

状況づけられたエージェントの推論活動 -アブダクションと常識推論をめぐって-

森 悠貴

Abstract

The main theme of this note is to explain why we can behave in goal directed and reactive way, and what kind of features of us make it possible, by referring to recent research results of philosophy, cognitive science and artificial intelligence. Besides that, we must take into consideration that our behavior is always influenced and restricted by our own body and the environment surrounding it. Specifically, this note deals with abduction and common-sense reasoning as important reasoning we are using on a daily basis. Because these nonmonotonic reasoning allows us to adjust the amount of information and revise information.

(1)研究テーマ

我々人間は合理的な行為を遂行することのできる行為主体・エージェント (agent) である。すなわち我々は、何か一定の目的を立て（その場合、我々はその目的を達成することが、自分にとって益になると普通信じているが）、その目的のためにはどのような手段を講じるのが適切であるかを考慮し（その際、多様な情報を収集し、それらに基づいて論理的帰結を導いたり蓋然的な推測を立てたりして）、実現可能な諸手段の内から、最も適當と思われるものを選択し、それを実行している。このような我々の合理的行為は、一方で、人間に固有の特性と考えられ、伝統的な認識論の文脈においては、この営みの根底にある知性 (intellectus)・理性 (ratio) といった（我々人間が本来的に備えているような）能力がどのようなものであるか、ということが哲学的な問い合わせであった。他方で、現代では、人工知能の発展に伴って、上記のような我々の活動のあり方をシミュレートし、より効率的な体系を作り上げるという課題が現実的なものとなってきた。

本稿では、以上のような 特徴的な活動のあり方を、目的指向的 (goal directed) で、かつ環境対応的 (reactive) という二つの特徴付けから捉えることにし、この二つの特徴を満たした活動を行いうるシステムないし行為主体・エージェントを、GDR 系[注 1]と呼ぶことにする。そして、我々の GDR 系としての営みがどのようなアーキテクチャ（設計原理）によって実現

しているのかを、伝統哲学的な文脈からではなく、人工知能や認知科学の近年の動向を踏まえて掘り下げることが、本研究のテーマである。

実際に具現している GDR 系のなかで、最も典型的なものは人間であると言つてよいだろう。そして、人工知能分野の課題とは、我々人間の合理的な、GDR 系としての振る舞いの再現であると位置づけることができるだろう。GDR 系を再現するにあたって、初期の人工知能分野において採用された設計思想は、まずもって人工知能に対して、達成すべき課題や周囲の環境についての可能な限り完全な情報を与えることで、状況を（一方的に）制御できるシステムをつくる、というものであった。このことから、当時の研究者たちの間で、行為に先立って完全な状況把握を行うことが合理的振る舞いの条件だと思われていたこと、また、典型的 GDR 系である我々人間も同様に、完全な状況把握から合理的な行為を導いていると考えられていたことが窺える。

例えば、迷路を抜けるという課題を果たすための人為的な GDR 系（すなわち人工知能）を設計する場合には、迷路の地図を端的に与えてしまう、あるいは、こうした完全な地図を作成する手順を組み込んでおくといった手法が、当然要求され、また最適なものであると考えられてきた。もちろん以上の述べ方は過度に単純化したものであるが、このように GDR 系として振る舞うためには、（伝統哲学の認識論が我々の合理的な活動の根幹に理性・知性といった能力があることを想定していたように）行為主体があらかじめ詳細で俯瞰的な情報や正確な学問的知識を有していることが不可欠あるいは望ましい条件だと考える立場をここでは、GDR 系についての理想主義的見方と呼んでおこう。

ところが興味深いことに、近年における認知科学・進化心理学・進化言語学などの発展、さらには人工知能研究における知識表現論の進展などによつて、GDR 系を理想主義的観点に基づいて実現することには著しい困難があるということが明らかになってきた。（その挫折のもっとも有名なものがフレーム問題であると筆者は考える）

これを受けて、現代の人工知能分野や認知科学の動向に基づいて GDR 系を捉えると、以下の重要な点が見えてくる。

まず、【1】我々 GDR 系が行う情報の入手・処理は非常に局所的・短期的で身体依存的・環境依存的である。この特徴は人工知能開発とその周辺の諸分野で強く意識され、人間的行為主体の捉えなおしの中で重視されるようになったため、状況づけられた認知（Situated Cognition）という標語で表現され、広く用いられている。

次に、【2】我々 GDR 系が遂行する行為は、われわれと我々を取り巻く環境

との相互作用（interaction）と捉えるべきである。というのも、伝統的な行為論の文脈において、行為とは、行為主体の意図が帰属された運動であり、行為主体から対象への一方的なものである（interactionではなくaction）、と特徴づけられていた。しかし、【1】のような特徴を持つGDR系が営む行為をより精確に分析するためには、行為主体の意図を出発点に行為を考えるのではなく、環境との相互作用（interaction）を出発点にして分析をすべきである。

そして、【3】状況依存的な我々GDR系が、環境との相互作用によって合理的に振る舞うことができるのは、その背後に（無意識に活用され、その重要性が見出されにくいような）洗練された情報処理メカニズムを持っているためである。我々は先に述べたように環境依存的ではあるが、それは環境に一方的に支配されているのではなく、その中で能動的に対応策を練るというリアクションをとりながら様々な創造的活動を行っている。例えば、他者との情報交換、目標達成に成功した他者の模倣、典型例からの予測、行為に失敗した後の分析・学習などは、環境の中に組み込まれ限られた情報しか得ることのできない我々が、GDR系として振る舞うために適応（adaptation）を重ねた結果会得し、体系化した行為であると言えるだろう。

そしてこのような環境へのリアクション・適応こそが、人間がGDR系として振る舞う（以下ではGDR的に振る舞う、等と略記する）ために本質的に不可欠なメカニズムであり、その特性について、周辺分野（認知科学や人工知能等）の研究成果を参照しながら探求することが本研究の目標である。

（2）研究の背景・先行研究

1956年のダートマス会議にて、コンピュータシステム上で人間の知的な振る舞い（合目的行為遂行）の再現を目標とする研究領域が人工知能（artificial intelligence）と名付けられた。前節で提示した研究のテーマにおいて理想主義的見方と呼んだ考え方は初期の人工知能研究（J.McCarthyやP.Hayesなど）に見ることができる。その一方で、Situated Cognitionの観点に立つ研究としては、M.Tomaselloによる人間同士のコミュニケーションの分析や、R.Langackerの像図式（Image Schema）などの、進化言語学や認知言語学の現代的分析を念頭に置いている。そして哲学上の先行研究としては、P.Gärdenforsによる人工知能の哲学的基礎付けの試みやConceptual Space論、そしてこの始原であると考えられるWittgenstein後期の言語ゲームやプロトタイプ理論が挙げられる。

(3)筆者の主張

先ほど紹介した近年の認知科学の動向では、我々の感覚知覚や情動といった情報処理の手法に注目が集まっている。しかし、本論文はこの潮流に倣うのではなく GDR 系である私たちが行う推論について掘り下げる。推論は(研究背景で紹介したところの)理想化された理性・知性によって可能になる(いわば感覚知覚や情動の対極にある)知的営みである、という認識が一般的であった。しかし、我々が状況づけられている行為主体でありながら、GDR 的な振る舞いが可能であるということをより正確に分析するには、推論に関しても理想的な見方を捨てるべきであると筆者は考える。このことを自覚できるようになったのは、近年の周辺学問(特に人工知能分野)の進展によるところが大きいだろう。と言うのも、我々が日常的に(しかも無意識的に)行っている、一見不完全に見える情報処理は、それだけで直接に観察すると単なる欠陥や不完全さのように見えてしまい、実はその根底に、我々自身が環境に対する積極的リアクションとして講じる方策・戦略といったものが隠されていることを見落としてしまいがちだからである。

そこで本論は【A】アブダクション・仮説形成(Abduction)、【B】フレーム問題(Frame Problem)と、それに密接に関係する常識推論(Commonsense Reasoning)について詳しく検討する。これらは、我々が日常的に用いる典型的な推論・情報処理であり、なおかつ、非妥当な推論が我々の GDR 的振る舞いにおいて(おそらく不可欠で)重要な役割を果たしていることを示しているにも関わらず、今までそのメカニズムや重要性が明確にされていなかったものであると言えるだろう。アブダクションや常識推論が我々の GDR 的振る舞いに対しどのような貢献をしているか、また、その働きから見てとれる哲学的興味はどのようなものか、以下で具体的に見ていくことにする。

3-A アブダクション

アブダクションとは大まかに言えば、命題 P によって表される事実が、行為主体に知られた際に、その事実が成立している理由を最も適切に説明すると思われる仮説 H を案出したり探し出したりして、次の行為を行うための前提として暫定的に採用すること、言い換えれば、観測した事実 P が成り立っていることをうまく説明できるいくつかの一般化の候補の中から、もっとも適切であると思われるものを選び、その一般化が背後で成り立っているという仮説を採用することである。

我々が日常生活においてアブダクションを行う典型事例を見てみよう。

【例 1】A 氏には、市役所に出かけるという用事があり、ついでにその途

中にあるレストランで食事をとろうと考えている。市役所に行く途中でレストランに寄るか、市役所から帰る途中でレストランに寄るかの選択肢があるが、A 氏はレストランでの待ち時間ができるだけ少なく済ませたいと思っている。そんななか A 氏が市役所に行く途中でレストランの様子を見ると、当のレストランの前に行列がないことに気付く。実はそのレストランにはいつも順番待ちの列ができていて、食事ができるまで待たされるのが常なのだが、その日は様子が違うようだった。そこで A 氏は「店の前に順番待ちの行列がないのは、おそらく今レストランがちょうど空いているためである」「従って今レストランに寄れば、待ち時間をかけずに食事ができる」と考え、市役所に行く前にレストランに寄ることにした。

つまり、A 氏はいつも行列ができるレストランの前に行列がないことを観測し、その事実を説明する一般化のうち「店が空いているならば、店の外に行列はない」という一般化が適していると考え、仮説として採用した。そして、その一般化が現状に適用できるということを前提にして、往路でレストランに寄るという行為を選択した。もし A 氏が採用した仮説が正しかった場合、市役所へ行くついでに、時間を無駄にせず効率的に食事をとることができただろう。しかし、仮説が間違いである場合も当然ありうる（実はレストランが休みで誰も並んでいないだけであった場合など）。

従来のアブダクション分析においては、我々がどのようにして仮説を立てているのか、そして複数たてた仮説の中からもっとも適切なものをどのように判別しているのか、といった問い合わせが議論の中心となっていた。しかし、アブダクションに関して筆者がより注目すべきであると考えたのは、(i)状況づけられた行為主体である我々は、必ずしも妥当な推論に基づき行動しているわけではない、という点と、(ii)我々は有限である手持ちの情報をアブダクションによって拡張し、一定の仮説が立てられれば、(仮にその仮説の正確性が保障されていなくても、またさらに、未決定の事項が保留にされたままであっても) 行為の実行に踏み切っている、という点である。先ほど指摘したように、状況づけられた行為主体である我々は身体的制約、環境的制約によって、合理的な振る舞いに必要な情報をすべて正確に直接入手することはそもそも不可能である。そのような中で、GDR的に振る舞うためには、不確実・未決定な事項を保留にしたままで行為の遂行へ踏み切ることは避けては通れない。

そこで我々は、身体によって観測した有限で断片的な情報を、仮説を設定することによって拡張し、次の行為のモティベートに利用している。さらに、もし仮説が間違いだった場合でも、観測に基づいて仮説の棄却をすることで、

(仮説を立てなかった場合よりも)手持ちの情報を増やすことが可能である。

しかし、アブダクションは、我々の GDR 的振る舞いにとってもっとポジティブな意味があると筆者は考える。それは、観察した状況についての情報に基づいて、行動を決定するという、我々の状況適応に貢献している点である。我々がある目標を達成しようとするとき、その成功が見込める手段はおそらく複数あることが常である。しかしその複数ある手段のうち、行為主体がどのような状況に置かれているかによって達成の容易さや実現度が変動する。そんな中、どの手段を選ぶべきか（目的実現のための手段にとって、状況がどのように影響を与えていたか）はすぐ明らかになる事ではない。そこで我々はアブダクションを通して得られた仮説に基づいて、どの手段が最も目的達成に適切であるか模索しているのではないだろうか。

つまり、アブダクションとは、我々の目的達成のためにどのような手段を選択すべきかを決定する基準・条件についての仮説を、観察した状況に基づいて提供する推論であり、状況づけられた行為主体である我々にとって、GDR 的に振る舞うために不可欠なものである。

以上が、本論文で強調したい、従来指摘されてこなかったアブダクションの重要な役割である。

3-B フレーム問題と常識推論

フレーム問題とは、人工知能分野から提示され、哲学者によって盛んに議論された、我々の合目的な振る舞いを人工知能で再現する際の困難についての議論である。しかし、フレーム問題がどのような問題であるか、どのような点で哲学的に重要であるか、ということについては精確な考察がなされてきたとは言えず、この点は M.Kamermans と T.Schmitz によっても指摘されている。詳細は参考文献に任せるが、言ってしまえばフレーム問題とは、妥当な推論だけを行う人工システムに、我々人間が GDR 的に振る舞うために行う推論（つまり、非妥当な論理的導出を含む推論）を再現させようとしたところ、追加しなければならない公理が大量に要求され計算爆発に陥ってしまう、という問題である。そしてこの問題を巡って、(a)人間にはフレーム問題は起こらない。従って人間には人工システムによって還元不可能な知的能力が備わっている、といった議論や(b)言語を用いた、つまりシンボル操作による表象主義的なアプローチでは人間の振る舞いの再現は不可能であり、コネクショニズム的アプローチをすべきである、といった議論が哲学者の間で行われていた。

しかし実際には、その後の人工知能研究の進展を参照すると、ある種の情

報の度外視の原理を組み込んだ論理的言語（Circumscription や Default 論理など）を採用することで、計算爆発は回避することができ、さらにこうした特別な種類の論理的原理とは、実は、状況づけられた GDR 系である我々が、環境へのリアクションとして日常的に用いている論理（いわゆる常識推論）そのものであることが明らかになってきた。従って、従来哲学者が行ってきたフレーム問題解釈は的外れなものであり、むしろ、フレーム問題が示唆している哲学的に興味深い問いとは、①我々が日常的に行う常識推論とはどのようなものであるか、そして、②なぜ常識推論が我々の GDR 的振る舞いにとって役に立つか、というものであるだろう。そして本研究の文脈からみると、デフォルト推論（Default reasoning）、閉世界仮説（Closed world assumption）などの常識推論は、（我々が行為を行う際に利用できる仮説を創出できるという点でアブダクションが有用であったのに対し）我々が行為をする際に考慮しなければならない情報を減らすことができる、という点で我々の行為に貢献していると考えられる。

我々は実際、日常的な推論において(I)（重要な影響を持たない事実に関しては）情報を得るまでは無いものと見なす、(II)報告を聞いてから対処すればいいこと（もしくは当面の目的遂行に関して関係ない話）は考慮しない、(III)もし新しく考慮すべきことが発生したという証拠を入手したら、その時に古い情報を棄却し新しい情報を採用する、といった特徴的な情報処理を行っている。このような情報処理が可能であるのは、我々が常識推論を駆使しているからであると筆者は考えている。つまり、常識推論には、（直接証拠を得ていない情報に対して、暗黙に、標準的な状況が成り立っていることを想定したり、考慮すべき状況は成り立っていないことを想定したりすることで）、目的達成のために扱わなければならない情報を減らす働きがある。そしてこれらの想定が正しくなかった場合でも、新たに情報が明示されれば、それに応じた手持ちの情報の改訂を可能にする。この働きこそが、我々の GDR 的振る舞いに対する常識推論の貢献である。

常識推論は我々の行為の遂行にとって不可欠なものであり、GDR 的振る舞いを人工知能で再現するならば避けて通ることはできないだろう。このことを明らかにしているという点で、フレーム問題は依然として興味ある問い合わせ我々に投げかけている主題である。

(4)今後の展望

かつての哲学史において、我々が行う妥当でない推論の重要性は指摘されてきたが、その重要性の所在はあまり明らかにされてこなかった。しかし本

研究を通して、状況づけられた主体が GDR 的に振る舞うためには、非妥当な推論を駆使することで、(情報を求めるときは不確かな仮説でも利用し、確かな情報であっても明示的に必要になるまで無視するといった) 考慮する情報量を調節する、もしくは状況に応じた手持ちの情報の改訂を行う、ということが情報処理上必須であることが明らかになった。

人工知能上の再現といった応用に耐えるような、我々の合理的行為者としての在り方についての分析をするためには、本研究が明らかにしたことが重要になってくると筆者は見込んでいる。さらに、人工知能分野や認知科学の成果と哲学的な分析のそれぞれの立場を双方向的に考慮しながら、我々の情報処理の傾向や状況適応の仕組みについて探究することは興味深い課題であると言えるだろう。

[注 1] “Goal directed” は行為論で、“Reactive” はプログラミング言語の意味論で使われている一般的の意味をそれぞれ念頭においている。

(5) 参考文献

- Dennett, D, 1984, *Cognitive Wheels: The Frame Problem of AI*. In *The Philosophy of Artificial Intelligence* edited by Boden, A, M, pp. 147-170.: Oxford U.P.
- Gärdenfors, P, 2004, *Conceptual space*: MITpress.
- Gärdenfors, P, 2014, *Geometry of Meaning*: MITpress.
- Harmelen, F, V; Lifschitz, V; Porter, B, Eds. 2008, *Handbook of Knowledge Representation*: Elsevier
- Kamermans, M; Schmits, T, 2004, *The History of The Frame Problem*
- Langacker, R, 1987, *Foundations of Cognitive Grammar* vol. I & II : Stanford U.P.
- Langacker, R, 2008, *Cognitive grammar — A basic introduction*: Oxford U.P.
- Wittgenstein, translated by G.E.M. Anscombe L, 1963 *Philosophical Investigations*: Basil Blackwell, Oxford
- McCarthy, J; Hayes, P, 1969,
Some philosophical problems from the standpoint of artificial intelligence
- Reiter, R, 1980, *A Logic for default reasoning*. *Artificial Intelligence* 13: pp.81-132.
- Tomasello, M, 2010, *Origins of Human Communication*: MITpress

(首都大学東京)