

価値が導く論理的思考の計算モデル化にむけて Toward the computational modeling of value guided logical inference

大森 隆司^{†‡}, 宮田 真宏[‡]
Takashi Omori, Masahiro Miyata

[†] 日本大学, [‡] 玉川大学
[†] Nihon University, [‡] Tamagawa University
omori@lab.tamagawa.ac.jp, 38tam.mail@gmail.com

概要

人の知的能力の一つの特徴が論理的推論である。これは他の動物にはなかなか見られない能力であり、そこにヒト脳の特殊性が表れている。一方で、人の脳は進化的な連続性に基づき他の動物の脳と大きくは変わらない。ヒト脳の何が他の動物と異なって論理的な思考を可能としているのか？その解明は認知科学に限らず科学の大きな問いであると同時に、社会的にもインパクトが大きいであろう。しかしこの問いに対する仮説はいまだ少ない。そこで本稿では、この現象に対する一つの仮説を提示して、その仮説の可能性と検証の方法について議論する。

キーワード: 論理推論, 連想記憶, 認知アーキテクチャ

1. はじめに

1.1. ヒト脳における論理的推論の特殊性

認知科学会の大会にはこれまで、大局的な認知モデルを論じてくださる名物会員がおられたが、残念ながら最近お見受けしない。目の前の課題に堅実に取り組む会員に大局的な視点や大胆な仮説を提示してくれる、未完成のあやしい理論もまた学会には必要と考える。そこで本発表では、著者が概念レベルで思い描く脳神経系で可能と思える論理的推論の認知アーキテクチャについて紹介し、脳認知過程の大局的なシステムデザインについての議論の材料を提供したい。なお、本稿は本年9月に発行予定の認知科学の論文特集「戸田正直『心理学の将来』から半世紀」に投稿した論文[1]の補遺的な位置付けである。本稿の前半は論文を整理しつつトレースして、後半は追加での説明をする。

論理的推論はヒト特有の知的能力とも言われており、論理学をはじめとする推論に関する学問はその典型的な頭脳であろう。それに伴い、20世紀半ばからのAI研究もまた論理的推論から始まった。しかし記号計算主義による知能観はニューラルネットの登場により大きく変わり、階層ネットワークによる入出力関係の近似が現代の知能観を形成している。

一方で、認知科学分野では1970年頃にPaivioにより二重符号化仮説が提唱され、それが現在までの認知ア

ーキテクチャ研究に引き継がれてきている[2]。その仮説とは、ヒトの脳には論理的推論と直感的推論という2つの推論過程が存在し、我々はその2つの過程を動的に組み合わせて思考している、というものである。そのメカニズムの理解はヒトという種が他の動物と異なる能力を持つ理由の理解につながると考えられる。

1.2. 論理的推論の脳メカニズムの仮説

人工知能(AI)の理論では、推論による意思決定の基本は予測と価値評価の組み合わせであるとされている。すなわち、現時点での状態に、可能な操作を加えた結果についての価値評価を行い、価値が最も高くなる操作を採用して実行するという方式である。論理的推論の場合については多くの研究があり、反復や相手の参加がある場合には複雑な計算処理となる[3]。

これを大量のデータによる入出力関係の学習で行うのがニューラルネットであり、予測と評価の反復をしない直観的推論による意思決定を実現している。AIでの論理的推論と直感的推論は計算原理が大きく異なるものであり、人の脳のように二つの過程を動的に組み合わせることは困難である。

一方で行動経済学では、人の意思決定はAIの論理的推論に似て価値の評価に基づくとされている。したがって、意思決定において強力な推論は可能な選択肢の効果を予測・評価するという意味で重要な過程である。ところが同時に、感情に駆動された意思決定は多くの場合に適切であることも指摘されている[4]。とするなら、意思決定における推論と価値と感情はどのような関係にあるのだろうか。

この問題に対して本研究は、意思決定に関わる脳プロセスについて以下の仮説を置き、そこから生まれる一見して論理的に見える記憶想起の系列からなる、推論過程のモデルを説明する。

- 感情とは価値計算システムの表出系である。

動物の進化において情動過程は基本的に同一であり、そこには生存に関わる必要性があるはずであ。本研

究ではそれは価値評価であると想定する[5].

- 脳を大規模な連想記憶系とみなし、推論とはそこに記録された記憶の間の連想的想起の連鎖と考える[6][7]. その想起のモードにより、直観的推論と論理的推論の二つの振る舞いが切り替えられる. 直観的推論は、入力からの並列的想起で実現され、論理的推論は、想起された記憶の収束の連鎖である.
 - 感情は、想起された記憶の価値評価を行ない、その想起収束過程にバイアスを与えて、記憶を探索する想起の連鎖を価値の高い方向に誘導する
 - 意識状態とは、一つの記憶が脳内全体に広がった状態として(Global Workspace Theory, GWT)[8], 本モデルの連想記憶の想起が収束した状態に対応する.
- 本モデルの特徴は2つある. 一つは連想記憶の動作モードによる推論過程の切り替え, もう一つは想起した記憶の価値評価による想起過程への介入である.

2. 意思決定と推論の関係

2.1. 脳内知識の二重符号化モデル

認知アーキテクチャとはヒトの脳の認知処理過程をモデル化しようという研究であり, 1970年代からAI研究と並行して始まり, 現在も続いている[9][10][11]. ここでの知識表現の基本戦略は二重符号化である. 脳内には, 連続的・分散的なパターンの情報と離散的なシンボリック情報の組み合わせた知識表現があり, 推論はそれらの間を動的に切り替えながら行われるという考え方である[2][12]. その表れが, 論理的推論と直観的推論であり(表1), 多くの研究がなされてきている[13].

表 1 直観的推論と論理的推論の二重符号化

直観的推論	論理的推論
作業記憶は使わず	作業記憶が必要
無意識的, 自律的	意識的, メンタルシミュレーション
速い	遅い
バイアスに影響されやすい	規範的, 公平
文脈依存	抽象的
確率的, 分散的	論理的, シンボリック
暗黙知(経験的確率)を利用	明示的な知識を利用
推論が浅い	深い推論が可能
進化的に古い	進化的に新しい

この2つの推論の脳過程について, 直観的推論はニューラルネットによる実現が可能が高いように感じられる. これまでの多くの脳神経系の研究は脳の階層

的認識機構のメカニズムを明らかにしてきており, その特性は行動研究とも矛盾が少ない[13].

しかし論理的推論の脳過程については不明な部分が多く, 前頭葉が深く関わることは間違いがないが, それ以上の詳細は明らかとは言えない. 脳研究において論理的推論を行うのはヒトのみであり, ヒトに対する研究の方法は限られている. ここからは, 脳科学の知見に幅広い分野からの知見を加えて, ありうるモデルを作っていかなければならない.

ここでの課題は意識である. 例えば脳神経回路による視覚認識は麻酔下でもある程度は働くことから, 直観的推論には意識は不要と考えられるが, 論理的推論には表1に挙げたように意識が伴う. 脳の神経回路がマクロな現象として論理的推論を生み出すメカニズムには, 何らかの形で意識のような現象が伴う意思決定の過程についての説明が求められる.

2.2. 連想記憶の相互想起と自己想起

神経回路による連想記憶は1972年に中野により提案され[6][7], 1990年前後に活発に研究された記憶モデルで, 記憶したパターンの一部からのパターン全体の想起(自己想起)や, 一つの記憶Aから別の記憶Bへの一撃での連想(相互想起)などの機能を持つ.

連想記憶モデルで我々が注目するのが, 自己想起の過程で神経回路の状態に対応して定義されるエネルギーである. 自己想起型の連想の反復計算過程では, 神経細胞群の状態はエネルギー最小化の過程として変化して, 記憶の安定想起状態へゆっくりと収束する. その際, 複数の記憶が重なった初期状態からでもそのうち一つの記憶が選択・想起される. 詳細は参考文献[14][15]をご覧いただきたい.

結果として, 相互想起の連想記憶は複数の記憶への分岐も含めて短時間で入力から関連する記憶を想起する, すなわち素早い推論と記憶探索の機能を実現し, 自己想起は複数の記憶が混在した状態から時間をかけて特定の記憶を選択的に想起する機能を実現する.

ここで, 自己想起型の連想記憶と相互想起型の連想記憶を組み合わせる動的にスイッチするなら, 表1に挙げた速い直観的推論と遅い論理的推論と類似の動作の動的組み合わせが可能となる.

2.3. 価値と感情と意思決定

通常の連想記憶では想起される記憶項目は事前に決

まっており、複数の記憶の競合場面でも分岐の方向を変えることはできない。そこで我々は、記憶項目の価値評価による記憶想起の分岐の制御を行った。その価値評価のメカニズムとして我々は感情を想定する[5]。

行動経済学では意思決定においてそれぞれの選択肢の効用、すなわち自己にとっての価値を評価し、その最大の選択肢を選ぶとする。そこに合理性が不足することが人の価値評価の特徴であるのであるが、ともかく選択肢の価値評価が意思決定を支配するというのが現在の認識である[4]。では、この評価のシステムとは何であろうか。それがここでの論点である。

これに対して本研究では「感情」を価値計算システムの顕れと想定する。意思決定の場合には可能な選択肢を実施した場合の結果の予測に対してであろうが、その予測に対しての価値評価が感情¹として表出されると想定する[16]。価値評価の計算自体は我々の脳内にて大部分は無意識のうちにに行われていると考えられ、その主要なものが感情として表出・認識されと考えると、感情に駆動された意思決定の多くが説明できるように思われる。

従来、感情は理性と反するものとされてきたが、人類がまだ狩猟生活をしていた時代には感情こそが生存に適していたものが、文明の発達とともに状況に適応できなくなって非合理的な判断が目立つようになったものと考えられる。しかし多くの場面ではいまだに感情の判断は適切であると主張されている[4]。また、動物の進化の過程で爬虫類以降の情動システムは基本的に不変であるように思われる。とすると、情動システムが動物の生存に必要なという理由があるはずである。それが価値評価である、というのがここでの仮説である。

3. 論理的推論の脳モデルと意識

3.1. 連想記憶と価値認識による推論モデル

感情は価値評価システムの表出とすることは受け入れられるとして、価値評価を連想記憶による推論に導入して想起の過程を修飾するメカニズムが必要である。

この問題に対して宮田は、自己想起型と相互想起型を組み合わせた連想記憶に価値評価のシステムを付加した推論モデルを提案した(図1)[17]。直観的推論で

は、現在の感覚入力から可能な連想を次々で行う。その際、一般に連想は一对多であり、Tree 探索的な分岐が発生するが、本モデルでは条件付確率を連想過程に取り入れて、並列的な記憶想起とその価値評価による高速な記憶探索を実現した。また自己想起による論理的推論では複数の記憶パターンが混合した状態から価値の大きい記憶を選択的に想起する価値駆動の記憶選択の機能を実現した。この詳細はやや専門的なので、参考文献[18][19]を参照されたい。

ここでの鍵は想起した記憶の素早い価値評価と連想過程への介入である。これが実現する論理的推論により、従来の AI 的な Tree 探索で起きていた無駄な探索を大幅に削減して価値のある方向にのみ推論する、深さ優先探索が実現できている(図2)。

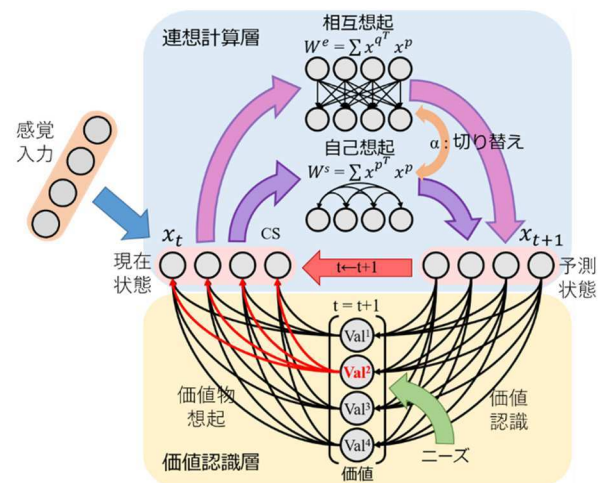


図1. 直観的推論と論理的推論を実現する連想記憶モデル ([17]より引用)

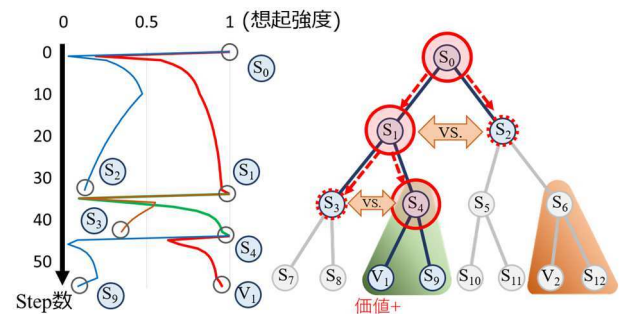


図2. 自己想起型の連想による Tree 探索的な記憶の探索過程 ([17]より引用)

¹ 本稿では感情という言葉を用いるが、心理学などでは情動という言葉が使われることが多い。情動は動物一般の価値評価システムと考えたときに、ヒト

の場合にはその価値評価に至るまでに予測などの知的処理が加わっているように感じる。そこで本稿ではヒトの価値評価にはそれを含めて感情と呼ぶ。

3.2. Global Workspace Theory と意識

先に述べた通り、現象としての論理的推論には意識が伴う。ここで言う意識は一般に言われる自己意識のような深い意味ではなく、意識のあるときにのみ論理的な推論が行われるという現象を言う。それにしても、連想記憶の自己想起過程を論理的推論と言うなら、それが意識様の現象を伴う理由の説明が必要である。

Baars により提案された Global Workspace Theory (GWT)は[8]、脳を大きな神経回路とみなし、外部からの刺激に対する応答から、脳の無意識状態と意識状態の違いを説明した。その説明によると、外部刺激が弱いときには脳はそれを認識せず、適度に弱いときにははかすかに反応するがその反応は局所的で脳に広く広がることはなく無意識的な反応となる。そして刺激がある程度強くなると、その刺激は認識されて反応が脳全体に広がる。この刺激に対する反応が脳全体に広がった状態を GWT では意識に登った状態としている(図3)。さらに GWT は、神経科学の立場からもサポートがある。意識に関わる脳過程は 2000 年代に研究が多くなされ、Dehaene は神経科学の実験と理論の立場から研究を進めてきている[20][21]。

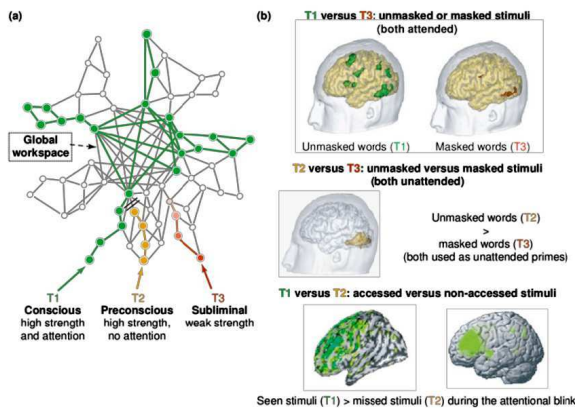


図3 Dehaene による GWT の説明 ([20]より引用)

この入力刺激が意識に登った状態とは、脳を一つの連想記憶とみなした場合には、特定の記憶が安定して想起できた、すなわち先述の推論モデルでいうところの論理的推論の安定想起と同等の状態と考えることができる。それに対して直観的推論は感覚入力からのすばやい連想であり、脳全体への認識の伝搬は不要な無意識的推論と言える。このように、宮田による推論モデルは GWT とのマッチングがよく、また表1に挙げた人の推論過程の現象との対応も比較的良好、すくなく

とも本稿で検討している粒度においては、人の推論過程のモデルであると考えることが可能と考える。

4. 論理的推論モデルから導かれる仮説

4.1. 意識状態の効用としての論理推論

ここまで、(1)推論の二重符号化モデル、(2)連想記憶モデルによる相互想起と自己想起の性質、(3)感情による価値判断と意思決定の関係、(4)宮田モデルによる価値に駆動される直観的推論と論理的推論、および(5)意識の理論としての GWT についてそれぞれ述べてきた。そしてこれらの組み合わせとして、ヒト脳が論理的推論を実現するメカニズムについて提案されたのが本推論モデル[17]である。

さて、このモデルから何がいえようだろうか？ヒト以外にもあると推測される直観的推論の特徴は、入力からの反射的な予測と評価によるすばやい意思決定である。その予測は広く・速く・浅く・自動的である。このような推論は生存をかけて瞬間的な意思決定を求められる厳しい環境での生存には必須であろう。

しかし浅い推論は遠い未来の予測には不向きである。我々が感じる知的能力の一つの指標に推論の深さがあるだろう。すなわち、ただちには価値評価が困難な場面で、深い予測により遠い未来の状態の価値を評価できることは、意思決定の精度を向上させることでその個体の生存確率を向上させると期待できる。そして、本モデルの論理的推論が可能としたものこそ、予測が分岐して複数の記憶が混合した場面でより確からしい分岐の一方の記憶を選んで選択想起して次の推論を始めるという、深い推論の過程である。

さらに、直観的推論と論理的推論の違いは、相互想起と自己想起という連想記憶の同じ計算メカニズムの利用法の違いでしかなく、両者の協調的相互作用は極めて自然に実現される[18][22]。直観的推論から論理的推論の創発は必然的な進化と言えよう。ただ、その脳神経的なメカニズムとしてヒト脳が他の動物の脳とどのように違うのかは不明である。

そしてその先に、意識様の脳状態としての GWT がある。GWT がいう意識状態とは、本モデルの論理的推論の記憶想起が収束した状態に相当する。この状態は論理的推論が深い推論を可能とする混合記憶からの選択想起の結果として得られるものである。先述のとおりこの状態が脳内に表れることは個体の生存に有利に

働くと考えられることから、これが意識のような機能が持つ進化的な効用となった、というのは言い過ぎであろうか。しかしこの考え方は、我々が感じるハードプロブレムとしての意識を説明するものではない。

4.2. 記憶想起の安定は作業記憶を可能に

論理的推論のもう一つの効果として作業記憶の実現の可能性がある[23]。作業記憶とは脳内の記憶想起状態の一時的な保持と考えられるが、直観的推論は記憶想起の過渡期の状態であり、脳波などの多くの揺らぎ要因がある脳内での保持は困難であろうと推測される。それに対して論理的推論は記憶想起の離散的かつ安定した収束状態が記憶の表現であり、それならば脳内で進行している他の処理とは別に一時的に保持できる可能性がある。現在、作業記憶の脳部位としては前頭葉・頭頂葉・前帯状皮質・大脳基底核や海馬など複数の部位が挙げられて議論が続いているが[24]、その保持の安定性もまた検討されるべき項目であろう。

作業記憶の実現は、複数の推論の同時遂行を可能とする。いま進行中の推論を停めて一時的に保存し、別の仮説に基づく推論を行って保存したものと比較するという処理は、脳の過程としても妥当である[25]。結果、作業記憶の利用は脳での複雑な推論、ひいては複雑な問題に対する意思決定を可能とするであろう。

4.3. 他者の価値を含む意思決定＝社会性

現生人類であるサピエンスが獲得したとされる特徴的な能力に社会性がある。社会性とは、インタラクション場面において自己の価値だけでなく他者の価値も含めて評価して自己の意思を決定することと言えよう。これは自己と他者の双方の予測と価値評価が必要な複雑な意思決定である。

我々がインタラクションを求めるのは、それに何らかの価値があるからであろう。質問をする行為は自己が知識を得るという価値がある。一方で他者に何かを教える行為は、自己の価値よりは他者の価値につながる。会話で相手を喜ばすことは、自己の行動決定の段階で他者の感じる価値を評価しているであろう。

そのような行為を可能とするには、自己の意思決定が他者の価値評価に及ぼす影響までも含む推定が必要である。社会的存在としてのヒトは、それを可能とする他者についての知識、すなわち他者モデルを持って意思決定に用いることで、友達や仲間などのグループを

作り維持することができているのであろう。同じ原理は敵対的な社会的関係にも適用できる。他者の嫌がること、他者の価値評価が低下する意思決定を行うことは、対戦ゲームなどで我々は日常的に行っている。

このような他者の価値評価を自己の価値評価に含んで意思決定を行う脳過程については、理化学研究所の中原裕之氏のチームの研究がある[26]。それによると、脳内で他者価値を意思決定に結びつける過程は3段階あり、第一段階は他者の価値を、第二段階はその価値の選択肢に対する影響を、そして最後は最終的な意思決定の価値をそれぞれ異なる脳部位で符号化している。またその符号化の強度は被験者の向社会性と関係している。社会性という一見複雑な意思決定もこのように脳過程として、さらに計算過程として表現可能であることが示されつつある。

4.4. 偽記憶による想像力の実現

もう一つ、連想記憶による推論モデルからの展開として、想像力を実現できる可能性がある。想像とは、これまで経験したことのない事物のイメージを作ることであるが、その事物の個々のパーツは過去に経験したことのあるものであることが多い。すなわち、過去に経験した事物群のパーツを集めて一つの事物イメージを作るとは、想像という行為になると考える。このような特性は連想記憶ではスプリアス記憶、あるいは偽記憶として知られており[14]、記憶の想起という面では間違いであるが、過去に経験した事物から新しい記憶イメージを作るという意味では想像、あるいは創造と呼ぶことができよう。人間の創造においても、まったく未経験の事物をゼロからイメージすることは難しく、その人物の知識の範囲内での創造がまず出発点であろう。記憶した知識の内容を制御できる連想記憶を用いたスプリアス記憶による想像機能の検証もまた、人の想像力のモデル的理解の一つの方法かもしれない。

4.5. 他の脳理論との関係

4.5.1. 統合情報理論 (IIT)

脳における意識に関する理論はそれほど多くは無い。その中で近年注目されたのが Tononi による IIT[27][28]である。この理論では、脳を大規模な相互結合型、すなわち本稿でいう自己想起型の神経回路の集合と考え、そこに含まれる情報量をもって意識の質とみなす。神

経回路を分割すればそのそれぞれに情報量が存在して意識があるとする。

この理論は、本モデルと極めて近い構造と考え方を持っている。連想記憶の機能は相互結合された神経回路に生まれ、回路が分断されたらそのそれぞれがまた連想記憶として機能してする。その個々の回路に含まれる記憶の量を IIT でいう情報量と考えるなら、神経回路の細胞数に比例する連想記憶の記憶容量は統合情報量と極めて近い関係にある。ただ、IIT は意識の質、すなわち複雑さを定量化する理論であるが、なぜこれが意識であるかという説明としては不足感がある。

実際の脳では、タスクに応じて脳回路が選択的に活性化されて個々の瞬間に異なる神経回路を構成しており、その活性化された領域の組み合わせに応じて幅広い記憶想起状態が生まれている。そのような場合に選ばれて組み合わされた神経回路が入力に対して推論する記憶想起過程は GWT でいう意識という解釈もありうるし、IIT でいう意識という解釈もありうる。ここは、関係者が集まっての議論が必要かと思われる。

4.5.2. 自由エネルギー原理

フリストンは自由エネルギー原理による脳の活動の理論を提案している[29]。その基本的な考え方は能動的推論であり、知覚において感覚入力とそれの解釈から生成される予測とが一致するように解釈を積極的に変えることが認識であるとしている。またその理論を運動や意思決定にも拡張し、脳の幅広い活動が共通の原理によって説明できるとしている。

自由エネルギー原理は極めて魅力的な理論であり、脳の病的状態も含めて極めて高い説明力がある。一方で、推論については直感的推論はみごとに説明しているが論理的な推論についてはいまだ不明確な部分もある。連想記憶の想起過程のエネルギー最小化の原理と自由エネルギー原理は抽象度が異なるようにも見えるが、現時点では両者の対応についての検討が不十分であり、今後の検討課題としたい。

5. 構成的による検証可能性

ここまで、連想記憶の想起過程に基づく推論モデルとそれに関連しそうな心の機能についての仮説を述べてきた。では、この仮説は検証できるであろうか。

一つの方法は構成的研究、すなわち作って動作を確認することでの評価である。連想記憶による論理的推

論は計算機シミュレーションで検証できている。しかしその課題は単純なトイタスクであり、より現実的で複雑な課題での検証が必要である。特に作業記憶や社会性の評価につながる課題の設定は重要であり、そこで GWT に関わるあたかも意識があるように見える動作生成の試みが必要であろう。

結果として、これは現実的な仮想世界で活動する行動主体を制御する認知アーキテクチャの構成的研究になろう。このような知的行動主体の研究はこれまでも多くあるが[30][31]、図4はそのような世界の一例である。この世界に生きるエージェントには、環境を認識して移動、事物を認識して目標設定、仲間や敵を認識して接近や回避、さらに他者への働きかけによるコミュニケーションや共同作業など、多くの選択可能なタスクが存在する。この世界での生存には、個々のタスクの達成による獲得価値を評価して、いま実施すべきタスクを選択し、さらにそれを実行する行動計画のための推論と遂行など、認知アーキテクチャとして推論を超えた幅広い機能が要求される[32]。

GWT を実装した認知アーキテクチャに LIDA があるが、LIDA は GWT の意識の機能は人手に組み込んだプログラムで実行することを想定しており、脳的な計算メカニズムでの GWT の実装はまたもう一つの課題である[33]。認知アーキテクチャの現状を考えるとチャレンジングな内容であるが、意識様のふるまいを伴う論理的推論を実行できる知的処理というのは、十分に挑戦しがいのある課題である。



図4 論理的推論の認知アーキテクチャに想定される現実に近い仮想環境の一例。食料・水などの生存に必要な材料だけでなく、仲間や敵などの社会性に関わる事物も含まれていることが望ましい。

6. まとめ

価値評価と連想にもとづく直感的推論と論理的推論の脳過程についての仮説的モデルについて述べた。計算論的にはありうるのだが、これが認知過程として人の中にあるのか、という点についてはヒトの行動レベルで多くの検証が必要である。

また、本稿では社会性の鍵となる他者の価値を含む価値計算メカニズムについても検討した。社会性は次世代 AI においても求められる重要な機能であり、認知科学の世界でもインタラクション研究として注目されている。そこでの鍵は他者モデルであるが、本モデルからは他者モデルには他者の行動予測だけでなく他者の価値認識を含む必要があることが示唆される。

以上、神経回路による連想記憶という古典的なモデルから出発して、直観的推論と論理的推論のメカニズムの提案、感情の価値評価システムとしての可能性、さらにそこから派生する思考における意識様の状態の効用、作業記憶、社会性の計算的理解、想像力などについて述べてきた。いずれも次世代 AI にて実現できればと夢見るものである。AI の進歩の背景には脳科学からの多くのヒントがあるが、同様に認知科学からも多くのヒントがあったと考えるべきであろう。それらが集約された認知アーキテクチャの開発に認知科学が力を発揮することを祈っている。

文献

- [1] 大森隆司 (2021), 人の心の過程を計算で表現する, 特集 戸田正直『心理学の将来』から半世紀, 認知科学 第 28 巻 第 3 号, 招待論文
- [2] Paivio, A (1986). Mental representations: a dual coding approach. Oxford University Press.
- [3] ラッセル: エージェントアプローチ人工知能第 2 版, 共立出版, 2008
- [4] ヴィンター・エヤル (2017), 愛と怒りの行動経済学—賢い人は感情で決める—, 早川書房
- [5] 信原 幸弘(2017), 情動の哲学入門: 価値・道徳・生きる意味 勁草書房
- [6] Nakano K.(1972), Associatron-A Model of Associative Memory, IEEE SMC,2, 380-388
- [7] 中野 馨(1979). アソシアトロン—連想記憶のモデルと知的情報処理, 昭晃堂
- [8] Baars, B. (1988). A Cognitive Theory of Consciousness. Cambridge, MA: Cambridge University Press.
- [9] Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. Psychological Review, 1036–1060
- [10] Sun R.(2016), Anatomy of the Mind: Exploring Psychological Mechanisms and Processes with the Clarion Cognitive Architecture. Oxford University Press
- [11] O'Reilly, R.C., Hazy, T.E. & Herd, S.A. (2016). The Leabra Cognitive Architecture: How to Play 20 Principles with Nature and Win!. S. Chipman (Ed) Oxford Handbook of Cognitive Science, Oxford: Oxford University Press.
- [12] Jonathan St. B. T. Evans et al. : How many dual-process theories do we need? One, two, or many?, Oxford Scholarship Online, 2009
- [13] 服部 雅史(2015). 思考と推論: 理性・判断・意思決定の心理学 北大路書房
- [14] 中野 馨(1990), ニューロコンピュータの基礎 コロナ社
- [15] Takashi Omori et al. : Emergence of symbolic behavior from brain like memory with dynamic attention, Neural Networks, Vol. 12, No. 7-8, pp. 1157-1172, 1999
- [16] 戸田 正直(1992), 感情—人を動かしている適応プログラム, 東京大学出版会
- [17] 宮田真宏・大森隆司(2019), 価値に駆動された連想記憶に基づく人の推論過程の統合モデルの提案, 知能と情報 31(3), 712-721.
- [18] 宮田, 大森(2019), 感情の価値計算システム仮説にもとづく人の推論システムの提案, 人工知能学会大会, [3G4-OS-18b-04]
- [19] 大森 隆司(2020), ヒト脳にシンボリックな思考を生み出す脳アーキテクチャについて, 人工知能学会研究会資料 SIG-AGI-014-07
- [20] Dehaene S., & Changeux J.P.(2011). Experimental and Theoretical Approaches to Conscious Processing, Neuron 70, 200-227.
- [21] ドゥアンヌ, スタニスラス(2015), 意識と脳—思考はいかにコード化されるか 紀伊国屋書店
- [22] Sompolsky (1986), Temporal Association in Asymmetric Neural Networks, Physical review letters, Vol. 57, No. 22, pp. 2861-2864
- [23] Baddeley, A.D., & Hitch, G.J. (1974). Working memory. In The Psychology of Learning and Motivation (Bower, G.A., ed.), 47–89.
- [24] <https://ja.wikipedia.org/wiki/ワーキングメモリ>
- [25] Collins A, Koechlin E (2012) Reasoning, Learning, and Creativity: Frontal Lobe Function and Human Decision-Making. PLoS Biol 10(3): e1001293.
- [26] Fukuda et al.(2019), Computing Social Value Conversion in the Human Brain, The Journal of Neuroscience, 39(26):5153–5172,
- [27] トノーニ,ジュリオ・マッスィミーニ,マルチェッロ (2015) 『意識はいつ生まれるのか 一脳の謎に挑む統合情報理論—』 花本知子訳, 亜紀書房
- [28] 大泉 匡史(2018), 統合情報理論から考える人工知能の意識, 人工知能 33(4),460-467.
- [29] 乾 敏郎・阪口 豊(2020). 脳の大統一理論 自由エネルギー原理とは何か 岩波科学ライブラリー
- [30] LIS(Life In Silico): <http://wba-initiative.org/1036/>
- [31] Lawrence W. Barsalou (1999), Perceptual symbol systems, Behavioral and Brain Sciences, Vol. 22, pp. 577-660
- [32] 大森, 栢沼, 宮田 (2021), 連想による価値評価が駆動する実時間タスクスイッチング, 人工知能学会研究会 SIG-AGI-017-02
- [33] Franklin S., (2011). Global Workspace Theory, SHANAHAN and LIDA, International Journal of Machine Consciousness, 3(2)