

# 人の心が生み出すエージェント

Agent that Human Mind Originates

竹内勇剛

Yugo TAKEUCHI

静岡大学 創造科学技術大学院 / 情報学部

Graduate School of Science and Technology / Faculty of Informatics, Shizuoka University

HAI researchers mostly endow an agent with behavioral and/or mental characteristics, distinct embodiment that functions as symbolic subject to interact, and role to perform. Nevertheless, people are able to conceive an imaginative agent even though the agent was not given such endowments. For example, people can easily believe in the existence of god as an omnipotent agent that is supposed to subjectively interact with human. In this paper, we discuss an issue concerning how do people cognitively identify the agent through human activities, and a suggestion of the next HAI studies is proposed.

## 1. はじめに

エージェントという言葉がコンピュータサイエンス分野、特に人工知能領域において、与えられた環境の中である目的を実現するために自律的な振舞いをする主体を表す象徴的なタームとして使用されるようになって久しい。とりわけ、一般に身体化エージェントやインタフェースエージェントと呼ばれるものや、近年ではコミュニケーションロボットと言われる人間との共生を目指したものは、人間とのインタラクションをすることを前提としたエージェントとして位置付けられている。HAI(Human-Agent Interaction) 研究はそのようなエージェントと人間とのインタラクションを成り立たせるための諸問題を中心的課題として取り組まれており、近年ホットな研究トピックとして注目されつつある [山田 07]。

HAI 研究では、一般にエージェントは実環境あるいは仮想環境において視覚的に知覚でき、空間的な位置を占める存在として取り扱われることが多い。そのためエージェントは工学的に構成される実在的な存在として捉えられることが多い。このようなエージェントの捉え方は、HAI 研究を確立化しその意義を学術的に認知させる上では大いに寄与してきた。しかし一方で、HAI 研究は構成論的なアプローチに基づいてクリーチャーを工学的にデザインすることと、観察されるインタラクションの解析をすることが主な研究目的であるというような、HAI 研究の意義を矮小化した認識を生み出す近因になっている可能性も否定できない。

HAI 研究はインタラクションに関する研究である以上インタラクションの当事者である人間とエージェントの相互的な理解が何よりも重要であり、その理解はさまざまな観点に基づいて議論することができる。したがって HAI の根本に立ち返り、人間にとってエージェントとは何かという存在論的問題を改めて問い直すことによって、HAI 研究を通して人間という個体の知性に関する問題に向き合うこともできるようになる。

本稿では、人間の心のはたらきとエージェント性の認知に関する議論を通して、これまであまり検討が進められていなかった HAI 研究ならではのインタラクションの観察の仕方と今後積極的に取り組むべき問題を提起し、我々が寄与することができる新たな HAI 研究の課題へのアプローチについて検討する。

## 2. インタラクションの観察と行為

人間は自我（内部状態）をもち、周囲の環境（外界）に対して能動的に行為を働きかけることができ、同時に外界から自己の内部状態を変化させていくといった外界とのインタラクションを行なう行為主体である。このことは人間自身も実世界で行動するエージェントの 1 つであることを含意している。人間の内部状態は一般に「心」と呼ばれ、人間は心のはたらき（認知活動）を通して外界と繋がっていると考えられている。

これまでの HAI 研究では、エージェントの身体性に基づいた引き込み現象や、Media Equation [Reeves & Nass 96] と呼ばれるコンピュータや高度な情報処理機器に対する対人的反応にみられるように、心のはたらきを顕在化させないインタラクションに関する報告が比較的多かった。これらのインタラクションは、基本的に人間の無自覚的な行為や反応に基づいているものであり、相手に対する意図推定など高次な心のはたらきを伴うインタラクションをしている際の人間の振舞いを説明するものではない。つまり明確な志向システムとしての人間やエージェントを取り扱う際には、これらとは異なるアプローチでの議論が必要になる。

通常、人間を含むエージェントの行為は成果 (result) と帰結 (consequence) に区別される [Wright 63]。成果とは為す (do) ことによって生まれ、その結果としてもたらされるものが帰結となる。つまりあるエージェントの行為は成果を因果的に生み出す、その成果によってもたらされる状態は帰結として結果とは区別される問題となる。この問題に対して、Wright は行為を次のように目的論的に説明している [Wright 63]。

- (1) エージェント I は帰結 b をもたらそうと意図する。
- (2) I は、成果 a によって b をもたらすことができると信じている。
- (3) それゆえ、I は a を達成しようとする。

これまで一般に、インタラクションを分析する際には行為者が為した結果の状態、すなわち上述の「成果」にのみ注目した観察が行なわれてきた。しかしエージェントの行為として注目されるべきは「帰結」であろう。なぜなら、人間であれエージェントであれその行為は帰結を意図した志向的なものとして捉えてしまう性質を人間は天然に有している [Dennett 96]。

この志向的な姿勢によって人間やエージェントによる行為は帰結を達成するための手段として理解される。一方、エージェントではない機械の行為はそれによって生じる因果的な成果を生み出すことが目的である。したがって意図的行為によって構成される HAI における人間やエージェントの行為に関する観察は、インタラクションによってもたらされた「帰結」を基盤にした枠組の中で行なわれるべきであると考えられる。

### 3. 人間の志向姿勢に基づく HAI

人間は動いているさまざまなものに対して、志向的な姿勢を適用する人間ならではの理解を行なう。例えば、エレベータに乗った男が 5 と書いてあるボタンを押した。という状況を想定してみる。2 章で述べた Wright の目的論的な説明を行なうと次のような理解 [U1] を人間は行なうだろう。

[U1-1] 男はビルの 5 階にあるオフィスに行こうと意図した。

[U1-2] 男はエレベータの中の 5 と書いてあるボタンを押すことで 5 階に行けると信じている。

[U1-3] それゆえ、男は 5 と書いてあるボタンを押す。

このとき、5 と書いてあるボタンを押すことによる因果的な成果はエレベータが 5 階で止まることであり、男が意図した帰結はそれにより目的のオフィスに到達できることである。

次に、男の目的が単に「5 と書いてあるボタンを押すこと」であった場合、つまりそれを為すためだけに行なわれる状況を想定してみる。この場合の我々の理解 [U2] は次のようになる。

[U2-1] 男は 5 と書いてあるボタンを押すことを意図した。

[U2-2] それゆえ、男は 5 と書いてあるボタンを押す。

ここで、もしこの男が乗ったエレベータにすでに別の利用者がおり、その者がすでに 5 と書いてあるボタンを押していたらこの男はどのような行動を起こすであろう。男の意図が [U1-1] であった場合には、男はボタンを押さずにエレベータに乗って 5 階で止まった時点でエレベータを降りることを予想される。一方、男の意図に関する理解が [U2-1] であった場合には、たとえ 5 と書いてあるボタンがすでに押されている状態であっても男はそれに関わらずそのボタンを押すことが予想される。

このことからわかることは、外部的な視点から観察される男の行動のみに基づいて意図を推定することは、しばしば妥当性を欠く分析となってしまうことである。前者のケースでは、[U1] の理解に基づき男がボタンを押さなかったのは、意図した帰結を導く成果がボタンを押さなくても達成できることがわかったからである。したがって男がボタンを押さなくても男は行為を遂行したとすることができる。また後者のケースでは、[U2] の理解が仮に事実だとしても男の行動を説明する上では説得力を欠く。なぜならば [U2] の理解では男の行動は志向的でないからである。言い換えれば、男の行動は人間（エージェント）らしからぬものと通常判断される。

人間は人間やエージェントはア priori に意図的な行為を遂行する存在として捉えていると同時に、行為を通して帰結に至るために成果を得ようとする志向的な姿勢を有した存在であると考えている。したがって、この前提のもとでの人間とエージェント間の相互行為（インタラクション）すなわち HAI を議論する際には、インタラクションに参加している者の間において志向姿勢に基づいて相互の意図や信念の主観的な仮定に関する分析が必要であることが強く示唆されるのである。

### 4. エージェントとしての存在の認知

これまでの議論に基づいて検討すると、HAI において人間が関わり合う対象を志向的な存在であるエージェントとして認知するためには、エージェントが意図していると予想される行為の帰結が人間とエージェントとが共有できる環境内で実現される必要があるかもしれない。

河村らは人間と情報提示装置との共同作業を通して、情報提示装置がエージェントとして認知される要因と、提示される情報に対する態度にどのような影響を与えるかを実験を通して検証している [河村 09]。その結果、実世界における「宝探し」のような課題を行なった場合には、情報提示装置の形態は CG キャラクターや音声のみの条件と比較して実世界で環境を共有できるロボット型の条件が、与えられた情報（全条件とも共通）に対する受容態度において有意に高かった（図 1）。

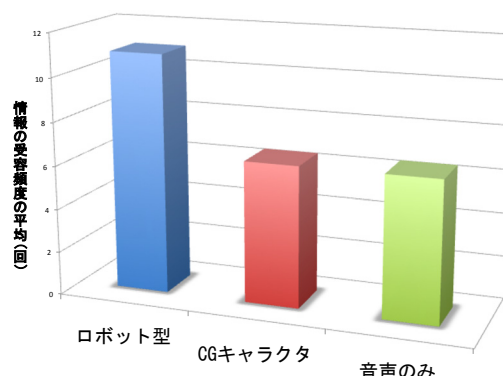


図 1: 情報提示装置の形態と情報に対する態度

今回この結果は実世界における課題に対するものであったが、知識や概念など観念的な世界を取り扱う課題の場合は、CG キャラクターあるいは音声優位の結果が得られる可能性も残されている。

### 5. おわりに

本稿では、人間の意図的な行為に関する認識とエージェント性の認知に関する議論を通して、HAI 研究ならではのインタラクション分析の観点に関して問題を提起した。

人間にとってエージェントとは何かという存在論的問題を問い直すことによって、HAI 研究が単なる現象の記述や構成論的なアプローチに基づくモノ作りだけでなく、知性に関する問題に向き合うことができるようになることが期待される。

### 参考文献

- [山田 07] 山田誠二 監修・著: 人とロボットの<間>をデザインする, 東京電気大学出版局 (2007).
- [Reeves & Nass 96] Reeves, B. & Nass, C.: The Media Equation, Cambridge University Press (1996).
- [Wright 63] von Wright, G. H.: "Norm and Action," Routledge (1963). (G. H. フォン ウリクト 著, 稲田静樹 訳: 『規範と行為の論理学』, 東海大学出版会 (2000).)
- [Dennett 96] Dennett, D. C.: "Kinds of Minds," Harper-Collins Publishers (1996). (D. C. デネット 著, 土屋俊 訳, 『心はどこにあるか』, 草思社 (1997).)
- [河村 09] 河村真吾・内藤久詞・竹内勇剛: 実世界指向インタラクションに基づく情報提示手法の提案, 電子情報通信学会論文誌 (D) (投稿中).