

回帰的意図理解をめざす共同注意ロボット

金野武司・柴田正良

Abstract

Human beings can convey their intentions to other people non-verbally, through looks and gestures. It is known that during development, infants acquire this ability before they learn to speak. We have focused on the phenomenon of “joint attention,” wherein infants try to look at objects by following their parent’s gaze. This behavior begins as a reflective one, but then progresses further into the next stage—understanding and sharing other peoples’ intentions. We try to clarify joint attention by constructing a simple robot that visually interacts with human beings. In this paper, we suggest two important mechanisms for understanding and sharing intentions; an association mechanism to produce an intentional gaze following, and an inference mechanism to recursively rectify the robot’s own intentions by understanding the interacting person’s intentions.

1. 視線の意図的な使用とその理解

我々はコミュニケーションの中で視線を意図的に交わすことがある。二人がテーブルに向かい合って食事をしている。雑多に並んだ食事皿と共に、テーブルには塩の瓶が置かれている。一方が手の届かない場所にある塩の瓶を注視したとき、もう一方はそれに気づき、塩の瓶を手渡す。この場面で重要なのは、人は言葉なしにでも視線を通じて互いの意図を共有できるという点だ。ここで我々が問いたいのは、人は視線のやり取りを通じて、どのようなメカニズムによって互いの意図を理解し合っているのかということである。

日常のコミュニケーションでは、視線やジェスチャといった非言語的な行動と言語は切り離すことができない。このため、意図の理解に言語が必要であるかどうかははっきりしない。ただし、乳幼児の発達過程においては、意図を理解し合う能力の原型は言語に先立って獲得されることがいくつかの

実験により確かめられている (Gergely et al., 2002; Itakura et al., 2008).

視線に対する乳幼児の理解に関して、共同注意という現象が注目されている。この現象は、身体動作としては「親が視線を向ける方向に子どもも視線を向ける行動」として定義されているが、重要なのは、乳幼児は早ければ12ヶ月程度の頃から、他者の視線を意図的に捉え始めるようになるということである (Tomasello, 1995)。

ところが、言語を獲得する以前の月齢 (24ヶ月以前) では、乳幼児が視線のやりとりから他者の意図を理解しているかどうかを示す明白な証拠を見つけることが難しい。なぜなら、その頃の視線のやりとりは、単純な刺激-反応パターンとして見れば、ほとんどが他者意図の理解を持ち出さずとも説明できるからである。しかし、視線のやりとりにおいて人が他者の意図を理解できることは、冒頭の例のような日常的な体験からは明らかである。このギャップを埋めていくためには、意図のような主観的な内部状態の形成メカニズムを客観的に記述しながら、行動の変化を内部メカニズムの違いとして説明し、さらにその行動の違いが持つ効果を検証していく必要がある。

我々は、ロボットを用いればこういった問題の一端を解決することができると考えている。我々が用いるのは、ロボットやコンピュータシミュレーションといった人工システムで対象のモデルを実作し、そのモデルの観察から概念分析を進める手法である。これは構成的アプローチと呼ばれている¹。

共同注意の発達に対して構成的アプローチを適用する。そのためにまず、共同注意の意図理解に関する発達過程を素描しよう。トマセロは、乳幼児が他者の視線を意図の観点から理解するようになる過程を3つの期間 (9ヶ月以前, 9~18ヶ月, 18~24ヶ月) に分けて説明している。9ヶ月以前の乳児では、親の視線を追従したり (Hood et al., 1998), あるいは親の視線の方向を見れば何か面白いものがあるという条件付けによって視線の移動が起こることが確認されている (Corkum and Moore, 1995)。しかし、この段階の乳児は自発的に親と対象の間で視線を行き来させないことや、大人の注意を確かめるために振り返るような行動が出現しないために、一般に、親の注意がどこに向けられているのかを気にかけている様子は伺えないとされている。この意味で、9ヶ月以前の乳児が見せるのは反射的な共同注意である。

9ヶ月を過ぎると、乳幼児はある対象に向けられた親の視線を追従した後で、親を振り返って親の視線がまだそこにあることを確認したり (参照視と呼ばれる), あるいは (しばしば指さしと共に) 自らが自発的に向けた視線に関して、親と対象の間で視線を行き来させる行動を見せるようになる (交

互凝視と呼ばれる)。このような行動は、単に相手の視線方向に自らの視線を向けているのではなく、他者の注意がどこにあるのかということに関心を持ち始めた証拠である。さらに12ヶ月頃を過ぎると、乳幼児はある対象を注視した後で親の方を振り返り、その表情によって自らの行動を調整するようになっていたり（社会的参照と呼ばれる）、自らが注意を向ける対象に親の注意が向くまで、ときおり呼びかけるような声を発しながら交互凝視を続けたりするようになる（教示的な交互凝視）。これらの行動では、親や対象を注視するという行為が、親の注意の所在を知ることや、親の注意を自らの注意に向けさせるといった目的の手段になっている。このように、ある段階から、幼児の注意行為は二階の注意行為を意図的に実現するための一階の手段的行為となる。またその際、一般的に幼児は、他者の行為の観察からその行為の目的も推論するようになって考えられている。ただし、この頃の幼児はいつでも他者の目的を正しく理解できるわけでもなければ、他者に自分の目的を正しく伝えて、自分の目的をいつでも達成できるわけでもない。

こういったことが適切に行なえるようになるのは、16ヶ月を過ぎて幼児が簡単な単語をしゃべる頃からである。幼児が他者の注意の所在を探ることで新しい単語を学ぶ場面（単語理解）や、他者に注意の焦点を向けさせるために言語を使う場面（単語産出）で、幼児は自分の意図や他者の意図をより深く理解していくようになる。幼児は簡単な単語を用いる場合であっても、単語理解の場面では対象に関する大人の意図を理解する前に、大人が自分の注意をどこに向けさせようと意図しているのかを理解していなくてはならない。また同様に単語産出の場面では、幼児は対象に関する自分の意図を大人に理解させる前に、自分が大人の注意を特定の対象に向けさせようと意図していることをその大人に理解させていなくてはならない。これを行なうためには、自分の意図を大人が理解しているかどうかを判断しながら行動を調整する必要が生じる。ここには、＜自らの意図的な行為に向けられる大人の意図＞を子どもがまた理解するという回帰的な構造があり、これが他者と注意を共有するための基本的なメカニズムになると言われている。

我々は以上の3つの過程を次のようにまとめ、そのモデル化を進めている。

- ▶ プロセス1: 視線追従や条件付けによる共同注意の段階から、意図的に注意を向けるようになる過程
- ▶ プロセス2: 自らの意図的な行為選択において、他者の行為を意図の観点から理解するようになる過程
- ▶ プロセス3: 自らの意図に向けられた他者の意図を回帰的に理解するようになる過程

3. 自らが意図的に注意を向けるためのメカニズム（プロセス1）

プロセス1で最初に作成するのは、9ヶ月以前の乳児に相当する反射的な共同注意のシステムである。この共同注意は、まず親の視線方向に自らの視線を向けて、その後で視界に映るオブジェクトを注視する。前者は視線追従、後者は視覚定位と呼ばれる。発達過程における発生順では視覚定位の方が早く、3ヶ月程度の乳児ですでに、視界の端に映る刺激に対する定位が起こり始める。そのため、発達の第一段階として視覚定位を、第二段階として視線追従のシステムを考えよう。

視線運動のモデルでは、基本的に視覚刺激を入力、視線の移動を出力としたシステムの内部メカニズムを検討する。このとき視覚定位は、外部空間（視野）と刺激の位置を何らかの形で内部表現し、かつ視界をその内部表現の中で移動させる仕組みを作れば、あとは刺激の位置に対して視界を逐次調整するメカニズムを加えることで実現できる。こういった調整の仕組みは、強化学習やニューラルネットワークのような学習アルゴリズムによって実現できることが知られている（Nagai et al., 2006）。ただし、その仕組みの中には意図的の行為を生成するメカニズムが含まれてはいないという意味で我々のテーマではないため、ここでは詳述しない²。

続いて第二段階の視線追従のモデルを考えよう。先述したように、乳幼児は6～8ヶ月頃になると、親の視線方向（発達の初期段階では視線よりも顔の向き）に反応してそちらの方向に自らの視線を向けるようになる。この行動のメカニズムもまた、視覚刺激として人の顔の向きや視線方向を検出できることを仮定すれば、先ほどの視覚定位のシステムと同じように考えることができる。こうして作成した視覚定位と視線追従のシステムを図1左のように並列接続してみよう。このシステムは、親を注視したときには視線追従によって親の視線方向に自らの視線を移動させ、視線を向けた先では視覚定位によってそこにある何らかの対象を注視する。これによって、親の視線の先にある対象を注視するという共同注意行動が実現される。

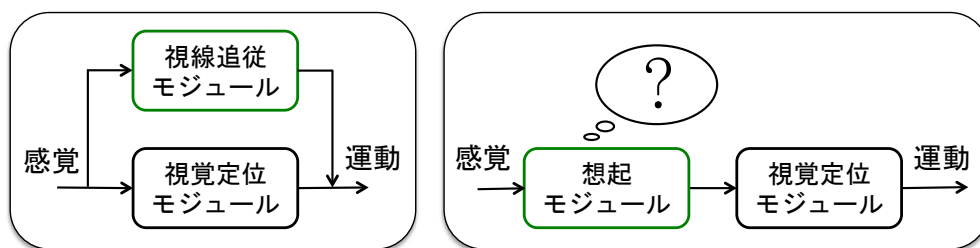


図1. 並列接続システム（左）と直列接続システム（右）

第二段階までのメカニズムでは、共同注意の発達過程において指摘されているような意図的行為は生成されていない。つまり、第二段階までのメカニズムは感覚刺激から直に行為目的が決定されており、感覚刺激が複数の行為目的を誘発して、その中からそのつど行為目的が選択されるようなシステムは実現されていない。したがって、プロセス1を実現するためには、例えば視覚定位のような行動を実行手段として用いる内部状態（行為目的）が形成され、そこで共同注意を実現するための目的選択が行なわれる必要がある。

これを実現するメカニズムとして我々が考えるのは、視覚定位のシステムに目的を形成するモジュールを直列に前置接続することである（図1右）。それでは、この前置接続するモジュールによって共同注意を実現するにはどうすれば良いか。ここで思い出さなければならないのは、視覚定位のシステムは、注視対象に由来する視覚刺激を入力として要求するということだ。したがって、親の顔の注視からその視線方向を検出したとき、前置接続のモジュールが出力すべきなのは、運動方向ではなく視覚刺激である。つまり、前置接続のモジュールが、親の視線方向からその先にあるだろう対象の刺激とその位置を仮想的に生成するならば、視覚定位モジュールによって視線はその方向を向くことになる。

このメカニズムの動作は次のようになる。まず親の顔を注視したときの刺激を前置接続のモジュールが受け取ると、その視線の先にあるだろう何らかの対象とその位置が想起される。そのデータが視覚定位モジュールに渡されると、システムはその想起された対象に向けて自らの視線を移動させる。この行動（視線移動）は注視対象の想起によって初めて可能となり、その意味で視覚定位は対象を注視するための手段として働く。

金野・橋本(2008)はこの仕組みの振る舞いをコンピュータシミュレーションによって確認した。直列接続システム（図1右）は、親の視線方向に複数のオブジェクトがある場合、想起された視覚刺激に基づいて注視対象が決定されるという特徴を持つようになる。これに対して並列接続システム（図1左）では、注視対象は視線を向けた先で決定されるので、ここに大きな違いが生じる。また直列接続システムでは、想起されたオブジェクトが視界にない場合に、その状態を次の行動を起こすための動機として利用することができる。これによって例えば、想起されたオブジェクトが見つからないことを理由に親を振り返る、といった参照視行動を生成することも可能になる。

ただし、直列接続システムが共同注意を行なうためには、親の視線の先にあるオブジェクトを注視する体験があらかじめ必要になる。もし、親が右側を見たときにそれとは逆の左側を見て、そこにあるオブジェクトを注視した

経験を記憶すると、次に右側を見ている親を見たときには左側に置かれたオブジェクトを想起して、結果として親とは逆の方向を見ることになってしまう。この意味で、反射的な段階の行動モジュールがまず機能を果たしていることが重要になる。反射的な共同注意（視線追従と視覚定位）は、親の視線から適切な対象を想起する記憶の形成に必要であり、この行動は大人になってからも日常的に重要な役割を果たしている。なぜなら乳幼児も大人も新たな文脈や突発的な状況では、他者が視線を向ける方向に反射的に視線を向けることで、他者の注意に自然に気付く機会を得ることができるからである。

モデルの作成から分かるのは、手段となるような行動はそれ自体が目的を形成するための体験を提供する役割を持っているということである。通常考えられるのは、情動に付随する報酬系の働きによって記憶が構成される仕組みだが（もちろんその仕組みも働いてはいるが）、意図的な行動へと向かう発達過程においては、そもそも反射的な行動の持つ特徴が、次のステップで獲得する能力を方向付けるといふ大きな役割を持っている。恐らく冒頭の塩の瓶のコミュニケーションでも、最初の視線のやりとりでは反射的な視線の運動が起こるのではないかと考えられる。しかし、視線を塩の瓶と相手の顔の間で行き来させることで、相手の視線から塩の瓶が想起され、それを取ってもらうことが相手の意図なのではないかという推論が働き始める。その推論を始めるきっかけであり材料となるのが、今回の設定状況では先述のような想起のメカニズムではないかと我々は考えている。

4. 人とロボットのインタラクション実験

我々は、プロセス 1 を構成する 2 つのシステム（並列接続と直列接続のシステム）が、人にどのような影響を与えるのかを調査した。この調査にあたって立てた仮説は、意図的な行動を生成するシステムが実装されたときの方が、人はロボットが何かを見ようとしていると感じることが多くなるのではないかというものだ。これを検討するため、我々は図 2 の吹き出しにあるようなロボットを開発した。このロボットは、目に相当するステレオカメラを備え、その視界を上下左右（パン・チルト）に動かすことができる。眼球に相当するものはなく、人はロボットのカメラの向きによって注視対象を判断することになる。人と視覚的なやりとりができれば良いので、ロボットの外観は極めてシンプルなものになっている。ロボットはテーブルに設置され、ロボットの前に実験の参加者が座る（図 2）。テーブルには 4 つのゴムボール（色は赤、青、黄、緑）が、参加者から見て「ハの字」型に置かれている。インタラクション中は、このボールの配置は変更しない。

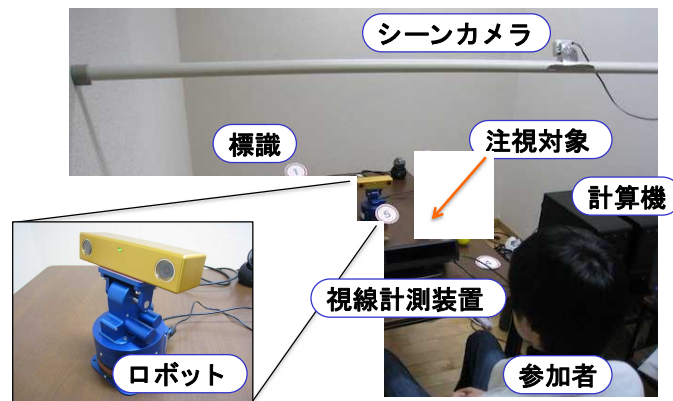


図 2. ロボットと実験環境

実験は極めてシンプルである。参加者は 2 つのタイプのロボットと 3 分間、視線のみでインタラクションした。このとき、参加者は 4 つのゴムボールのいずれかを注視するように指示された。参加者には、ロボットは視覚的なコミュニケーションだけを行ない、参加者の視線を読み取ってその方向を見ることが伝えられた。さらに、実験にはタイプの違うロボットが用意されており、それぞれは参加者が見るものをどのように理解するかという点に違いがあるだけだ、ということが参加者に告げられた。

実験の最初では、ロボットは参加者の顔を視界に収める状態で正面を向き、参加者の顔を検出してその位置（自分の首の角度）を憶える。参加者の視線はロボットのカメラによってではなく、参加者の前に設置された視線計測装置によって計測される³。この装置は参加者が視線をボールに向けると、その視線方向を 30 度の解像度で検出する⁴。参加者は最初に並列接続システムを体験する。これは、並列接続システムに関する参加者との経験を、直列接続システムの学習データとするためである。直列接続システムにおける記憶は、参加者の視線方向にロボットが追従し、その後に注視した各ボールの頻度を蓄積して、その頻度に従ったルーレット選択によってロボットが注視ボールを想起するというやり方で形成される。これによってロボットは、人が最も多く注視したボールを自分の見ようとするボールとしてほぼ常に想起するようになる。ロボットはいずれかのボールを注視すると、最初に憶えた人の顔の位置に視界を戻すように設定されている。これは、参照視行動をロボットに埋め込んだものであり、その位置づけは反射的な行動になる。

大学生 6 人と研究者 2 人を合わせた 8 人にこの実験に参加してもらい、実験後にアンケートを実施した。アンケートは 5 段階尺度を用いて、5 を質問に対する肯定、1 を否定、そして 3 をどちらでもないとした。結果を報告する前に断っておきたいのは、この実験は厳密な仮説を検証する前の予備的

なものであるという点だ。通常はこのような段階の実験結果を報告することはできないが、構成的アプローチの研究においては哲学者との間でのこういった段階からの議論が有効となる。以下は、それを1つの事例として示すことを目的とした報告である。

アンケートでの最初の基本的な問いは次の2つである。一つは「ロボットにあなたの見ているものが伝わっていたか？」であり、もう一つは「あなたはロボットの見ているものが分かったか？」である。この2つの質問に対する回答の平均は、それぞれ3.48と3.41(標準偏差はそれぞれ1.70と1.18)だった。この回答からおおよそ把握できるのは、ロボットは自分(参加者)の見るものをだいたい見ていたと参加者は感じており、かつ、参加者はロボットの見ているものもだいたい分かったと感じていたということである。

そして、検証する仮説に関して「あなたはロボット自らが何かを見ようとしているように感じたことはあったか？」を質問してみると、その回答の平均は3.86(標準偏差=1.04)だった。この結果から、参加者は、我々が用意したシンプルなロボットに対しても何かを見ようとする主体であるかのように接したことが伺える。続いて、「感じることはあった」と回答した5人の参加者に、どのタイプのロボット(何番目に接したロボット)にそれを感じたかを複数回答を許して尋ねてみると、並列接続システムが3ポイント、直列接続システムが4ポイントと、どちらのシステムに対しても、自ら見ようとしている状態をロボットに感じていることが分かった。

この結果は、ロボットの見ようとする意図性の受け取り方に関して、システムが並列型か直列型かによっては違いがないことを示している。ところが、それぞれのロボットのどのような行動に意図性を感じたかを自由記述で尋ねると、そこには若干の違いがあることが判明した。その報告で我々が特に注目したのは、次のような2つの相反する報告である。一つは、「自分がボール以外のところを見たときに、ロボットもそちらを見たとき」というものであり、もう一つは、「自分がボール以外のところを見たときに、ロボットがボールを見たとき」というものだ。前者は並列接続システムに対する報告であり、後者は直列接続システムに対する報告である。

ここで、2つのタイプのロボットが実際にどのような動きを見せるのかを整理しておこう。並列接続システムでは、まず視線追従が働いて、参加者の視線方向に視線を向け、その後で視界に映るボールを注視する。このため、ボールのないところを参加者が見えていても、ロボットはまず参加者の見ている方に視界を向ける。その視界内にボールが映っていればロボットはそのボールを注視する。他方、直列接続システムでは、参加者の視線方向からその

方向にあるだろうボールを想起することで、そちらに視線を向ける運動が生成される。このため、参加者がボールのないところを見た場合にも、ロボットは特定のボールを想起して、そこに一気に視界を移動させることになる。

それぞれのロボットが参加者に対して同じように「何かを見ようとしている」と感じさせた行動には、2つのシステムが働く状況に違いがある。並列接続システムは「ボールを見る」という文脈が形成される以前から機能しており、直列接続システムは、「ボールを見る」という文脈が形成された後で機能している。つまり、前者は文脈のない状況（コミュニケーションの最初や突発的イベント）において機能するシステムであり、後者は、ボールを見るという文脈において機能するシステムだと考えられる。ここで我々が重要視するのは、やはり後者である。もちろん、後者のシステムだけではコミュニケーションの最初や突発的なイベントにうまく対処することはできない。しかし前者のシステムだけでも、文脈に沿った注視対象を的確かつ迅速に定位して円滑なコミュニケーションを実現する、ということはいかならないだろう。そしてこれは、単に視線の解像度が上がれば解決するという違いではない。なぜなら、人は他者の視線方向を毎回のようには幾何学的に分析して注視対象を定位するのではなく、数回のインタラクションの中で文脈を形成し、その文脈の中で見ているだろうものを推定し、それを定位するスタイルを持っているからである。ただし、当然のことながらこの段階でのロボット（以降、これを第三段階システムのロボットと呼ぶ）は、頻度分布という形で相手が最も多く見ているボールを見るという文脈は形成できても、他者の意図に沿った文脈を理解できるわけではない。こういったことができるようになるためには、プロセス 2, 3 の発達過程を検討していく必要がある。

5. 他者と注意を共有するためのメカニズム（プロセス 2, 3）

5-1. 他者の意図を理解しようとするロボット（プロセス 2）

我々が現時点で示すことができるのは、あくまで、具体的なシステムとプロセス 2, 3 の意図理解との間にある概念的な分析図である。まずは、プロセス 2 における第四段階ロボットの作動原理を考えてみよう。

プロセス 2: 自らの意図的な行為選択において、他者の行為を意図の観点から理解するようになる過程

ここで我々は、これまでの第三段階ロボットがもつ 3 つの重要な制約を乗り越えなければならない。一つは自分と他者の区別である。実は、これまでのロボットにおいては、他者の意図と区別された限りでの自分の意図というものは存在していなかった（そしてその逆でもある）。それは、他者の視線

から想起されたオブジェクトが直ちに自分の意図的な注視対象になる、という点からも明らかだ。したがって、第四段階のロボットにおいては、この自他の区別が何らかの形で行為選択の内部メカニズムに組み込まなければならない。第二は、ロボットの行為選択の自律性をこれまでより高める必要があるという点だ。これは、自分の見たい対象が必ずしも共同注意の経験によって供給されずとも、例えば単なる外部空間探索など、共同注意とはまったく別の文脈から生ずることもある、ということの意味する。第三は、行為選択に影響を及ぼすような内部状態、以下で「情動タイプ」と意味づけるような状態を、ロボットに組み込む点である。これは、意図的行動が、現在の刺激から相対的に独立して、(内部状態としての)ロボットの過去の経験から生成されることを可能とするための仕掛けだ。

さらに、こうした区別を明確にするために、第四段階システムの説明に新たな表現を導入しよう。第四段階システムでは、「ある対象が自分の〈見たい対象〉であること」を、「欲求ボックス」へのその対象(の内部表現)の出現として表現し、「自分の〈見ようとする対象〉であること」を、「意図ボックス」へのその出現として表現する。例えば、ある想起対象 O1 が自分の〈見たい対象〉の一つとなり、さらにそれが〈見ようとする対象〉として選ばれるということは、O1 が「欲求ボックス」の中に出現し、それから「意図ボックス」に送られる、というように表現される。

すると、この第四段階システムの動作は次のようになるだろう。まず、他者の視線を捉えると、直列接続システムと同様に、このシステムは過去の共同注意経験からある特定の対象を想起する。しかし、第三段階システムとは異なり、これが直ちに自分の〈見ようとする対象〉となるわけではない。自分の意図する注視対象は、自分が推論した結果として選ばれる〈他者の注視対象〉である。推論による自己の注視対象のこの優先的な選択が、「他者の行為(対象注視)を意図の観点から理解する」ことの内実だと我々は考える。

さて次に、この第四段階システムは、どのようにして他者の注視対象を推論するのだろうか。ここで我々は、この段階の最も重要なポイントとして、自分の〈情動－意図〉の内的連関性を他者へと投影することによって他者の意図を理解しようとするメカニズムを、第四段階システムに組み込むことにする。これは、自分の経験した対象と自分の情動状態との対応を手がかりに、現在の他者の情動状態から、それと対応する対象を他者の意図的注視対象として推論するメカニズムだ。システムが持つ情動タイプは極めて少数の単純化されたものであり、他者はそれと同一の情動タイプを表情タイプとして持つというように設定しておく。ここでは、情動と対象の関係は、「タイ

プ」対「タイプ」の対応づけによって与えられる。例えば、タイプとしての「赤いボール」にはシステムの「面白い」という情動タイプが対応し、「面白い」は、他者が示す表情タイプ「笑い顔」と同一の情動タイプとする。また別のタイプの対象にはシステムの「怖い」という情動タイプが対応し、それは他者の「怒り顔」と同一の情動タイプである、等々。つまり、第四段階システムは、他者の視線方向と同時にその人の表情（情動）タイプを読み取り、また同時に、注視対象に対応した自分の情動タイプも呼び起こす。

そこで、他者の視線方向がこのシステムによって捉えられたとき、どうなるか。まず、このシステムでは、その視線方向と過去においてペアになったことのある複数の対象が想起によって「欲求ボックス」に生ずる。しかし、このシステムは、ここで直ちにその対象群の中から第三段階と同じやり方で対象を選び出すのではなく、もう一つ別のデータと対象群を照らし合わせる。そのデータとは、システムが同時に読み取っている他者の表情タイプだ。この表情タイプを手がかりに、システムは、それと同じタイプの自分の情動とペアになっている（タイプの）対象を、候補となった対象群の中から選び出す。こうして選択された対象が他者の意図的な注視対象として「意図ボックス」の中に送り込まれ、それを見るためのメカニズムが作動を開始する。

もう少し具体的に述べよう。例えば、システムが他者の視線方向 G1 と表情（＝情動）タイプ E3 を検出すると、それによって、G1 と過去においてペア（<G1, O1>, <G1, O2>, <G1, O3>）になっている対象群（O1, O2, O3）が想起される。それらの各々には、システムのこれまでの経験全体を通して、特定の情動タイプが割り当てられている（<O1, E1>, <O2, E2>, <O3, E3>）。そこでこのシステムは、他者の表情タイプ E3 を基に、それに割り当てられている O3 をその人の意図的な対象だと推論するわけである。この推論プロセスに託した我々の狙いはこうだ。「この人が見ているものは、この3つのうちのどれだろう？ この人は笑っている。ということは、この人は面白がっているに違いない。この3つのうちで自分が面白かったのは赤いボールを見たときだから、この人もきっとそれを見て面白いから笑っているんだろう。だから、この人は赤いボールを見ているんだ」。

ところで、このシステムはどうやって、それぞれのタイプの対象と自分の情動タイプを結びつけるのか？ またどうやって、自分の情動タイプに他者の表情タイプを結びつけるのか？ このシステムはまだ自律的な情動メカニズムを持っていないので、「面白い」とか「怖い」といった通常の意味を持った情動タイプを対象に対応させることは不可能だ。そこで、このシステムは、共同注意を含め過去に経験したすべての対象タイプに、その注視頻度順

に従って、一連の情動タイプを「機械的に」割り当てる。したがって例えば、システムが最も多く見た対象タイプ「赤いボール」に「面白い」という情動タイプが結びつけられ、次に注視頻度の高い対象タイプ「青いコーン」に「怖い」が結びつけられても、そのこと自体に情動本来の意味はない。むしろ重要な点は、他者との共同注意という文脈はこのシステムの過去の全経験の一部をなすにすぎないという点だ。したがって、他者との共同注意の経験において「赤いボール」は2番目の注視頻度しか持たないとしても、それを含めた経験全体の中では「赤いボール」が最も注視頻度が高い場合には、それにシステムの情動タイプ「面白い」が結びつけられることになる。では、この情動タイプ「面白い」は、どうやって他者の表情タイプと結びつくのか。それは、共同注意の文脈において最も出現頻度の高い表情タイプと結びつけることによって果たされる。一般的に言えば、ある時点において、過去全体における対象タイプの注視頻度順によって決定された情動タイプの序列（E1, E2, E3, ...）に対して、共同注意文脈で経験した表情タイプが出現頻度順に「機械的に」対応づけられる。つまり、その文脈の表情タイプは、出現頻度順に E1, E2, E3, ... と名づけられることになる。

我々が第四段階システムに許した自律性は、共同注意以外の文脈において経験された対象を、位置とは無関係に、対象の「タイプ」として捉えることを可能とするだろう。それゆえ、例えば「赤いボール」はいろいろな場所や文脈であらかじめ何度も経験されている。もし共同注意の文脈においても、経験全体におけるのと同じ頻度順で対象タイプが注視されたなら、このシステムは、実質的には次のような推論をうまくやってのけるだろう。「この人が見ているものは、この3つのうちのどれだろう？ この人は<最高頻度>の表情タイプを示している。ということは、いま<最高頻度>の情動タイプを経験しているに違いない。この3つのうち自分が過去の全経験において<最高頻度>の情動タイプを経験したのは赤いボールを見たときだったから、この人も、この共同注意の文脈において、きっと赤いボールを見て<最高頻度>の情動タイプを経験し、それと同一の表情タイプを示しているんだろう。だから、この人は赤いボールを見ているんだ」。

しかし、上のように経験全体と共同注意文脈の両方において注視対象の頻度順が同一の場合でさえも、とくに、ある表情タイプを示しながら、それと同一の情動タイプとペアに設定されていない対象をたまたま他者が実際に意図した場合には、このシステムは「自分の情動タイプの片割れ」を選択し、その人の意図的な注視対象を O3 ではなく O2 だと誤って推論するかもしれない。しかし、このシステムでは、推論の結果が誤っていようとも、それを

「意図ボックス」に送り込み，それに従って視界移動が行なわれる。

ここで，3つほど関連のコメント．一つは，ここでシステムが理解する他者の意図は，最終的には，＜最も高い頻度＞に従った自己の情動タイプからの推論の結果であって，心の理論を用いて推論されるような意図ではないということだ．これは，こうした貧弱なマインド・リーディングのメカニズムしか持たないシステムにとっては，いかんともしがたいところである．しかし，「そこのお塩の瓶，取れますか？」という発話が，食事中に本当に相手の運動能力を確かめようとしたリハビリ医の問いかけであったにもかかわらず，それにうっかり気づかずに塩を手渡そうとした患者，といった例を容易に想像できるように，高頻度解釈によって他者の意図を推論するのはそうバカげたことではない．第二に，このシステムは他者の意図理解に関して誤りを犯すが，我々はそれを好ましいことだと考えている．反射行動には基本的に＜誤り＞を犯す余地がない．少なくとも我々の文脈では，誤りは，システムが複雑化し，自分以外の意図が評価の基準になりうる場合に初めて可能となる．第三に最も重要なことだが，我々は，いわゆる共同注意現象一般においてこの種の推論が必ず生じているとは考えていない．この種の推論は，それに先立つ反射的共同注意と想起的共同注意を前提として，他者の意図理解をなるべくシンプルな道具立てで実現しようとした結果である．さらに言えば，我々の想起的共同注意すらも，見慣れた経験ずみの対象の想起を本質としている点で，共同注意現象一般のパラダイムケースと言うには強すぎる主張だろう．

ところで，この第四段階システムは，実は，他者の意図（注視対象）に関しての自分の推論が正しいかどうかをまったく「気にかけていない」．つまり，推論の結果として自分の注視対象としたものが他者の実際の意図的对象であるかどうかは，このシステムにおいてはまったく問題とされないのだ．それを「気にかける」システムを開発することが，次のプロセス3へと我々を導く．それは，他者に自分の意図を確認してもらい，そのことによって他者と意図を共有しようとする第五段階のシステムである．

5-2. 他者と意図を共有しようとするロボット（プロセス3）

プロセス3: 自らの意図に向けられた他者の意図を回帰的に理解するようになる過程

第四段階システムが他人の意図的な注視対象を誤って推論した場面に戻ろう．このシステムは例えば，たまたま他者のそのときの意図的对象O3を，そのときの表情（情動）タイプからの推論によって，誤ってO2と理解して

しまったのだった。第五段階システムは、この時点で自分の推論が他者の意図に適っているのかどうかを確認しようとする。そして、それが適っていることを確認して初めて、このシステムはその推論対象を自分の「意図ボックス」に送り込む。しかし、それが確認できない場合は、自分の推論をキャンセルし、他者の意図通りの対象を探り当てるまで推論をやり直す(あるいは、1節で述べた「教示的な交互凝視」に相当するものによって、自分の推論対象をその人に承認させる)。

そのために必要となるプロセスは1節で触れた社会的参照であり、それは、乳幼児がある対象を注視した後で親の方を振り返り、その態度によって自らの行動を調整するような行動であった。我々の第五段階システムが、第四段階で行なった推論の正しさを確認するには、推論された対象を他者に表示するメカニズムと、他者が示す少なくとも2つの評価的態度、是認(Y)と否認(N)を見分けるメカニズムが必要となる。これらが実装されれば、第五段階システムは次のような流れで自分の推論が正しいかどうかを他者の態度から知るようになり、それに従って推論をやり直すことができるだろう。

まず、第五段階システムは、他者の視線の補足から注視対象を推論し、それを試行的な対象として注視する。その際、このシステムが複数の候補対象のうちのどれを注視対象として選択したかが分かるように、例えば、その対象にロボットからレーザー光が照射されるなどの表示方法が必要である。これは、乳児が対象を見るだけではなく、指さしたり実際に対象を手を取ったりすることに対応する。相手はそれを見て、それが実際に自分の意図した対象であれば是認の態度(Y)を、そうでなければ否認の態度(N)をロボットに示す。この評価的態度は第四段階で人が示す表情タイプとは異なるものであり、ロボットの試行的対象への注視の後に人がロボットに対して表すものだ。評価的態度の表示方法は、表情タイプと同様に、たとえシステムによる判断の間違いが多かろうとも、システムの今後の応用可能性を拡大する上では、人の表情や身振りに依拠するのが最善であろう。しかし、人間の表情を的確に識別することは、我々の目下の課題とは別の大きな課題である。我々の文脈では、表情や態度は他者(ロボット)にそれとして理解されればよいので、なにも内面を伺わせる曖昧な表現や微妙なニュアンスを伴っている必要はない。容易に識別・解読できるような仕方でコード化され、外部に表示されればよいのだから、極端な話、「笑顔」や「驚き顔」をコード化した数字や、「Y」や「N」と記したボードを首からぶら下げてもいいだろう。

第五段階システムが行う共同注意のバリエーションとして、我々は、先に述べた「教示的な交互凝視」を考えることができる。「交互凝視」は幼児が

親の注視方向を確かめるために親と対象の間に視線を行き来させることであるが、「教示的な交互凝視」は、自らが注意を向ける対象に親の注意が向くまで、ときおり呼びかけるような声を発しながら交互凝視を続けることであつた。そこで、試みに、我々のシステムが教示的な交互凝視をするきっかけとなる場面を作りだしてみよう。それは、我々のロボットが自分の推論の正しさを確認しようとした時、相手側が是認「Y」でも否認「N」でもない第三の態度を表明するような場合だ。この第三の態度「-」を相手側の「意図の曖昧さ」と解釈するのも、あるいは「意図の撤回」と解釈するのも自由である。ロボットは、この評価に接すると、是認「Y」の際と同様に、自分が推論した対象の注視を続行するが、さらにそれだけでなく、相手側に自分の注視対象を（単純な音声などによって）繰り返しアピールし、最終的に相手側に是認「Y」を表明させようとする。このような場面設定はいくらでも工夫できるが、我々にとって大事なことは次のことである。

第五段階システムにおいては少なくとも、自分が推論した対象を自分の注視対象としようとする「意図 1」がロボットの側に存在し、その「意図 1」を是認もしくは否認しようとする「意図 2」が他者の側に存在し、さらにその「意図 2」を理解した上でそれを共有しようとする「意図 3」がロボットの側に存在する。そして場合によってロボットは、他者との意図の共有を、他者の「意図 2」を自分の「意図 1」に同調させることによって達成しようとする（ロボットの「意図 4」）。今回の我々のプロジェクトに関する限り、これが、プロセス 3「自らの意図に向けられた他者の意図を回帰的に理解する」における〈回帰性〉の内容だと我々は考える。

残された課題は、第四および第五段階のシステム（ロボット）を実作し、被験者に把握される「意図性」の程度とその内容を検証することである。

註

¹ モデルの作成においては現象をいかに捨象するかが重要である。例えば石黒浩(大阪大学)は、ロボットの外観を可能なかぎり人に近づけたアンドロイド(ジェミノイド)を制作することによって、人が他者をどのように認識しているのかを議論している。通常は〈人間〉という性質複合体のうち、重要なのは心や思考や感情といった内面的性質群だと見なされるが、石黒は外見という次元だけを残して後はすべてを捨象したようなモデルを制作する。興味深いのは、こういった極端な捨象が行なわれたモデルが、他の性質を残したモデル以上に〈人間性〉をリアルに再現しているように思われることだ。

² 詳細は金野・橋本(2008)を参照のこと。

³ ロボットから取得する画像解析では、人の視線を得ることが技術的に難しいので、常に人の注視点を計測することのできる専用の装置を用いた。

⁴ 視線計測装置はより高い解像度を実現できるが、8ヶ月程度の乳幼児の視線解像度を想定し、あえて粗い設定にした。

文献

- Corkum, V. and C. Moore (1995) “Development of joint visual attention in infants,” in C. Moore and P.J. Dunham eds. *Joint Attention: Its Origins and Role in Development*: Lawrence Erlbaum, pp. 61–84.
- Gergely, G., H. Bekkering, and I. Kiraly (2002) “Rational imitation in preverbal infants,” *Nature*, Vol. 415, No. 6873, p. 755.
- Hood, B.M., J.D. Willen, and J. Driver (1998) “Adult’s Eyes Trigger Shifts of Visual Attention in Human Infants,” *Psychological Science*, Vol. 9, No. 2, pp. 131–134.
- Itakura, S., H. Ishida, T. Kanda, Y. Shimada, H. Ishiguro, and K. Lee (2008) “How to Build an Intentional Android: Infants’ Imitation of a Robot’s Goal-Directed Actions,” *Infancy*, Vol. 13, No. 5, pp. 519–532.
- 金野武司・橋本敬 (2008) 「乳幼児の視線:交互凝視行動の計算論的研究」, 『認知科学』, 第 15 巻, 第 2 号, 233–250 頁.
- Nagai, Y., M. Asada, and K. Hosoda (2006) “Learning for joint attention helped by functional development,” *Advanced Robotics*, Vol. 20, No. 10, pp. 1165–1181.
- Tomasello, M. (1995) “Joint attention as social cognition,” in C. Moore and P.J. Dunham eds. *Joint Attention: Its Origins and Role in Development*: Lawrence Erlbaum, pp. 103–130.

(北陸先端科学技術大学院大学, 金沢大学)

なお、本稿は科学研究費課題:「意図的主体性のロボットの構築に向けて」(No.23320002, 柴田正良, H23~26), 「視覚的注意の共有化メカニズムの解明, H21~23」(No. 21700285, 金野武司), および財団法人日産科学振興財団学術研究助成課題:「数理モデル構築による意図的コミュニケーションメカニズムの解明」(金野武司, H21)の研究成果の一部を含んでいる。