan クラウドソーシング」にて募集され

た 620 人が調査に参加したが，そのうち 14 名が無意味な回答のために除外され，残り 606 人が人間以外の六種類のエージェントとの実験シナリオに割り当てられた．具体的には 606 人のうち，104 人がスマートフォン，102 人がノート PC，102 人が掃除ロボット，96 人がヒューマノイド，105 人が生成 AI，97 人が対話型 AI のシナリオに割り当てられた（参加者間計画）．

なお上記の六種類のエージェントに対する回答に，調査 1 の人間シナリオの回答を加えた合計 706 人（男性 474 人，女性 223 人，無回答 9 人；20〜87 歳，M = 51.03 歳 (SD. = 11.91)）の回答を分析対象とした．報酬としては一人 50 PayPay ポイントが付与された．

結果

図 3 に参加者の回答に対するコレスポンデンス分析の結果を示す．この図より，外部変数のうち「人間」「掃除ロボット」「ノート PC/スマートフォン」が原点を中心とする三角形の頂点付近に配置されていることが分かった．さらに，「生成 AI」と「対話型 AI」が「ノート PC/スマートフォン」と「人間」とを結ぶ直線上の「ノート PC/スマートフォン」にかなり近い位置に配置され，「ヒューマノイド」がどのエージェントとも距離を置いた三角形の中心付近に配置されていることが分かった．このことから，本調査で注目した七種類のエージェントは，「人間」「掃除ロボット」「ノート PC，スマートフォン，生成 AI，対話型 AI」「ヒューマノイド」という四つのグループに分類されることが分かった．

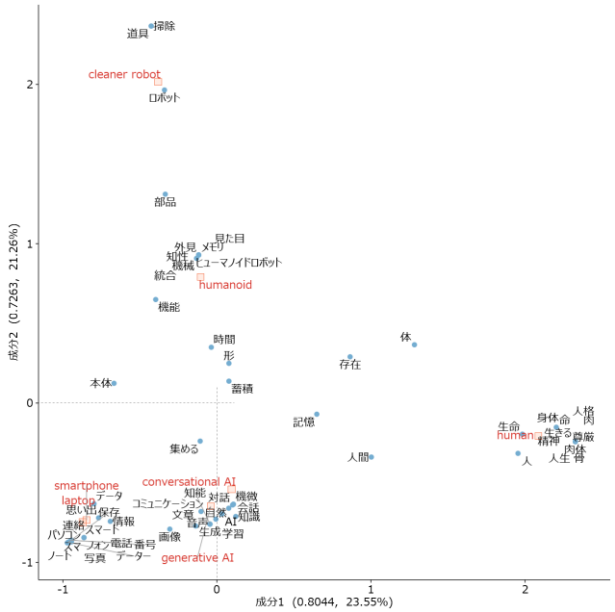

図 3：四種類のエージェントへの回答に対するコレスポンデンス分析結果

この図 3 を細かく見ていくと，「ノート PC」「スマートフォン」の周囲には「データ」「写真」「思い出」などこれらの端末が保持しているデータや「連絡」「保存」といった機能に関する単語が配置されていた．特にこれらの単語は『思い出の写真』『大事なデータ』『昔からの連絡先』のように思い入れを表現する形容詞とともに使用されていたことから，これらのエージェントは私たちの日常生活において非常に重要なツールと認識されていることが理解できた．

その少し右に同じく隣接している「生成 AI」「対話型 AI」の周囲には，「画像」「文章」「音声」「対話」など実際に AI によって生成される情報や「知識」「学習」「知能」といった AI の知性に対する単語が配置されていた．これらの単語は，『自然な対話』『人工的に作られた文章や画像』『元の画像や音声』『学習された文章や音声』など，生成 AI や対話型 AI を実際に使用しているユーザならではの回答が多く観察された．このことからも，生成 AI や対話型 AI も私たちの日常生活においてすでに有効に活用されていることが明らかとなった．また「生成 AI」「対話型 AI」というエージェントは，「ノート PC」「スマートフォン」上で使用されることが多く，さらにはこれら四種類のエージェントは多くのユーザにとってすでに日常生活において無くてはならない存在となっているため，これらのエージェントは隣接した位置に配置されて一つのグループを構成していると考えられた．

一方，図 3 上部に配置されている外部変数「掃除ロボット」の周囲には「道具」「掃除」といった掃除ロボットの具体的な機能に関する単語が配置されているが（具体的な回答としては『掃除機能』『掃除としての道具』など），そこから下方向に離れた位置に配置されている「ヒューマノイド」の周囲には，調査 1 における「ロボット」と同様に「外見」「見た目」といったロボットの外見に関する単語や，「統合」「機能」といったロボットの機能に関する単語が配置されていることが明らかとなった．具体的な回答としては『ヒューマノイドの存在自体』『構成する部品』とったものの他に，調査 1 と同様に，『学習して獲得したデータ・機能など』『記憶した全データ』などという回答が多く観察された．ここから，「掃除ロボットは掃除をする道具」として調査参加者に具体的に認識されている一方，「ヒューマノイド」に対してはそのような具体的な認識はされておらず，調査 1 と同様の抽象的な存在として認識されていることが示唆された．この二者は同じロボットでもあるにもかかわらずその位置は大きく離れており，さらに「ヒューマノイド」は他の三つのグループ（「人間」「掃除ロボット」「ノート PC，スマートフォン，生成 AI，

対話型 AI」）とも離れた位置に配置されていることから，人間でもなく，「掃除ロボット」のように掃除をする道具でもなく，「ノート PC，スマートフォン，生成 AI，対話型 AI」のように日常生活において不可欠な機器でもないという，非常に曖昧な存在として認識されていると考えられた．

議論

本稿では，思考実験課題「ポール・ワイスの思考実験」にエージェントを登場させることで，ユーザがこれらのエージェントをどのように認識しているのか，特にロボットや AI システムといった新しい技術がどのように認識されているのかを把握するために実施した二つの調査について報告した．

一つ目の調査 1 では，Komatsu and Shirai [18] が着目したコンピュータ，ロボット，人間という三つのエージェントに加えて AI システムというエージェントを追加して合計四種類のエージェントに対する認識を把握する調査を実施した．調査 2 では，調査 1 で注目したエージェントのコンピュータを「スマートフォン」「ノート PC」に，ロボットを「掃除ロボット」「ヒューマノイド」に，AI システムを「生成 AI」「対話型 AI」と，より日常的に使用されているエージェント名に置き換えて人間というエージェントを加えた合計七種類のエージェントに対する認識を把握する調査を実施した．

本稿の冒頭で，それぞれ調査 1 および調査 2 に対して，以下のような RQ を設定した．

- RQ1: ポール・ワイスの思考実験に，コンピュータ，ロボット，人間，AI システムという四種類のエージェント登場させた場合，参加者はこれらのエージェント，特にロボットおよび AI システムをどう認識しているのか？
- RQ2: 調査 1 よりもより具体的なエージェントを登場させた場合（例．コンピュータではなくノート PC およびスマートフォン），参加者はこれらのエージェント，特にロボットおよび AI システムをどう認識しているのか？

そして，調査 1 および調査 2 の結果，上記の RQ に対して，以下のように回答することができると考えられる．

- **RQ1 への回答**：「AI システム」は人間でもコンピュータでもロボットでもなく，過去から蓄積された静的な知識を操る知的な存在，さらには「ロボット」は人間や AI システムでもなく，コンピュータと同じく学習によって，さらには自身の経験によって情報を動的に獲得していく物理的な身体をもつ存在として認識されていたことが示唆された．
- **RQ2 への回答**：「生成 AI」「対話型 AI」「ノート PC」「スマートフォン」は，多くのユーザにとってすでに必要不可欠な重要なツールとして，「掃除ロボット」は掃除の道具として，「ヒューマノイド」は人間でもなく，「掃除ロボット」のように掃除をする道具でもなく，「ノート PC，スマートフォン，生成 AI，対話型 AI」のように日常生活において不可欠なツールでもないという，非常に曖昧な存在として認識されていると考えられた．

調査 1 と調査 2 の違いは，調査対象とするエージェントの名前が抽象的か具体的かという点のみである．そのような操作だけでも，AI システムおよびロボットは調査間にてその認識が大きく異なっていたことが明らかとなった．

まず AI システムの場合，エージェント名に「AI（人工知能）システム」を使用した調査 1 では，調査参加者の回答に「学習」という単語が使用されず，「過去」から蓄積された静的な知識を操るエージェントと認識されていた．そのため，回答に「学習」という単語が使われていた「コンピュータ」とは異なる位置に外部変数が配置されていた．しかしながら，「生成 AI」「対話型 AI」を使用した調査 2 では，回答に「学習」という単語が使用され，それにより「コンピュータ」（調査 2 の場合は「ノート PC」「スマートフォン」）と同じような位置に外部変数が配置されていたことが明らかなった．この理由としては，調査 1 で使われた「AI（人工知能）システム」というエージェントよりも，調査 2 の「生成 AI」「対話型 AI」というエージェント対して調査参加者が適切な知識を有していたことが挙げられる．特に調査 2 では，「生成 AI」「対話型 AI」と併せて ChatGPT や Siri など日常で多く使われている具体的なアプリケーション名を提示したため，これらのエージェントが実際に学習しながら動作をしていることを参加者がすでに経験することで，それらに対して適切に理解した上で回答をしていたと考えられた．その一方，調査 1 の「AI（人工知能）システム」というエージェント名は非常に抽象的で大きい意味を含む名前であるため，具体的にどのようなシステムでどのような動作をするのかイメージすることが難しく，ソーシャルメディアの記事や投稿などでよく見かけるような「人工知能が・・・」という伝聞で得た知識やステレオタイプによって回答がなされていたと考えられた．つまり具体的にそのエージェントを経験したことなどによって得られた適切な知識の有無がこのような異なる認識を生じさせていたのではと考えられた．

この「エージェントに対する適切な知識の有無」がエージェントの認識に与える影響は，調査 2 における「掃除ロボット」「ヒューマノイド」への認識にも当てはまる.「掃除ロボット」は『掃除をする道具』という回答に代表されるように適切かつ具体的に調査参加者に認識されていた一方,「ヒューマノイド」はその見た目やロボットの一般的な機能について抽象的に回答されており，これは「ヒューマノイド」が具体的にどのようなものでどのような動作をするのかが適切に理解されていないと示唆された．そのため,「ヒューマノイド」は人間でもなく,「掃除ロボット」のように掃除をする道具でもなく,「ノート PC, スマートフォン，生成 AI，対話型 AI」のように日常生活において不可欠なツールでもないという，非常に曖昧な存在として認識されていると考えられた.

調査 2 の結果から,「生成 AI」「対話型 AI」は「ノート PC」「スマートフォン」と同じようにユーザの日常生活において非常に重要な地位を占めていること，そして「掃除ロボット」は調査参加者に掃除の道具として正しくその位置づけが浸透している一方,「ヒューマノイド」はその認識が曖昧でいまだステレオタイプ的な認識にとどまっていることが明らかとなった．このステレオタイプ的な曖昧な認識をされているエージェントは，いわば日常生活にいまだ普及していないが故の認識だととらえると，この調査2からはロボットおよびAIシステムといった新しい技術は, 実はロボットよりも AI システムの方がすでにユーザの日常生活に浸透しているという現状を表したものだとも考えられる．また，調査 1 および 2 を通じて,「コンピュータ」および「スマートフォン」「ノート PC」は同じような位置に配置されており，このことはこれらのエージェントが，AI システムやロボットなどよりも，すでにユーザの日常生活に深く根付いていることを明確に示しているといえよう．その一方で，AI システムおよびロボットなどのエージェントは調査によってその位置を大きく変えているため，これらのエージェントの普及はまだその途上にあることを示しているとも考えられる．このようにポール・ワイスの思考実験はエージェントに対するユーザの現状認識を把握するのに非常に有効な方法であると言えよう.

おわりに

ポール・ワイスの思考実験にエージェントを登場させて，調査参加者がこれらの対象をどのように認識しているのか, 特にロボットおよび AI システムといった新しい技術がどのように認識されているのか把握する二つの調査を行った.

やや抽象的な「人間」「コンピュータ」「ロボット」「AI システム」というエージェントを登場させた調査 1 では,「AI システム」は人間でもコンピュータでもロボットでもなく，過去から蓄積された静的な知識を操る知的な存在，さらには「ロボット」は人間や AI システムでもなく, コンピュータと同じく学習によって，さらには自身の経験によって情報を動的に獲得していく物理的な身体をもつ存在として認識されていたことが示唆された.

また調査 1 よりも具体的な名前を使用した調査 2 では,「生成 AI」「対話型 AI」「ノート PC」「スマートフォン」は，多くのユーザにとってすでに必要不可欠な重要なツールとして,「掃除ロボット」は掃除の道具として,「ヒューマノイド」は人間でもなく,「掃除ロボット」のように掃除をする道具でもなく,「ノート PC, スマートフォン，生成 AI，対話型 AI」のように日常生活において不可欠なツールでもないという，非常に曖昧な存在として認識されていると考えられた.

これら二つの調査から,「エージェントに対する適切な知識の有無」がそれらのエージェントに対する認識に大きな影響を及ぼしていることが明らかとなった．例えば，参加者にとってなじみ深い ChatGPT や Siri といった「生成 AI」や「対話型 AI」は,「AI（人工知能）システム」よりも正しく理解されているが故に日常生活に不可欠な存在として認識されていた．逆に言うと，ステレオタイプ的な理解など適切に理解されていないエージェントに対しては，参加者は曖昧な認識をするにとどまっており (例えば, 調査 2 における「ヒューマノイド」)，このことはこのエージェントが日常生活空間に普及していないという証左と成り得ると考えられた．このようにポール・ワイスの思考実験はエージェントに対するユーザの現状認識を把握するのに非常に有効な方法であると言えよう.

しかしながらこの研究手法にも，まだ慎重に検討しなければならない事項があるのは言うまでもない. 具体的には，調査参加者にシナリオを提示しその回答を「言語化」することを求めているため，本提案手法は「暗黙的に」エージェントに対するメンタルモデルを把握しているとは言えないのではという事項である．この問いに答えるためには，潜在連合テスト，質問紙調査や半構造化インタビューなどの明示的な把握方法を, 本手法と併せて実施することで, 各調査においてどのような情報が抽出可能なのかを把握することが, まずは必要不可欠だと考えている.

また，ポール・ワイスの思考実験以外の思考実験課題の応用可能性についても検討していきたい．す

でに,「トロッコ問題」[17, 24]や「テセウスの船」[16]などを利用してエージェントの認識を把握できることについては報告してきているが，このような思考実験課題によるエージェントのメンタルイメージベースの認識の把握という枠組みの一般化およびそのメリットを強化するべく，その他の思考実験課題の応用可能性についても積極的に検討していきたいと考えている.

謝辞

本研究は JSPS 科研費 JP25K15310 の助成を受けた. ここに謝意を記す.

参考文献

[1] Naoko Abe, Yue Hu, Mehdi Benallegue, Natsuki Yamanobe, Gentiane Venture, and Eiichi Yoshida. 2024. Human Understanding and Perception of Unanticipated Robot Action in the Context of Physical Interaction. J. Hum.-Robot Interact. 13, 1, Article 9 (March 2024), 26 pages. doi:10.1145/3643458

[2] S. Alhouli, N. Almania, and D. Sahoo. 2024. Designing a Positive Initial Experience with a Companion Pet Robot for Older Adults in Kuwait. In Human Interaction and Emerging Technologies (IHIET-AI 2024): Artificial Intelligence and Future Applications (AHFE Open Access, Vol. 120), Tareq Ahram and Redha Taiar (Eds.). AHFE International, USA. doi:10.54941/ahfe1004558

[3] Minja Axelsson, Nikhil Churamani, Atahan Çaldır, and Hatice Gunes. 2025. Participant Perceptions of a Robotic Coach Conducting Positive Psychology Exercises: A Qualitative Analysis. J. Hum.-Robot Interact. 14, 2, Article 36 (March 2025), 27 pages. doi:10.1145/3711937

[4] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. International Journal of Social Robotics 1 (2009), 71–81. doi:10.1007/s12369-008-0001-3

[5] Anton Bogdanovych, Tomas Trescak, and Simeon Simoff. 2016. What Makes Virtual Agents Believable? Connection Science 28, 1 (2016), 83–108. doi:10.1080/09540091.2015.1130021

[6] Courtney M. Carpinella, Aaron B. Wyman, Michael A. Perez, and Steven J. Stroessner. 2017. The Robotic Social Attributes Scale (RoSAS): Development and Validation. In Proceedings of the 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI ’17). Vienna, Austria, 254–262. doi:10.1145/2909824.3020208

[7] Nathan Caruana, Ross Moffat, Alba Miguel-Blanco, et al. 2023. Perceptions of Intelligence & Sentience Shape Children’s Interactions with Robot Reading Companions. Scientific Reports 13 (2023), 7341. doi:10.1038/s41598-023-32104-7

[8] Herbert H. Clark and Kerstin Fischer. 2023. Social Robots as Depictions of Social Agents. Behavioral and Brain Sciences 46 (2023), e21. doi:10.1017/S0140525X22001486

[9] Francesca Diana, Masahiro Kawahara, Ilaria Saccardi, et al. 2023. A Cross-Cultural Comparison on Implicit and Explicit Attitudes Towards Artificial Agents. International Journal of Social Robotics 15 (2023), 1439–1455. doi:10.1007/s12369-022-00917-7

[10] Nicholas Epley, Adam Waytz, and John T. Cacioppo. 2007. On Seeing Human: A Three-Factor Theory of Anthropomorphism. Psychological Review 114, 4 (2007), 864–886. doi:10.1037/0033-295X.114.4.864

[11] Julia Fink. 2012. Anthropomorphism and Human Likeness in the Design of Robots and Human-Robot Interaction. In Proceedings of the International Conference on Social Robotics (ICSR 2012). 199–208. doi:10.1007/978-3-642-34103-8_20

[12] Heather M. Gray, Kurt Gray, and Daniel M. Wegner. 2007. Dimensions of Mind Perception. Science 315, 5812 (2007), 619. doi:10.1126/science.1134475

[13] Anthony G. Greenwald, Debbie E. McGhee, and Jordan L. K. Schwartz. 1998. Measuring Individual Differences in Implicit Cognition: The Implicit Association Test. Journal of Personality and Social Psychology 74, 6 (1998), 1464–1480. doi:10.1037/0022-3514.74.6.1464

[14] Anthony G. Greenwald, Brian A. Nosek, and Mahzarin R. Banaji. 2003. Understanding and Using the Implicit Association Test: I. An Improved Scoring Algorithm. Journal of Personality and Social Psychology 85, 2 (2003), 197–216. doi:10.1037/0022-3514.85.2.197

[15] Cynthia Kidd and Cynthia Breazeal. 2004. Effect of a Robot on User Perceptions. In Proceedings of the 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2004), Vol. 4. 3559–3564. doi:10.1109/IROS.2004.1389947

[16] Takanori Komatsu. 2024. Understanding Humans’ True Perception of Robots by Means of a Thought Experiment “Ship of Theseus”. In Proceedings of the 16th International Conference on Social Robotics (ICSR 2024). 395–408.

[17] Takanori Komatsu, Bertram F. Malle, and Matthias Scheutz. 2021. Blaming the Reluctant Robot: Parallel Blame Judgments for Robots in Moral Dilemmas Across U.S. and Japan. In Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (HRI ’21). Association for Computing Machinery, New York, NY, USA, 63–72. doi:10.1145/3434073.3444670

[18] Takanori Komatsu and Karen Shirai. 2025. Understanding Humans’ Perception of Robots by Means of "Paul A. Weiss’

Thought Experiment". In Proceedings of the 2025 ACM/IEEE International Conference on Human-Robot Interaction (Melbourne, Australia) (HRI '25). IEEE Press, 1418–1422.

[19] Christian U. Krägeloh, Jeyasankar Bharatharaj, Santhosh Kumar S. Kutty, P. R. Nirmala, and Lijuan Huang. 2019. Questionnaires to Measure Acceptability of Social Robots: A Critical Review. Robotics 8 (2019), 88. doi:10.3390/robotics8040088

[20] Ashish Kumar and Avinash Paul. 2016. Mastering Text Mining with R. Packt Publishing, Birmingham, UK.

[21] Mengyao Li, Feng Guo, Xin Wang, Jie Chen, and Jaap Ham. 2023. Effects of Robot Gaze and Voice Human-Likeness on Users' Subjective Perception, Visual Attention, and Cerebral Activity in Voice Conversations. Computers in Human Behavior 141 (2023), 107645. doi:10.1016/j.chb.2022.107645

[22] Yifan Liu, Fan Li, Lin H. Tang, Zhiqiang Lan, Jing Cui, Olga Sourina, and Chi-Hung Chen. 2019. Detection of Humanoid Robot Design Preferences Using EEG and Eye Tracker. In 2019 International Conference on Cyberworlds (CW). 219–224. doi:10.1109/CW.2019.00042

[23] Karl F. MacDorman, Sriraam Vasudevan, and Chin-Chang Ho. 2009. Does Japan Really Have Robot Mania? Comparing Attitudes by Implicit and Explicit Measures. AI & Society 23 (2009), 485–510. doi:10.1007/s00146-008-0181-2

[24] Bertram F. Malle, Matthias Scheutz, Thomas Arnold, John Voiklis, and Cynthia Cusimano. 2015. Sacrifice One for the Good of Many? People Apply Different Moral Norms to Human and Robot Agents. In Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI '15). Association for Computing Machinery, New York, NY, USA, 117–124. doi:10.1145/2696454.2696468

[25] J. D. McDonald. 2008. Measuring Personality Constructs: The Advantages and Disadvantages of Self-Reports, Informant Reports and Behavioural Assessments. Enquire 1, 1 (2008), 75–94.

[26] Jonathan Mumm and Bilge Mutlu. 2011. Human-Robot Proxemics: Physical and Psychological Distancing in Human-Robot Interaction. In Proceedings of the 6th International Conference on Human-Robot Interaction (HRI '11). Association for Computing Machinery, New York, NY, USA, 331–338. doi:10.1145/1957656.1957786

[27] Tatsuya Nomura, Kazuki Sugimoto, David S. Syrdal, and Kerstin Dautenhahn. 2012. Social Acceptance of Humanoid Robots in Japan: A Survey for Development of the Frankenstein Syndrome Questionnaire. In 2012 12th IEEE-RAS International Conference on Humanoid Robots (Humanoids 2012). Osaka, Japan, 242–247. doi:10.1109/HUMANOIDS.2012.6651528

[28] J. C. Nunnally. 1978. Psychometric Theory (2nd ed.). McGraw-Hill, New York, NY.

[29] Mildred L. Patten. 2014. Questionnaire Research: A Practical Guide (4th ed.). Routledge, Abingdon, UK.

[30] Delroy L. Paulhus and Simine Vazire. 2007. The Self-Report Method. In Handbook of Research Methods in Personality Psychology, Richard W. Robins, R. Chris Fraley, and Robert F. Krueger (Eds.). The Guilford Press, London, 224–239.

[31] Adriana Peca, Mark Coeckelbergh, Ramona Simut, Cristina Costescu, Simona Pintea, Daniel David, and Bram Vanderborght. 2016. Robot Enhanced Therapy for Children with Autism Disorders: Measuring Ethical Acceptability. IEEE Technology and Society Magazine 35 (2016), 54–66. doi:10.1109/MTS.2016.2554438

[32] Tuğçe Nur Pekçetin, Seyda Evsen, Serkan Pekçetin, Cengiz Acarturk, and Burcu A. Urgen. 2024. Real-World Implicit Association Task for Studying Mind Perception: Insights for Social Robotics. In Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction (Boulder, CO, USA) (HRI '24). Association for Computing Machinery, New York, NY, USA, 837–841. doi:10.1145/3610978.3640706

[33] Aaron Powers and Sara Kiesler. 2006. The Advisor Robot: Tracing People's Mental Model from a Robot's Physical Attributes. In Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction (HRI '06). Association for Computing Machinery, New York, NY, USA, 218–225. doi:10.1145/1121241.1121280

[34] Katie Seaborn and Satoshi Nakamura. 2025. Quality and Representativeness of Research Online with Yahoo! Crowdsourcing. Frontiers in Psychology 16 (2025), 1588579. doi:10.3389/fpsyg.2025.1588579

[35] Nicolas Spatola and Olga A. Wudarczyk. 2021. Implicit Attitudes Towards Robots Predict Explicit Attitudes, Semantic Distance Between Robots and Humans, Anthropomorphism, and Prosocial Behavior: From Attitudes to Human–Robot Interaction. International Journal of Social Robotics 13 (2021), 1149–1159. doi:10.1007/s12369-020-00701-5

[36] Leila Takayama and Caroline Pantofaru. 2009. Influences on Proxemic Behaviors in Human-Robot Interaction. In 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2009). 5495–5502. doi:10.1109/IROS.2009.5354145

[37] J. Gregory Trafton, Caroline R. Frazier, Kristin Zish, Brian J. Bio, and J. Michael McCurry. 2023. The Perception of Agency: Scale Reduction and Construct Validity. In 2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN). 936–942. doi:10.1109/RO-MAN57019.2023.10309318

[38] J. Gregory Trafton, J. Michael McCurry, Kristin Zish, and Caroline R. Frazier. 2024. The Perception of Agency. ACM Transactions on Human-Robot Interaction (2024). doi:10.1145/3642133

[39] Büşra A. Urgen, Michael Plank, Hiroshi Ishiguro, Howard Poizner, and Ayşe P. Saygin. 2013. EEG Theta and Mu Oscillations During Perception of Human and Robot Actions. Frontiers in Neurorobotics 7 (2013), 10. doi:10.3389/fnbot.2013.00010

[40] Yuting Wang, Fang Liu, and Xue Lei. 2024. Neural Correlates of Robot Personality Perception: An fNIRS Study. In Cross-Cultural Design: 16th International Conference, CCD 2024, Held as Part of HCII 2024. Springer, Berlin, Heidelberg, 332–344. doi:10.1007/978-3-031-56742-1_25

[41] Adam Waytz, Kurt Gray, Nicholas Epley, and Daniel M. Wegner. 2010. Causes and Consequences of Mind Perception. Trends in Cognitive Sciences 14, 8 (2010), 383–388. doi:10.1016/j.tics.2010.05.006

[42] Adam Waytz, Carey K. Morewedge, Nicholas Epley, Greg Monteleone, J. H. Gao, and John T. Cacioppo. 2010. Making Sense by Making Sentient: Effectance Motivation Increases Anthropomorphism. Journal of Personality and Social Psychology 99, 3 (2010), 410–435. doi:10.1037/a0020240

[43] Paul A. Weiss. 1968. From Cell to Molecule. In Dynamics of Development: Experiments and Inferences. Academic Press, 24–95.

[44] Ralf Wullenkord and Friederike Eyssel. 2019. Imagine How to Behave: The Influence of Imagined Contact on Human–Robot Interaction. Philosophical Transactions of the Royal Society B: Biological Sciences 374, 1771 (April 2019). doi:10.1098/rstb.2018.0033