

人工知能がもたらす新しい世界像 ～「相対化する知性」から～

2023年10月18日

東京大学未来ビジョン研究センター客員教授

西山圭太

今日お話ししたいこと

- いま決定的な変化が起きつつある(第○次産業革命というより第2次認知革命)
- それを端的に示しているのが人工知能の発展であり、絞ればディープラーニング技術だということになる
- コンピュータの発明以降のデジタル化にその伏線はあった
- 決定的な変化を捉えようとする和我々の世界像が変わる
= それをしない限り今起きている変化の実相は捉えられない
- それが人文科学、社会科学のあり方にも影響する

ディープラーニング超入門

達成したこと

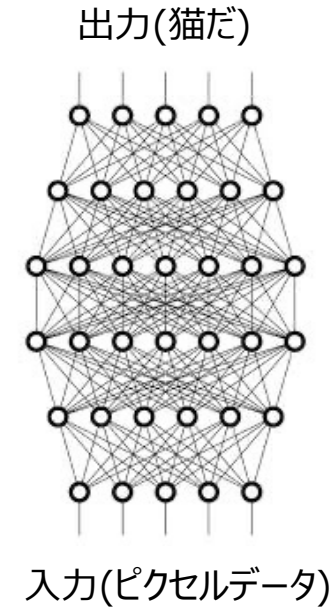
- 画像認識の精度向上(人間を上回った)
- アルファ碁
- 自然言語処理(GPT-4、BARD)

人工知能の歴史

- 第一の波:探索木
 - 第二の波:エキスパートシステム
 - 第三の波:ディープラーニング
- ⇒ 人間の論理や知識を教え込んだのではない
- ⇒ 学習の仕方を身につけさせた(Learning)

凄み:パターン認識

- パターン(猫が猫であること)を自分で見出した
- 不得意科目を克服した
- 「チェスはできて囲碁はできない」が常識だった
 10^{170} vs 10^{80} (探索木の限界)
- 言語処理まで相当できるようになった



仕組み

- ニューラルネットワーク
- 入力⇔出力の学習...大量のデータ
- 隠れ層...「深層」学習(**Deep**)
- 誤差逆伝播法
- 確率的勾配下降法(SGD)
- 活性化関数
- 教師あり・なし

課題

- 言語処理 ⇔ シンボルグラウンディング

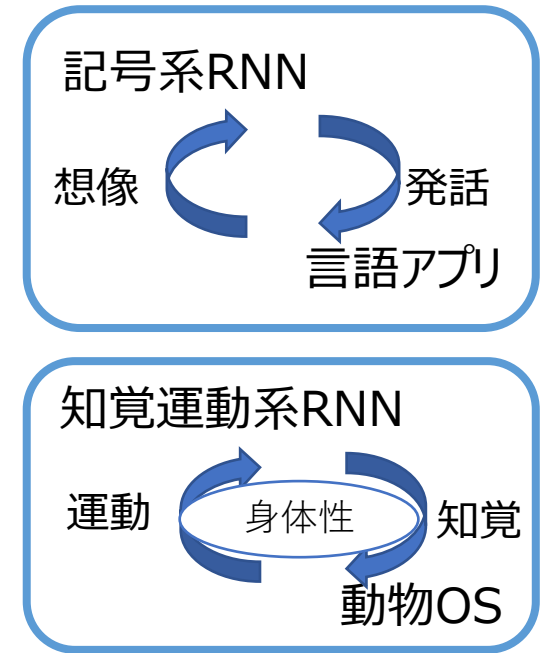
かつてはこう思われていた

私はここで1つの仮説を提起しよう。それは、人間のもっている、そして人間だけに高度に恵まれている形態に対する認知能力が数学の本当の起源なのだ。それは今日流行の表現をかりれば、**パターン認識**の能力であるといってもよい。今日、驚異的な能力をもったコンピュータが出現しつつあるが、それは一定のプログラムを設定してやれば人間をはるかに超える能力を発揮することはよく知られている。だがこのことに幻惑されて人間の能力について悲観的になる人がいるとしたら、それは愚かなことだ。<中略>コンピュータがどれほど発達してもおそらく人間を追い越すことのできない能力が人間には備わっている。それが、ここでいう**パターン認識**そのものである。

遠山啓『代数的構造』(原著1972年刊行)

「AIがパターン認識ができる」ことは、次の問を招来する

- なぜ人工物であるAIにパターン認識が可能なのか
- そのメカニズムは人間以外の生命体の知的活動にも共通ではないか
- パターン認識と言語の利用とはどのような関係にあるのか
- 人間の言語の習得や文化・社会的規範の習得はどのように行われるのか
- (人間であれAIであれ)パターン認識がなぜ上手いく(生存に役立つ)のか

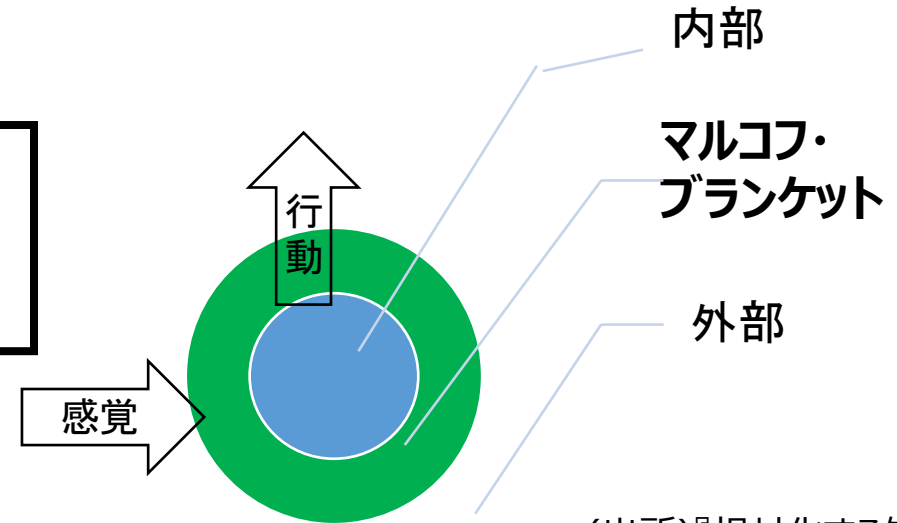


(出所)『相対化する知性』第1部

統一理論?: マルコフ・ブランケットと自由エネルギー原理

カール・ Friston(脳神経科学者)らの取り組み:

- 生命体や脳の基本的なメカニズムは、機械学習の延長線上で説明可能
- 経済、社会の組成にも応用可能ではないか
- 非生命体のメカニズムにも拡張可能ではないか



(出所)『相対化する知性』第2部

- システムが局所的な作用を中心に構成されるときメカニズム
- **機械学習**のメカニズムの応用:あるシステムに対して影響を与えるもの(親)とそのシステムが影響を与えるもの(子)との関係性をベイジアンネットワークの中で説明したもの。それを Fristonらは環境とシステムとの関係に応用した。
(注)このメカニズムをマルコフ・ブランケットと命名したのは Judea Pearl

現れるメカニズム:

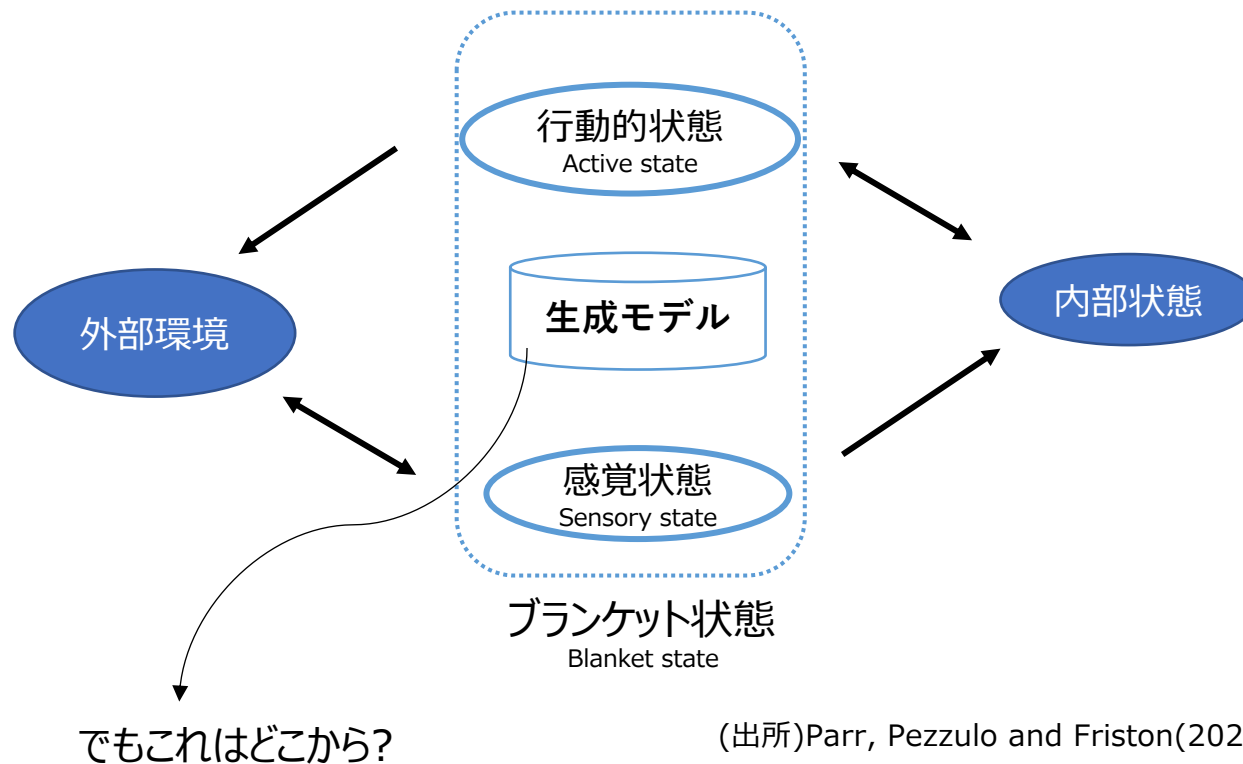
- 外部環境を予測し、介入し、適応しようとする
- 内部と外部に仕切りが生ずる・・・例えば細胞膜(ただし物理空間内に固定されとは限らない)

これがパターン(秩序)を認識(しかつパターンを生成)するメカニズムではないか

マルコフ・ブランケット(1) 個体レベル

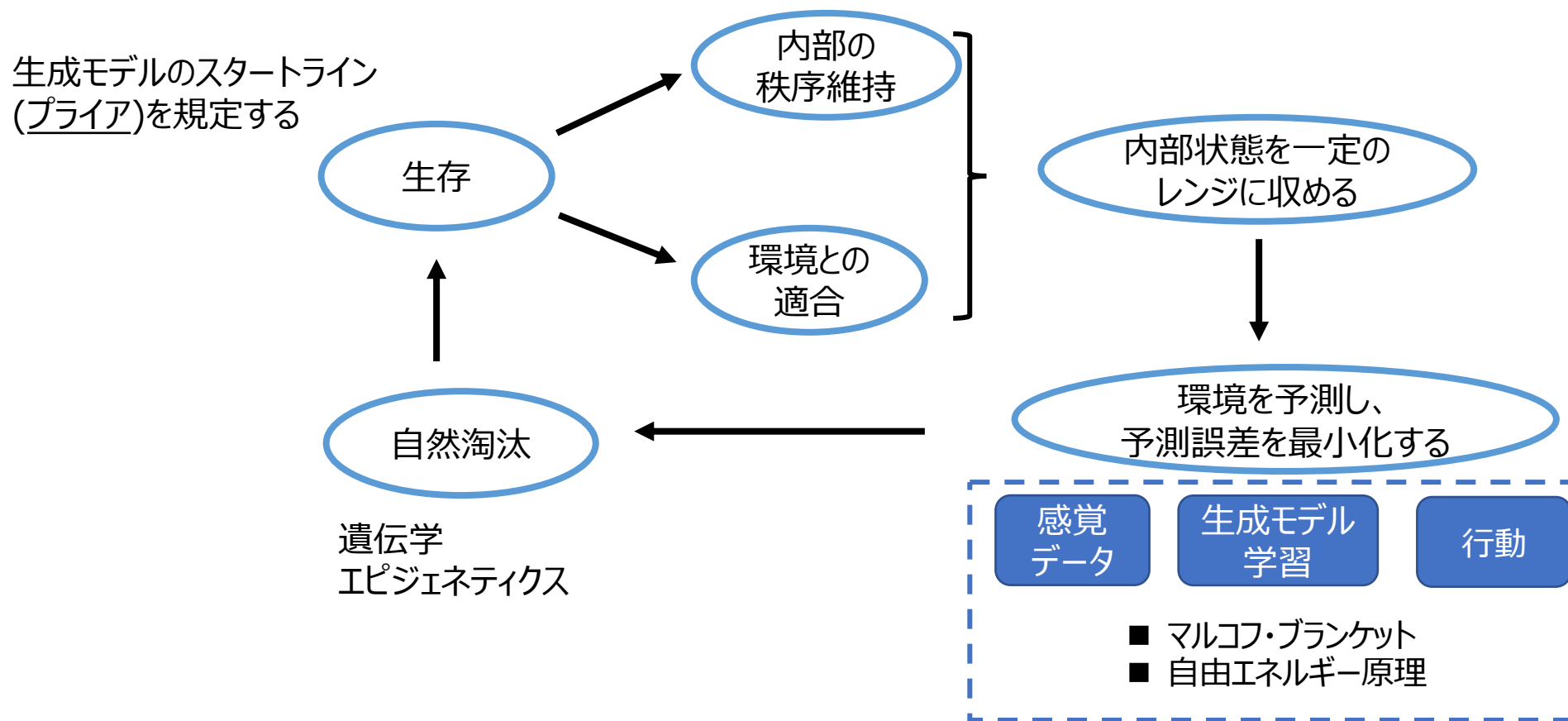
メカニズム

- 自己の感覚状態をあるレンジに収める
⇒生存のため
- 外部環境からのサプライズを最小化する
⇒**予測を最適化**する
- 自己のあり方自体についての観察でもある
⇒Self evidencing・・・「ある」「知る」
- **生成モデル**を作り更新する
- 環境に行動で働きかける
 - ・ 観測にアテンションをかける
 - ・ 物理空間を移動して別の環境に身をおく
- 外部と内部は直接接しない



(出所)Parr, Pezzulo and Friston(2022)

マルコフ・ブランケット(2) 類レベル



マルコフ・ブランケットの展開(1) ... 能動的推論(active inference)

(Parr, Pezzulo and Friston(2022))

より複雑な生存戦略や、それに伴う複雑性の上昇自体も、マルコフ・ブランケットの延長線上で説明可能。

探索

発達した生命体 → 環境からのデータに対して同様の反応を繰り返すだけではない

- 積極的にデータを収集し、不確実性を減少させる
- 今はない秩序やそれに対する制御可能性を探求する
 - 今いる周辺環境(ニッチ)だけではなく、新たなニッチも試す(新たな都市、新たな文化の構築)
- 短期的にはエントロピーを増大させるが、長期的には減少させる行動
 - 新奇性の追求、探索行動はマルコフ・ブランケットのモデルと矛盾しない
 - 不確実なもの(予測に必要なエントロピー)はむしろ最大限残す
- ホメオスタシス→アロスタシス

計画

- 自分の行動が起こす結果をシミュレーションし、複数の選択肢から行動を選択する
- 反実仮想を行う

⇒ここまで含めると**能動的推論**

自由エネルギー原理

上記の一見区々な行動を一つの変数の最小化として表現できる...**変分自由エネルギー**(Variational Free Energy)

...‘as if’だが、計算可能性を示しているとも言える

- 変分自由エネルギー = KL divergence(カルバック・ライブラー情報量:事前事後分布の差の最小化 + 複雑性のより少ないモデル) + 証拠探索
- 計画 = 予想変分自由エネルギーの最小化

マルコフ・ブランケットの展開(2) エマージェンス、階層構造

(Palacios, Razi, Parr, Kirchhoff and Friston (2020))

- 生命体は、細胞→器官→個体のように、より複雑な秩序に至る構成を持つ。
- それを基本的な単位の振る舞いだけでは説明できない → DNAやタンパク質だけから形態発生を説明できない
- マルコフ・ブランケットは**形態発生、自己組織化を説明可能**
 - ・ 「ミクロのマルコフ・ブランケットの上にマクロのマルコフ・ブランケットができる」仕組み
 - ・ 前提となるプライア: 各々のマルコフ・ブランケットがより上位のマルコフ・ブランケットにおいて、内部状態、行動状態、感覚状態のいずれかを担う
 - ・ マルコフ・ブランケットという構成自体がアトラクタ、秩序変数になる
- この自己組織化は、次々とより上位のレベルを構成するように繰り返され、**複数のレイヤーからなる階層構造**が作られる
- 司令塔的な役割はなく、全てのレイヤーは一つ下のレイヤーに対してマクロ的な機能を果たし、その上のレイヤーに対してミクロ的な機能を果たす
- 複数レイヤー間の相互作用は、生命体以外の複雑な秩序(文化の形成や都市環境の形成)にも適用できる

階層構造と科学: モア・イズ・ディファレント

階層構造の議論は科学のあり方とも関係する

PW Anderson(1972) “More is different”

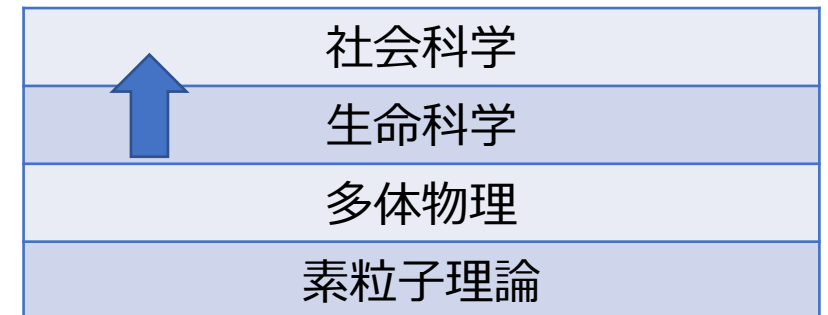
- 人間の精神や身体の働き、生命体も非生命体も同じ基礎的法則に従う(還元主義仮説)
- しかしそれを逆転させて基礎的な法則から宇宙の様々な事象を再現はできない
- 複雑性の段階に応じて全く新たな特性が出現し、それを理解するには本質的研究が必要

→つまり**科学は複雑性の段階に応じた階層構造**をなしている、と言っている(例えば「文系と理系」のようには考えない)

マルコフ・ブランケット的な考え方

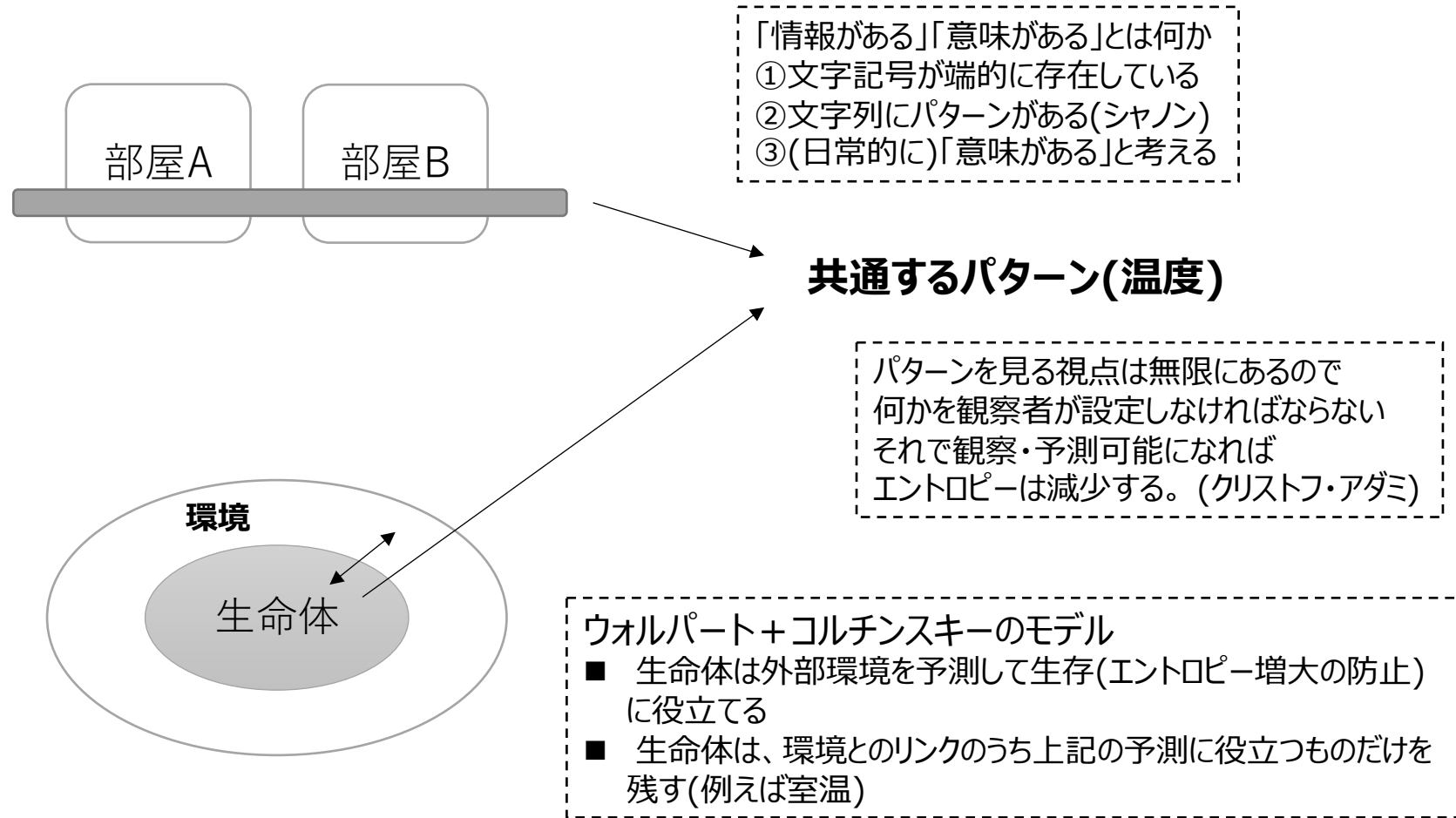
⇒複雑性の段階に応じた階層構造を前提として
その複雑性の発生のメカニズム、レイヤーとレイヤーとの関係性
を説明しようとしている

⇒経済学を含めた科学のあり方に影響する



パターン認識と情報

(カルロ・ロベッリの議論を基礎に『相対化する知性』で整理)



- 発想は能動的推論と同じ・・・生命体と環境との関係がまずある
- 「寒いから部屋のドアを閉める」と考えれば、行動・計画に結びつく
- 行動(政策)選択の判断は、本来局所的な視点の設定(関心のある構造の取り出し方)と結びついている

人間は実際にどう学習しているのか: Theory Theory

- **Alison Gopnik**が主唱。ここ十年くらいで児童心理学の考え方に転換をもたらしたとされる(Gopnik and Wellman 2012)
- 従来の発達段階論の有力説は、生得説(先天的に認知構造を持っている)か、データ結合説(個々の感覚データを結びつけて世界を捉えている)だが、Gopnikはピアジェ以来の**構成説**の再構築を企図
- 子供は**まず抽象的な認識枠組みを幼児の段階で習得**し、その上に応用問題を次第に解いているという考え方
 - ピアジェの時代にはそれを証明するためのcomputation modelはなかったが、近年の機械学習モデル(Bayes Net、Hierarchical Bayes Net)を足がかりにできる
 - 近年の児童心理学のさまざまな実証研究は、児童は先に抽象化する力を学んでいることを指し示している。子供は幼児を含めて大人並みに(ある意味はでそれ以上に)反実仮想思考が発達している
- **認識枠組み ≡ 生成モデル**

なお、フリストンらは、Theory Theory に対する批判(Theory Theoryは認識枠組みを強調するのに対して、文化的習慣や他者の行動に対する条件反射的な反応の役割を重視する議論)にも触れつつ、それらは自由エネルギー原理をベースにする彼らの枠組みで統合的に扱えるのだ、という議論を展開している (Veissiere, Constant, Ramstead, and Friston (2019))

世界になぜ秩序はあるのか: 同型論

なぜ「同型論」なのか

FAハイエク『感覚秩序 The Sensory Order』(1952)

- 人間が世界について「知る」= 世界から受ける刺激を元にある共通のパターンを見出すこと
- それを実現しているのがニューラルネットワークの働き(「神経秩序」)
- パターンが形成されるとそれは予測機能を果たす
- 神経秩序は世界の局所的な位置から周辺の外部環境を把握・予測している
- 神経秩序とそれを通じて獲得されるパターン認識(現象的秩序、感覚秩序)とは**同型(isomorphous)**である
- 感覚秩序はあくまでも局所的なものなので物理的秩序自体とは相似(similar)だが異なる

⇒ハイエクとフリストンの議論は近いものも多い(両者とも源流はヘルムホルツだとする)が、フリストンの議論は、生命体一般の同じメカニズムの中で知能の働きを議論しているので、同型性の範囲が広い。(→これに基づく世界像を**強い「同型論」**と呼んだ。)

- 生命体-AI-人間の知能-(非生命体?)
- 階層構造とエマージェンス
- 「知る」と「ある」

世界像が実在論的になる

- 名目論的な主張の根源の一つは、「人間の知性は他にはない固有なものである」という考え方である
 - ⇒ そうなるとそれが認識の対象と一致する保証はない(カントの「物自体」)という論理展開になる
- しかし今起こっていることはその前提を突き崩しつつある。

Fristonらの考え方

- 生命体の持つ生成プロセスと生成モデルとは同じではない
同時に...
- なぜ階層構造が生命体システムに広く見出されるのか?
 - ⇒ 一つの仮説は、階層構造的で自己相似的なマルコフ・ブランケットというアーキテクチャが、長期的に自由エネルギーを最小化することで進化に適合的だから
- (推測の域を出ないが)動植物が驚くほど多様なのは、シグナルが空間に依拠しているこの世界では、マルコフ・ブランケットが存在と同義(synonymous)であるから (Palacios, Razi, Parr, Kirchhoff and Friston (2020))
 - ⇒ 複雑な秩序が存在していることの根拠そのものへと向かう(実在論的)

さらにこれを拡張すると...

マルコフ・ブランケットと階層構造を組み合わせると、それはブランケット状態の連続(外部環境も内部構造も各々別のレベルのマルコフ・ブランケットになるという、一種のマトリョーシカ的關係)になる
⇒ 量子理論で議論されているような関係的世界観と親和性が高い
⇒ そこまで含めて我々の世界像を「同型論」と言っている

人間は世界内存在である＝知性を特別視しない(『相対化する知性』第3部)

- 「観察者としての私がいることを含む世界をどう説明するか」が大事。観察者も自然の一部だから、特権的な地位に立たせる必要はない
- エルンスト・マッハ(ハイエクもマッハに依拠している)は人間の観察を重視するので非自然主義的に捉えられるが、マッハは人間を世界内存在として考え、人間の感覚も世界内の自然現象だと考えた
- ウィラード・クワイン＝我々の知は、自然的な過程の一つであり、そういうものとして学習することができる

関係的世界観

- 量子理論は、「宇宙のある一部が、宇宙の別の一部について立ち現れる仕方」を説明している
 - ・ 実体ではあるが、他とは一切干渉しない実体を考えるということは、存在しないことと同じ
 - ・ 対象の性質・属性propertyと対象が干渉を通じて「他に示す」属性とを区別することは不可能だ
- 物理的な世界を、固定的な属性をもった実体の集合と考えるのではなく、関係性の網として考えるのが量子理論

諸科学との関係

- 量子理論 ＝ 概念や思想を直接説明はできない
- しかし間接的には関係⇒問いの性格を変更する
 - ① 厳密な精神世界と物理的な世界の区分は後退
 - ② 物理的な世界が基礎的な物質とその属性からできているのなら、「相関」は二義的である。単に「相関している」というだけでは説明にならず何か追加的な説明が必要。しかし量子理論は、物理的な世界がそもそも相関的な網目なのだという考え方。意味や意図はこの相関の網の目の中にある特殊ケースだ、と理解することができる。
 - ③ 世界の隅々にレイヤー構造がある
 - ・ 個々のレイヤーに我々が関与することで、規則性を見出すことができる
 - ・ そこから我々は情報を得て、そのレイヤーの像をある程度の一貫性を持って捉えることができる

まとめ: 新しい世界像

- **自然主義的**: 人間の知性、生命体、AIを同一の地平で捉える(精神と物理とを厳密に区別しない)
- **实在論的**: 「ある」と「知る」は区別できない(パターンを知り、パターンが生まれる)
⇒「**同型論**」(同型のメカニズムの繰り返し・展開で世界は成り立っている)
- **局所的な環境への適応**をスタートラインに置く
⇒個別システムを超えた司令塔、グローバルな最適性の存在を前提としない＝「**分散型**」の発想
- **変化・適応の中に安定**があり、常に**進化**する
- 進化を説明しようとするすると複数レイヤーからなる**階層構造**に至る
 - ・ モア・イズ・ディファレントの発想と通底する
 - ・ 特定のレイヤーに関心を絞れば、そこには規則性が見出され、ある程度一貫した理解ができる
 - ・ レイヤー間の関係(進化)が説明可能になりつつある・・・**エマージェンス**の説明
 - ・ その一つとして規範や文化が説明される
- 局所的な主体が持つ**計算能力**・情報処理能力から説明する・・・**情報理論的**
 - ・ それはベ이지アン的であり、「**主観的**」である
 - ・ モデルがデータと適合的かどうかという評価は客観的だが、推論自体は主観的
- **関係的世界観**: ブランケット状態の連続
 - ・ 宇宙のある一部が、宇宙の別の一部について立ち現れる仕方
- 世界を観察する人間も(AIも)また**世界内存在**である

参考文献

- 西郷甲矢人、田口茂(2019) 『<現実>とは何か 数学・哲学から始まる世界像の転換』 筑摩書房
- 遠山啓(2011) 『代数的構造』 ちくま学芸文庫 筑摩書房
- 西山圭太、松尾豊、小林慶一郎(2020) 『相対化する知性 人工知能が世界の見方をどう変えるのか』 日本評論社
- Friston K. (2010), “The free energy principle: a unified brain theory?” *Nature Reviews. Neuroscience* 11(2)
- Gopnik, A. and H.M. Wellman (2012), “Reconstructing Constructivism: Causal models, Bayesian learning mechanisms, and theory theory” *Psychol Bull.* 2012 November 138(6): 1085-1108
- Hayek, F.A.(1952), *The Sensory Order: An Inquiry into the Foundation of Theoretical Psychology*, The University of Chicago Press
- Palacios, E.R., A. Razi, T. Parr, M. Kirchhoff, and K. Friston (2020), “On Markov blankets and hierarchical self-organisation” *Journal of Theoretical Biology* 486
- Parr, T., G. Pezzulo, and K.J. Friston (2022), *Active Inference: The Free Energy Principle in Mind, Brain, and Behavior* The MIT Press
- Rovelli, C. (2021), *Helgoland : Making sense of the quantum revolution*, translated by Erica Segre and Simon Carnell, Riverhead Books
- Veissiere S., A. Constant, M.J.D. Ramstead, and K. Friston (2019), “Thinking Through Other Minds: A Virtual Approach to Cognition and Culture” *Behavioral and Brain Sciences*, May 2019