

志向姿勢再考

—ロボットと人間の違いはどこにあるのか?—

A Reconsideration of “Intentional Stance”:
What is the Difference between Human and Robot?

寺田 和憲 伊藤 昭
Kazunori Terada Akira Ito

岐阜大学
Gifu University

A robot is sometimes construed as a designed artifact. In the present study, we address whether and how can a robot be construed as a human being. Through an experiment investigating the human stance toward various behavior of sixty entities, we found that human construe them in terms of two axes: chaotic and intentional axis. This imply that intentions emerged from inside the robot would be a cue for a robot to be construed as a human being.

1. はじめに

哲学者 Dennett は人間が他者の振舞を理解し予測するために用いる認知バイアスとして意図スタンス (*intentional stance*), 設計スタンス (*design stance*), 物理スタンス (*physical stance*)*¹ と呼ばれる 3 つの心的姿勢の存在を指摘した [Dennett 87][Dennett 96].

このうち意図スタンスは心の理論 (*theory of mind*) と呼ばれる心的機能として発達心理学, 認知心理学の分野で広く研究されている。意図スタンスとは他者の振舞が信念 (*belief*), 願望 (*desire*), 意図 (*intention*) などの心的状態によって駆動されていると仮定し, 観察した振舞に対して具体的な心的状態を帰属することによって振舞を理解する戦略である。意図スタンスは脳内の心の理論ネットワークの賦活として捉えられており [Gallagher 03], そのような能力の欠如は自閉症として知られている。信念, 意図, 願望などの心的状態は外部から決して観察することはできず, 観察者が自身の頭の中に想定することによってのみ存在する。想定によってしか存在しえないということは, 帰属すべき心的状態のバリエーションが無数に存在することを意味する。観察者が直接知覚可能なのは表面的な振舞だけであり, 観察した振舞を駆動している心的状態を想定するためには, 発話やジェスチャーなどの明示的な情報だけでなく, 状況や表情, 仕草などの非言語情報も含め, 得られる限りの様々な情報を駆使して妥当なものを推論する必要がある [Sperber 86]。冗談や嘘, 皮肉は表面的な振舞 (発話) と心的状態が異なった典型的な例である。これらの語用論的解釈が必要な発話を理解するためには心の理論が必要不可欠だと考えられている。

我々は通常人間に対しては意図スタンスを取る。なぜなら, 他者の多様な振舞を理解するためには意図という単一のシンボルのもとに観察された行為を抽象化することが合理的だからである。そして, 意図というかなり大きな粒度で振舞を抽象化することによって, 認知的負荷の少ない認識を可能にし, 心

的な目的である意図を帰属することで, 異なる状況における新規な行動の予測を可能にする [Baron-Cohen 95][Dennett 87]. 複雑な振舞が意図を想定するだけで非常に単純なものとして理解できるようになる。

一方で, 我々は機械に対しては設計スタンスを採用する。設計スタンスによる振舞理解では心的な目的である意図を想定するのではなく, 振舞を規定しているアルゴリズムを想定する。アルゴリズムによって規定される振舞は入力を固定すると出力が一意に決まり (複雑な分岐があったとしてもアルゴリズムとして記述できるということ), システムチックに振舞を予測可能である。また, 設計的振舞の特徴として失敗や例外に対処できないということがある。一方, アルゴリズムではなく目的によって駆動される主体 (意図の主体) は同一の目的を達成するために状況に応じて手段を変えることができる [Meltzoff 95][Gergely 95]。ロボットの振舞が多様になると振舞のアルゴリズム的な解釈は破綻する。そのために, 振舞を意図という単一のシンボルのもとに抽象化する振舞理解戦略 (意図スタンス) が有効になってくるのである。

意図的な振舞や設計的な振舞は物理スタンスによる解釈も可能である。例えば意図的な主体である人間の振舞はエネルギーの燃焼や動きのダイナミクスなど物理的な観点から解釈可能である。また, 設計スタンスを使用すべき対象である機械についてもエネルギーや電子の流れといった物理的な解釈が可能である。しかしながら, 人間が通常人間や機械に対して物理的な解釈を行わないのはそのような解釈よりもより良い解釈が存在するからである。

このように人間は多様なスタンスを使い分けて対象の振舞の理解を行っている。今後, ソフトウェアエージェントやロボットなど知的な人工物の振舞いは益々人間らしくなることが予想される。そのようなロボットに対して人間はどのようなスタンスを採用するのであろうか。この問題は換言すると, ロボットが道具としての延長上にしか存在しえないのか, もしくは, 人間と同等の存在になり得るのかという問題である。設計スタンスと意図スタンスが完全に独立したものであるならば, 道具としてのロボットが意図的主体へと格上げされるためには何が必要なのであろうか。本稿の目的は, 設計スタンスと意図スタンスを連続的に論じることが可能かどうか, そして可能であるならば, それはどのような変数によってなされるのかを模索することである。

2章では, まず, 様々な対象の多様な振舞に対して人間がど

連絡先: 寺田和憲, 岐阜大学工学部応用情報学科, 〒 501-1193 岐阜市柳戸 1-1, Tel & Fax: 058-293-2792, terada@info.gifu-u.ac.jp

*1 intentional stance は [Dennett 87] の訳書である [Dennett 96] では志向姿勢と訳されている。本稿のタイトルは, この訳書からの引用で志向姿勢としているが, 本文中では intentional をより一般的な意図と訳し, 姿勢はスタンスと訳す。

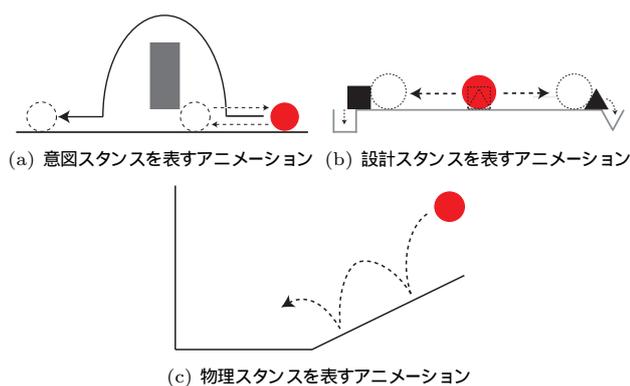


図 1: 実験で用いたアニメーション

のような心的姿勢を採用するかを網羅的に調べた実験について説明し、3章では、スタンスを規定する新たな軸である、規則性と意図性に基づいて、意図スタンスと設計スタンスの違いについて論じる。

2. 振舞理解を調べる実験

我々は Dennett の思想に基づいて、人間がどのような対象のどのような振舞に対して 3 つのスタンスのうちいずれを採用するかを調べる心理実験を行った。実験は、様々な対象の振舞が記録された 60 個の動画それぞれについて、我々の作成したスタンスを表わす 3 つのアニメーションと比較し、振舞を適切に表わすと思うアニメーションを一つ選択してもらうことによって行った。60 個の動画は、振舞う主体を動物、人間、機械、自然物、植物、ロボット、単純物、乗り物の 8 カテゴリーから選び、振舞の主体が単独であるものを用いた。同一のカテゴリーの中では、例えば動物では鳥や犬、イルカなど、できるだけ主体の種類に変化を持たせるようにした。また、振舞の選定においては、それらの主体が生成する典型的な振舞の他に他のスタンスによって解釈が可能であろう振舞が入るようにした。例えば、人間の振舞については意図的スタンスが採用されることが予想されるが、単純作業など繰り返し動作を行っている場合は設計スタンスが採用される可能性が、また落下している場合は物理スタンスが採用される可能性がある。

2.1 3 つのスタンスを表わすアニメーション

被験者が動画との比較に用いたアニメーションは次のようなものである。いずれのアニメーションにおいても赤い円が運動しており、運動はできるだけ具体性のない振舞でかつ各スタンスをうまく表現するような典型的なものになるようにした。各スタンスを表現するアニメーションの詳細は次の通りである。意図的な振舞を表すアニメーション

意図スタンスを表わすアニメーションは Gergely ら [Gergely 95] の用いたアニメーションとほとんど同じである。アニメーションは、円が長方形の物体を越える様子を表している (図 1(a) 参照)。まず、円は長方形の物体の下側を通り抜けようとする。しかし円は、長方形の物体の下側の隙間が小さいため通り抜けることができず、最初の位置へ戻る。この動作を二度繰り返した後、円は長方形の物体の上側を飛び越えて障害物の左側へ移動する。最初の 2 回の衝突は行為が完遂しないものの、その後の飛び越えによって障害物の左側へ行くための動作として理解される。この一連の振舞は異なる動作によって単一の最終状態に到達するという等結果性 (equifinality) を持っていると言える。Heider[Heider 58] は等結果性が意図帰属の

キューになると考えた。しかし、Gergely らは意図帰属のためには等結果性だけでは不十分だと考え、障害物が取り除かれた状態において無意味な跳躍をして左側へ移動するアニメーションを作成し、この振舞が奇異に感じられるかどうかによって乳児が合理性に基づいて意図を理解するかどうかを調べた。

我々は、成人の被験者が振舞から意図性を読み取るためには等結果性を備えた振舞の提示だけで充分と考え、このアニメーションとの類似性が知覚された場合に意図スタンスを採用したと考えることにした。

このアニメーションから分かる振舞の特徴は、目的志向性だけでなく、それを実現するための試行錯誤、努力、固執が挙げられる。

設計スタンス

従来の設計スタンスを表すアニメーションでは、円は単純な往復運動を繰り返す様子を表現し、振舞の機能的側面が強調できなかった。そこで本実験のアニメーションでは、特定のタスクとして、円が別の機能をもつ様子を表している。画面中央に配置された円に重なるようにして、四角形と三角形のいずれかが現れる。円は、現れた図形に応じて、それらを左右のくぼみへと押しゆく。図形がくぼみの中へ落ちると、円は再度中央に戻り、以下同様のプロセスを繰り返す (図 1(b) 参照)。意図スタンスを示すアニメーションにおいては、円が試行錯誤を経て目的に到達する様子表現することで被験者を円の内部状態 (意図) に注目させることを目標とした。一方、設計スタンスでは対象の振舞を設計者が規定した機能に基づいて解釈する [Dennett 87][Matan 01] ため、このアニメーションでは対象そのものが意図を持つのではなく、特定の入力に対して特定の出力を行うという「設計原理」によって振舞が規定されていることを表している。

この振舞の特徴は等速での移動、対象物の形状に応じた移動であるため、周期性、規則性、アルゴリズム的である。また、より高い抽象レベルで見ると、分別という機能が特徴となる。同じ振舞を繰り返すという意味では固執であるが、数回の失敗の後に成功に至るといったような目的志向的な固執ではない。

物理スタンス

物理スタンスを表現するアニメーションは、円が左上から出現し、坂を弾みながら落ちてくる。左下の平坦な部分では左の壁と右の斜面に何度か衝突しながら跳ね返りを減衰し、最終的には静止する。このアニメーションでは、円が重力や跳ね返りという物理法則に従っていることを表現している (図 1(c) 参照)。また、跳ね返りながら動きが収束することで、振舞のランダムさも表している。

2.2 実験結果

19 歳から 25 歳の男女計 32 名に対してこの実験を行った。各動画についての回答結果の類似性を定量的に明らかにするため、数量化 III 類を実施した。数量化 III 類のためのデータ化は次のように行った。横軸を動画の番号とし、縦にその列の動画に対して各被験者が選択したアニメーションを 3 ビットで数値化したものを並べる。3 ビットの数値化は、被験者が意図スタンス、設計スタンス、物理スタンスのアニメーションを選択した場合、それぞれ、100, 010, 001 とした。この結果、動画が 60 個、被験者が 32 人なので、 60×64 のクロス表が出来上がる。このデータに対してエクセル統計を用いて数量化 III 類を実施した。その結果、得られた固有値は一軸が 0.33、二軸が 0.24、三軸が 0.10 であった。二軸と三軸の間に比較的大きな差があるといえるため、ここでは一軸と二軸を採用する。一軸と二軸の累積寄与率は 28.59% であった。その結果得られた散布図を図 2 に示す。この図では、必ずしもスタンスの分

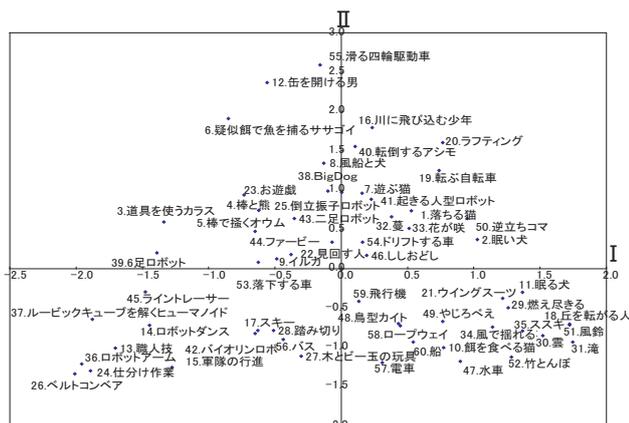


図 2: 数量化 III 類による回答結果の分析

布が類似しているものが近くに位置するのではない。異なった動画に対しても同一の選択をした被験者が多いものほど近くに位置している。言い換えると、分布が同一であっても、その分布を構成する被験者が異なる場合には離れた点に位置する。

3. スタンスを規定する新たな軸

Dennett の説明する三つのスタンスはそれぞれ独立した心的な構えであり、スタンス間の関係を記述する変数や連続性についてはあまり言及されてこなかった。我々はこれらについて明らかにするために、数量化 III 類によって各動画に対する被験者の評価を分析した。

図 2 は全 60 個の動画を被験者が採用したスタンスの類似性に基づいて二次元の連続した空間にプロットしたものである。図を見ると、意図スタンスの採用率の高いもの (6, 12, 55 など) は II 軸正方向に、設計スタンスの採用率の高いもの (24, 26, 36 など) は I 軸負方向に、物理スタンスの採用率の高いもの (18, 31, 51 など) は I 軸正方向に集まっていることが分かる。また、いずれの採用率も高いといえないもの (9, 22, 46 など) については、原点付近に集まっている。

3.1 規則性

I 軸に注目してみると、設計スタンス採用率の高いものが負方向に分布し、物理スタンス採用率の高いものが正方向に分布している。さらに、その間に意図スタンスの採用率の高いものが分布している。振舞の抽象度の高さを基準にすると、意図スタンス、設計スタンス、物理スタンスの順に並ぶのが自然だと思われる。また振舞理解戦略としての優位性という観点からも、意図スタンスが設計スタンスと物理スタンスの間に位置するとは考えられない。従って、この軸を説明するためには抽象度や戦略の優位性とは別の解釈を考えなければならない。我々は、この分布を統一的に説明するのは規則性、もしくは振舞の予測可能性だと考える。すなわち、負方向に行くほどなんらかのアルゴリズムによって振舞が決定論的に規程されているため容易に予測可能であり、正方向に行くほど、振舞に規則性がなく予測が困難だということである。丘を転がる人や風鈴、滝が同じようなところに位置しており、また落下して跳ね返る円との高い類似性を得たことは、これらの振舞を重力や反発などの法則的な物理現象として捉えたのではなく、物理的振舞のカオス性に注目したためだと考えられる。Dennett の説明においても我々のアニメーションの設計においても、設計スタンスでは機能性が、物理スタンスでは物理法則性が振舞を同定す

るための主となる特徴だと考えていた。しかしながら、この軸から読み取れることは、振舞理解において第一に考慮される特徴は機能性や物理法則性といった解釈に意味づけを要するものではなく、振舞が予測可能か否かという極めて単純な特徴だということである。

3.2 意図性

次に II 軸に注目する。II 軸では意図スタンスが正方向に分布しており、原点から負方向にかけて設計スタンスと物理スタンスが分布している。このため、この軸は意図的なものかそれ以外分類する軸だと言える。また、値が大きくなるにつれて意図スタンスの採用率の高い動画が分布している。これらのことから、この軸は意図性を表す軸だと考えることができる。意図性の高いものは I 軸上では原点付近に分布しており、意図性の低いものは I 軸上に一様に分布している。言い換えると、規則性の軸 (I 軸) において規則的でもカオス的でもないものは意図性の軸によって解釈されるのではないかと考えられる。

規則性とカオス性の中間に位置するところに意図性の軸が存在するという事はカオスの縁というアナロジーによって説明することができる。カオスの縁とはセルラーオートマトンの振舞を評価するパラメータのある範囲をさす [Langton 90]。このパラメータは 0 と 1 の間の値を取り、0 に近いところではセルラーオートマトンは静止もしくは周期的振舞をし、1 に近いところではカオス的な振舞をする。その間に位置する値においては、定常状態でもカオス状態でもない複雑な振舞をする。規則的な振舞からカオス的な振舞へと転移する領域であることからカオスの縁と呼ばれる。また、生命が決定論的でもなくカオス的でもなくその中間のカオスの縁に位置する存在だと考える研究者も多い。

我々の発見した意図性の軸はまさにこのカオスの縁における系の振舞と同じものを表していると言って良いだろう。カオスの縁でセルラーオートマトンは過渡状態が長い複雑な振舞をし、予測可能な部分もあれば予測不可能な部分もある。その意味においては目的志向的な振舞と同じである。なぜなら、目的のもとに振舞っている主体の行動は目的さえ同定してしまえばアルゴリズム的に解釈が可能だからである。しかし、目的がどのように遷移するかは全く予想できない。すなわち、人間を含む生物は遷移する短期的な目的を次々に辿っていく存在であり、その意味で規則的でもなくかと言ってカオス的でもない。この目的志向的な振舞における予測可能性と予測不能性の間がかオスの縁における振舞との共通性だと考える。

3.3 振舞の機能的解釈と設計スタンス

人工物の振舞理解においては機能的解釈が欠かせない。しかし、数量化 III 類による分析の結果、人工物を同定するために機能という軸が必ずしも必要ではなく、むしろ規則という側面の重要性が示唆された。では、機能とは何なのか？

認知科学における人工物解釈の研究では、人間は人工物の分類や名前付けのために、その人工物の現在の用途ではなく、設計者によって最初に意図された機能を重視すると言われていた [Matan 01][Kelemen 07]。しかし、Dennett は古いアイロンがブックエンドとして使われるように、人工物を規定するのは現在の使われ方だと言う [Dennett 90]。いずれにしても、人工物解釈においてはそれがどのような目的を持っているかを知ることが重要になってくる。ここで言う人工物の目的は意図スタンスで言うところの心的な目的とは少し異なる。「アイロンは服の皺を伸ばすためのもの」や「傘は雨を避けるためのもの」といったような目的論 (teleological) 的解釈を指す。この目的論的解釈は意図スタンスと密接に関連している。Kelemen

は、幼児が「山は登るためのもの」「雲は雨を降らせるためのもの」などのように、様々な主体に対して偏執的に目的論的解釈をしてしまうことと、創造主としての意図的主体を想定することの間に関連があることを主張し [Kelemen 05], 設計スタンスが意図スタンスから分化すると述べている [Kelemen 07]. これは発達心理学の知見からも妥当なものである. 近年の研究によって 4 から 6 歳ぐらいの間に設計スタンス的理解をするようになっていわれているが, これは意図スタンスの萌芽よりも随分遅い. そこで, Kelemen は次のような 3 段階を考えた. 子供は, まず最初に人工物が意図的な終端の手段であることを理解する. そして, 人工物がその目的を体現するものとしての理解がなされる. そして最終的に, それほど明確に意図を想定しなくても目的そのものだけを直接理解可能な設計スタンスが確立するというものである.

振舞としての設計はそもそもかなり目的志向性の強いものである. すなわち, 設計スタンスとは暗黙的に意図を想定するものなのである. 設計スタンスにおける意図は本来不可分であるが, 我々の実験結果は人間の設計的振舞の理解において, 規則性そのものにだけ注目している可能性を示唆している.

そこで, 我々は, 機能とは規則性の意図の利用だと考える. 人工物は利用状況と密接に結びついているため利用者を想定しないことは困難である. しかし, 利用者を想定しなければ, そこに存在するのは規則でしかない. 例えば自動販売機は, 特定の質量と外形を有する円盤を投入し, 特定の可動突起を押すとその突起の位置に依存した缶入りの液体が出力される装置である. 2 つの入力によって一つの出力を得るという規則を持っている. この規則はその装置が存在する限り, 全ての行動主体に対して同じく存在している. しかし, その法則が機能として働くのは人間などの行動主体がジュースを手に入れるという目的を持って行動したときだけだ.

一方で, 物理的な振舞であっても利用者を想定すると機能が発現することがある. 例えば沸騰という現象は, 水を熱してもある特定の温度までしか上昇しないという規則を持つ. この法的な振舞は同条件下では再現性を持って同じ結果を得ることができ, 大気圧下で 100 と大幅に異なる温度で沸騰することはない. この法則は人間の存在に関係なく普遍的に存在する. しかしひとたびこの法則を調理という特定の目的のために利用すると, その法則は機能を持つことになる.

このように, 目的なき機能は存在せず, 機能から目的を除去すると法則しか残らない. 以上のことから, 設計スタンスは利用者としての行動主体を想定しなければならないために, 振舞理解戦略のモデルとしては複雑だと言わざるを得ない. より簡潔に説明可能なのは上述の通り, 規則性と意図性の軸による振舞理解戦略モデルだと考える.

4. おわりに

Dennett の提案する 3 つのスタンスはそれぞれ独立したスタンスであると考えられている. しかし本稿では, 我々の実験結果に基づき, 規則性と意図性という軸を導入することによって 3 つのスタンスを連続的に説明できる可能性を示した. そして, 人間が設計の対象の振舞を理解する際に, 意図性を排除して規則性だけに注目している可能性を示した.

ロボットが人間の意図を実現するための手段から意図の存在へと格上げされるためには, 他者によって埋め込まれた目的ではなく, ロボットの内部から湧き出してくるような目的を持つことが必要である.

参考文献

- [Baron-Cohen 95] Baron-Cohen, S.: *Mindblindness: An Essay on Autism and Theory of Mind*, The MIT Press (1995)
- [Bloom 96] Bloom, P.: Intention, history, and artifact concepts, *Cognition*, Vol. 60, No. 1, pp. 1 – 29 (1996)
- [Dennett 87] Dennett, D. C.: *The Intentional Stance*, Cambridge, Mass, Bradford Books/MIT Press (1987)
- [Dennett 90] Dennett, D. C.: The Interpretation of Texts, People and Other Artifacts, *Philosophy and phenomenological research*, Vol. 50, pp. 177–194 (1990)
- [Dennett 96] Dennett, D. C.: 「志向姿勢」の哲学 –人は人の行動を読むのか?–, 白揚社 (1996)
- [Gallagher 03] Gallagher, H. and Frith, C.: Functional imaging of 'theory of mind', *Trends in Cognitive Science*, Vol. 7, No. 2, pp. 77–83 (2003)
- [Gergely 95] Gergely, G., Nádasdy, Z., Csibra, G., and Bíró, S.: Taking the intentional stance at 12 months of age, *Cognition*, Vol. 56, No. 2, pp. 165–193 (1995)
- [Heider 58] Heider, F.: *The Psychology of Interpersonal Relations*, Lawrence Erlbaum Associates (1958)
- [Kelemen 05] Kelemen, D. and DiYanni, C.: Intuitions About Origins: Purpose and Intelligent Design in Children's Reasoning About Nature, *Journal of Cognition and Development*, Vol. 6, No. 1, pp. 3–31 (2005)
- [Kelemen 07] Kelemen, D. and Carey, S.: The Essence of Artifacts: Developing the Design Stance, in *Creations of the mind: Theories of artifacts and their representation*, Oxford University Press (2007)
- [Kelemen 09] Kelemen, D. and Rosset, E.: The Human Function Compunction: Teleological explanation in adults, *Cognition*, Vol. 111, No. 1, pp. 138 – 143 (2009)
- [Langton 90] Langton, C. G.: Computation at the edge of chaos: phase transitions and emergent computation, *Physica D*, Vol. 42, No. 1-3, pp. 12–37 (1990)
- [Matan 01] Matan, A. and Carey, S.: Developmental changes within the core of artifact concepts, *Cognition*, Vol. 78, pp. 1–26 (2001)
- [Meltzoff 95] Meltzoff, A. N.: Understanding the Intentions of Others: Re-enactment of Intended Acts by 18-Month-Old Children, *Developmental Psychology*, Vol. 31, No. 5, pp. 838–50 (1995)
- [Sperber 86] Sperber, D. and Wilson, D.: *Relevance: Communication and Cognition*, Oxford: Blackwell (1986)
- [Watson 05] Watson, J.: Causal logic and the intentional stance., G. Origg and A. Reboul (Moderators) Causality: Internet Conference at www.interdisciplines.org/causality (2005)