

A wide-angle photograph of a lush green field under a clear blue sky. In the foreground, several dark purple, bell-shaped flowers with green leaves are in focus. The field extends to a flat horizon line. In the distance, a traditional windmill with a blue roof and white sails is visible on the right side. The overall scene is bright and open.

創発について考える

# はじめに

- GPT-4との対話から考えたこと
- セミナーの構成について

# GPT-4との対話から考えたこと

セミナー「創発について考える」  
はじめに 1

## GPT-4との対話から考えたこと

この間GPT-4と「対話」を楽しんできたのですが、印象に残ったやりとりがありました。

それは、僕がGPT-4に、「記憶」や「学習した情報」を、GPT-4は、内部ではどのように組織しているのかと尋ねたときです。

## GPT-4が答えたこと

「具体的な「記憶」の概念について話すと、GPT-4は単語や情報を個々に「覚えている」わけではありません。...

「学習結果」は、モデルの全体的な構造とその重みとバイアスによって組織化されています。

これは人間が個々の事実や情報を脳に直接保存するのとは対照的で、より抽象的な表現を生成します。」

## 多分、GPT-4は人間のことに誤解している

これは面白い答えです。なぜなら、GPT-4は、人間のことに誤解しているからです。

我々人間は、「事実や情報を脳に直接保存」しているわけではありません。そういうスタイルで情報の記憶をしているのは、コンピュータのメモリーやデータベースであって、人間ではそうではありません。

我々人間が、脳の内部で「記憶」や「学習した情報」を、どのように組織しているのかをGPT-4に答えるとすると、次のようになるとおもいます。

# 人間がしていること

「人間の脳内の最下層の「記憶」のメカニズムについて話すと、人間は単語や情報を個々に「覚えている」わけではありません。...

「学習結果」は、脳のニューロンの全体的な構造とそのシナプスの重みとバイアスによって組織化されています。

我々人間は、コンピュータのメモリーやデータベースのように、個々の事実や情報を脳に直接保存しているわけではありません。それから、より抽象的な表現が生成されます。」

## GPT-4と人間は、 これらの点では、とてもよく似ている

GPT-4と人間の両者は、これらの点では、とてもよく似ています。

GPT-4のベースとなっているディープラーニングのニューラル・ネットワークのモデルが、もともとは人間を含めた生物のニューラル・ネットワークをモデルにしていることを考えれば、両者の類似は不思議なことではないかもしれません。

ただ、両者の間には、最下層のニューラル・ネットワークのレベルでの類似を超えた類似があるように見えます。両者ともに、「記憶」や「学習」や「知能」といった高次元の構造が、最下層の構造のうえに存在するようになって見ると言うことです。

## 両者はよく似ている

### GPT-4:

「具体的な「記憶」の概念について話すと、GPT-4は単語や情報を個々に「覚えている」わけではありません。... 「学習結果」は、モデルの全体的な構造とその重みとバイアスによって組織化されています。これは人間が個々の事実や情報を脳に直接保存するのとは対照的で、より抽象的な表現を生成します。」

### 人間:

「人間の脳内の最下層の「記憶」のメカニズムについて話すと、人間は単語や情報を個々に「覚えている」わけではありません。... 「学習結果」は、脳のニューロンの全体的な構造とそのシナプスの重みとバイアスによって組織化されています。我々人間は、コンピュータのメモリーやデータベースのように、個々の事実や情報を脳に直接保存しているわけではありません。それから、より抽象的な表現が生成されます。」

## 「創発」とは何か？

このように、低次元の構造の上に、まったく新しい働きをもつ高次元の構造が生まれることを「創発」と言います。

# 「創発」の奇妙な性質

「創発」という現象は、奇妙な性質を持っています。

第一に、低次元の構造（例えば、ニューロンのネットワーク構造）だけに注目すると、高次元の構造（例えば、「記憶」「学習」「知能」といった）は、目に入らずに隠れてしまいます。あるいは、そうした高次元の構造は、謎めいた説明のつかないものに見えます。

第二に、今度は、高次元の構造に注目してみましょう。こうした高次元の構造は、それを生み出した低次元の構造と強く結びついていて、そのただなかから出現しているのは明らかです。それにもかかわらず、高次元の構造を低次元の構造に「単純に還元」することができないのです。

# 今回のセミナーの問題意識

今回のセミナーは、この「創発」がテーマです。

冒頭のGPT-4との対話の例は、我々は、我々人間が作った構造の上で、初めて新しい構造が「創発」されていることを目撃しているのかもしれないという問題意識によるものです。

ただ、今回のセミナーは、直接、この疑問に答えることを目的としたものではありません。

# エントロピーと創発

なぜなら、「創発」という現象については、20世紀に活躍した科学者たちが、既に多くのことを語っているからです。

それらは、意外に思えるかもしれませんが、「エントロピー論」と深く結びついています。

今回のセミナーは、そうした議論を紹介することを主な目的に設定しようと考えています。



# セミナーの構成について

セミナー「創発について考える」  
はじめに 2

# 「創発について考える」 セミナーの基本的問題意識

基本的な問題意識は、最近の大規模言語モデルの高い言語能力に触発されたもので、我々人間が作った大規模で複雑なシステムの上でも、新しい構造が「創発」されることがあるのかというものです。

それは、「人工知能」が、どのように人間と同様の知能を獲得することができるのかと言う問題意識と繋がっています。

# 「エントロピーと創発」 今回のセミナーの目的

残念ながら、僕には、こうした問題にすぐに答えることはできません。

今月のセミナーの主要な目的は、「創発」という現象を考える上で基礎になるとおもわれる、次の二つの研究を紹介することです。

- Prigogineの非平衡熱力学と散逸構造論
- Jaynesの最大エントロピー原理とベイズ推論

# Prigogineの非平衡熱力学と散逸構造論

Prigogine以前の熱力学は、エネルギーも物質も外部と交換することのない隔離され孤立した系の、最終的には熱的平衡状態に達する系を対象にしていました。

Prigogineは、こうした従来の熱力学を、非平衡状態を取り込んだ熱力学に拡張します。現代の熱力学は、このPrigogineの非平衡熱力学に基礎を置いています。

# Prigogineの非平衡熱力学と散逸構造論

エントロピーとエントロピーを生み出す不可逆的なプロセスは、一般的には、無秩序を生み出すと考えられていました。

驚くべきことに、平衡状態から遠く離れた非平衡状態では、不可逆なプロセスは、内部のエントロピーを最小の状態に保つ「自己組織系」を生み出すことを彼は見出します。彼はこうした構造を「散逸構造」と呼びます。

重要なことは、いわば、エントロピーの力で「創発」された「秩序」をもつ構造は、自然のいたる所に存在するということです。

# Jaynesの最大エントロピー原理とベイズ推論

Jaynesは、Gibbsらの統計力学でのエントロピー概念と Shannonの情報理論でのエントロピー概念が基本的には同一であることを明確に示します。

彼は、この二つのエントロピー概念の対応のもとで、統計力学的に最大のエントロピーを与える確率分布が、情報理論的に最小の情報を与える確率分布であることから、それがベイズ推論での事前確率を与えるという解釈を提案します。

# Jaynesの最大エントロピー原理とベイズ推論

Jaynesが提唱した、こうした解釈の強い形は、原理的に統計学的エントロピーが論理的な推論の形式を導くというもので、彼はそれを「最大エントロピー原理」と呼びます。

彼は、事前確率と事後確率の間の「相対エントロピー」が、ベイズ推論の構造を反映しているとして、その重要性を強調します。

「相対エントロピー」は、現在のDeep Learning 技術の中で大きな役割を果たしています。

# 生命と知能 「創発」論の二つの焦点

こうした20世紀に活発に行われた探究と、「人工知能」での「創発」は起こりえるのかという冒頭の問題意識との関係は、わかりにくいかもしれません。

ただ、こうした「エントロピーと構造の形成」をめぐる議論は多岐にわたるのですが、二つの重要な焦点があったことがわかります。一つは、物質の世界からの「生命」の創発、もう一つは、生命の世界からの「知能」の創発です。

少し単純化して言うと、Prigogineの散逸構造としての「自己組織」構造への注目は、前者の「生命」の創発に、Jaynesのベイズ推論の構造への注目は、後者の「知能」の創発に結びついています。

# Fristonの最小自由エネルギー原理と脳のモデル

今回のセミナーでは、より直接的に、現代の理論家たちが、脳のモデルをどのように構成しようとしているかを紹介しようと思います。代表的なものに次のものがあります。

- Fristonの最小自由エネルギー原理と脳のモデル

Helmholtzの自由エネルギー $F$ は、内部エネルギー  $U$ 、熱力学温度  $T$ 、エントロピー  $S$  として  $F = U - TS$  で与えられますので、Fristonの「最小自由エネルギー原理」は、Jaynesの「最大エントロピー原理」に対応するものです。

# 創発について考える **Agenda**

- **はじめに**
- **Part 1** : プリゴジンが考えたこと  
-- 非平衡熱力学と散逸構造論 --
- **Part 2** : ジェインズが考えたこと  
-- 最大エントロピー原理とベイズ推論 --
- **Part 3** : フリストンが考えたこと  
-- 最小自由エネルギー原理と脳のモデル --
- **おわりに**





Part 1 : プリゴジンが考えたこと

# Part 1 : プリゴジンが考えたこと

## 非平衡熱力学と散逸構造論

- 古典的熱力学から非平衡熱力学へ
- エントロピーの生成
- 「局所平衡仮説」の導入と不可逆過程の熱力学の基本式
- 不可逆過程の線形熱力学の二つの基本定理と「散逸構造」
- 機械と散逸構造の違い



# 古典的熱力学から非平衡熱力学へ

セミナー「創発について考える」  
Part 1 : プリゴジンが考えたこと 1

# 熱力学の基本法則

## Clausius 1865年 古典バージョン

熱力学の基本法則には、さまざまな定式化がありますが、19世紀のClausiusの次の定式化は興味深いものです。

- 宇宙のエネルギーは一定である。
- 宇宙のエントロピーは、最大にまで到達する。

19世紀の古典論においても、熱力学は「熱機関から宇宙まで」をカバーする非常に広い範囲のマクロな現象を説明するものとして意識されていたことを示しています。

古典的な熱力学は、マクロな観点から、例えば、結晶のような平衡構造の概念をほぼ解明することに成功します。

## 熱力学のミクロな側面

もともと、熱力学の理論的妥当性をめぐる当時の認識は、まだまだプリミティブなものでした。19世紀末から20世紀初頭にかけて、熱力学の理論的リーチをミクロの世界に広げるさまざまな科学的発見があいつぎます。

Boltzmannは、ランダムに運動する気体分子の運動から、エントロピーを導くことができることを示します。ミクロな世界の物質の運動の統計的性質からマクロの世界の現象を説明したのです。それは統計力学への道を開く画期的な発見でした。

Plankは「黒体輻射」の問題から、「量子」の概念を提案します。Einsteinの「ブラウン運動」や「光電効果」の研究も、熱力学と深い結びつきがありました。

# 不可逆な過程としてのエントロピーの増大

熱力学の第二法則は、エントロピー $S$ という関数が存在し、エントロピー $S$ は熱力学的平衡状態で最大になるまで単調増加するという主張に他なりません。

エントロピー $S$ の変化を $dS$ で表すと、第二法則は、次の簡単な式で表されることになります。

$$dS/dt \geq 0$$

$dS/dt = 0$  のエントロピーが最大の状態が、熱力学的平衡状態です。

エントロピーは、減少することがないまま一方的に増大します。それは、不可逆的な過程として進行します。Eddingtonは、エントロピーの増大する方向に時間が流れるという「時間の矢」というたとえを用いました。

# 熱力学的平衡状態は、どこにあるか？

熱力学的平衡は、また、次の式で定義されるヘルムホルツ自由エネルギー $F$ の最小値によって特徴づけられます。

$$F = E - TS$$

私たちの回りに、このような熱力学的平衡状態にあるものが数多く存在するのでしょうか？

その問への答えが、否定的なものであることは、すぐにわかります。多くの車が行き交う街も、緑の自然も、人間も動物も命ある全てのものは、決して静的な平衡状態にあるわけではありません。

# 非平衡の熱力学と「散逸構造」

それでは、それらは無秩序な混沌の中にあるのでしょうか？ そうではありません。それらは、動的ですがある秩序を保っています。

Prigogineが関心を持ったのは、静的な平衡状態にはないが、整合的な機能的秩序を持った、こうした構造です。彼は、こうした構造を「散逸構造」と呼びます。

彼は、この種の構造に対する熱力学的理論を構築するためには、非平衡が秩序の源となりうることを示さなければならないことに気づきます。

## 現代の熱力学 – 非平衡熱力学へ

先のClausiusの熱力学の第二法則の定式化「宇宙のエントロピーは、最大にまで到達する」は、外部との物質やエネルギーのやり取りのない孤立した系の法則として語られていることに注意が必要です。なぜなら、宇宙には「外部」が存在しないからです。

ただ、生命のシステムを考えれば、そこでは外部との物質やエネルギーの交換が、「代謝」として普通に行われていることがわかります。しかも、それは、生命にとって本質的に重要なものです。

Prigogineは、エントロピーの増大をもたらす不可逆なプロセスと、彼が「散逸構造」と呼ぶ、新しいタイプの動的な構造の生成を結びつける探究を通じて、現代の熱力学 – 非平衡熱力学を作り上げます。

# 非平衡熱力学が取り組んできた課題

現代の熱力学としての非平衡熱力学が、どのような理論構成をもつのかは、もう少し詳しく紹介していきたいと思います。非平衡熱力学は、化学反応論、生命科学、環境科学の分野で大きな成功を収めます。

特筆したいのは、20世紀の自然科学の一大飛躍であった、量子論と相対論と新しい熱力学との驚くべき関係です。20世紀の量子論も相対論も、古典物理学と同様に、時間を反転させても理論が成り立つ可逆な理論です。不可逆な熱力学とは大きく異なります。

21世紀の物理学の大きな焦点は、量子情報理論と量子論・重力理論との統合なのですが、それは、20世紀の熱力学が取り組んできた課題を引き継いだものと考えられます。



# エントロピーの生成

セミナー「創発について考える」  
Part 1 : プリゴジンが考えたこと 2

# 古典的な熱力学でのエントロピー $S$ の特徴付け

$$dS \geq 0$$

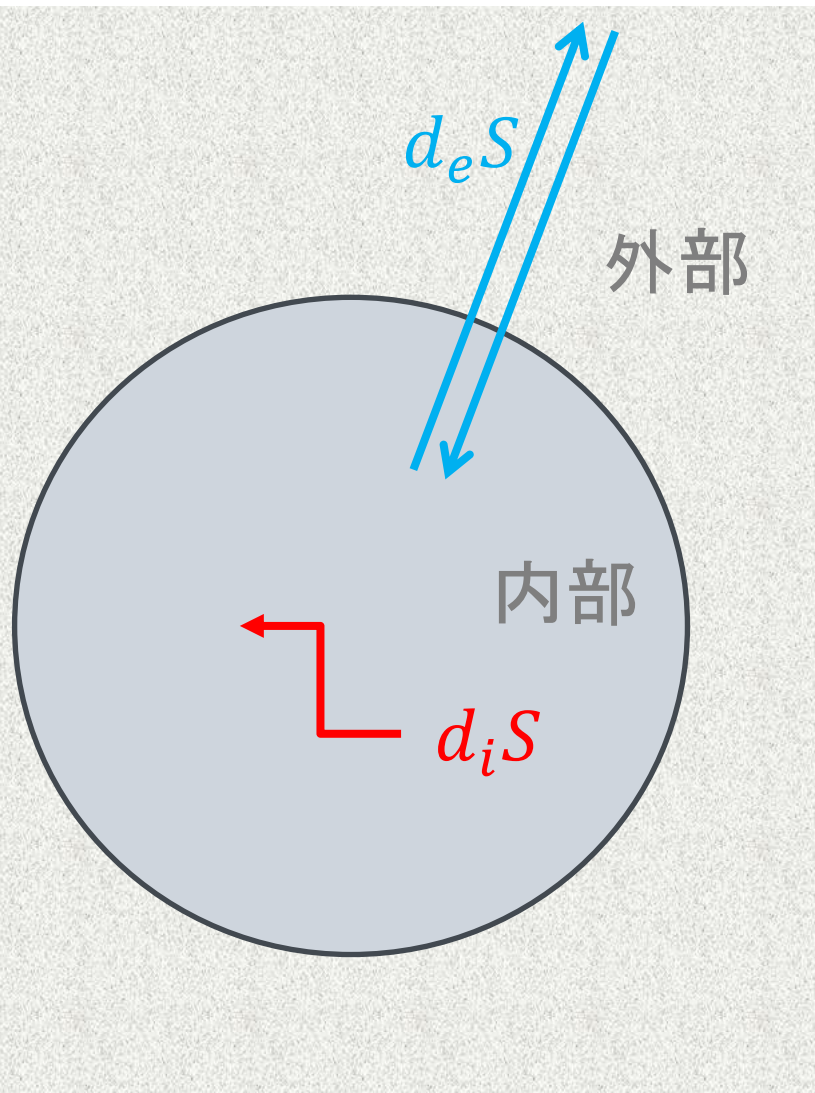
前回見たように、古典的な熱力学では、ある系のエントロピー $S$ は、次の式で特徴づけられます。それを熱力学の第二法則と言います。

$$dS \geq 0$$

古典的な熱力学は、エントロピーが定義されている系が外部の世界とエネルギーや物質を交換することを想定していません。

ですので、古典的な熱力学での第二法則は、孤立した系では、熱力学的平衡状態と呼ばれるエントロピーが最大の値を取るまでエントロピーは単調に増大することを意味します。

# 「外部」を持つ系のエントロピーの変化



古典論での孤立した系のエントロピーの変化 $dS$ を、「外部」を持つ系に拡張することは容易です。

$d_e S$ を、系が外部との境界を通じて外部と交換するエントロピーの変化とし、

$d_i S$ を系の内部で生成されるエントロピーの変化として、

$$dS = d_e S + d_i S$$

と表せばいいのです。

## 「外部」を持つ系の第二法則

エントロピーの変化が  $dS = d_e S + d_i S$  で表される時、一般的には、 $dS \geq 0$  は成り立ちません。

この時、外部を持つ系のエントロピーの変化を表す第二法則は、次の式で表されることになります。

$$d_i S \geq 0$$

# エントロピーの生成と不可逆過程

ここでは、可逆過程と不可逆過程の区別が重要です。不可逆過程のみが、エントロピーの生成に寄与します。

明らかに第二法則は、不可逆過程が時間の流れの一方向性を導くという事実を表現しています。時間の正の方向がエントロピーの増大に関連しています。

次に、その中で第二法則の時間の一方向性が表れる、もっと強くかつ特別な方法を紹介したいと思います。

# Lyapunov function (リアプノフ関数)

この定式化に従えば、孤立した系では、それは時間の中で増加するだけであるという事実で表現される特別な性質を持つある関数の存在が導かれます。

このような関数は、Lyapunovの研究によって始まった「安定性理論」において重要な役割を果たします。このため、これらの関数はLyapunov関数と呼ばれています。

エントロピー $S$ は、孤立した系のLyapunov関数です。HelmholtzやGibbsの自由エネルギーといった熱力学的ポテンシャルも、ある「制約条件」のもとで、Lyapunov関数になります。



# 「局所平衡仮説」の導入と 不可逆過程の熱力学の基本式

セミナー「創発を考える」  
Part 1 プリゴジンが考えたこと 3

# 「アトラクター」としての平衡状態

エントロピー $S$ は孤立系に対するLyapunov関数です。ヘルムホルツ自由エネルギーやギブス自由エネルギーなどの熱力学ポテンシャルも、また、Lyapunov関数であることは、先のセッションで見てきました。

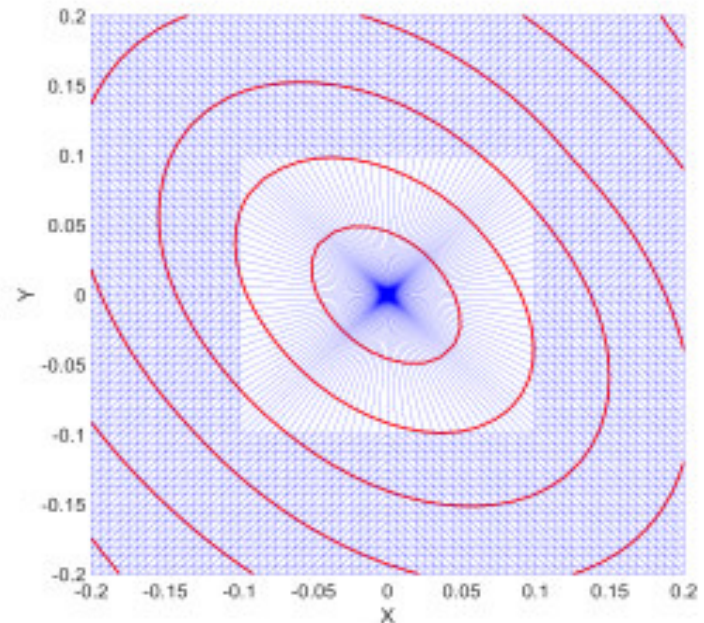
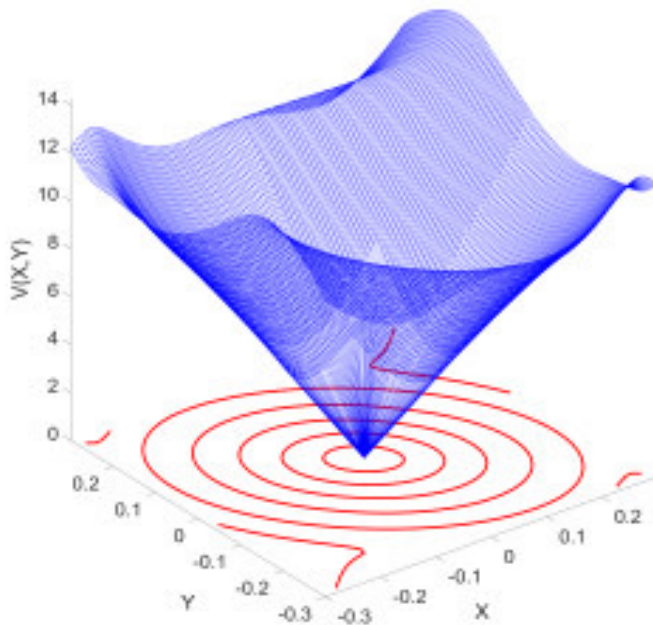
これらすべての場合において、系は熱力学的ポテンシャルの存在を特徴とする平衡状態へと発展します。平衡状態は、非平衡状態がそれへと引き寄せられる「アトラクター」なのです。

# Lyapunov function

力学系とは、あるなめらかな関数  $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$  に対して、 $\dot{y} = f(y)$  で表される系のことです。力学系で定義された連続なスカラー関数  $V: \mathbb{R}^n \rightarrow \mathbb{R}$  が、

$$V > 0 \quad (y \neq 0 \text{ の時})$$

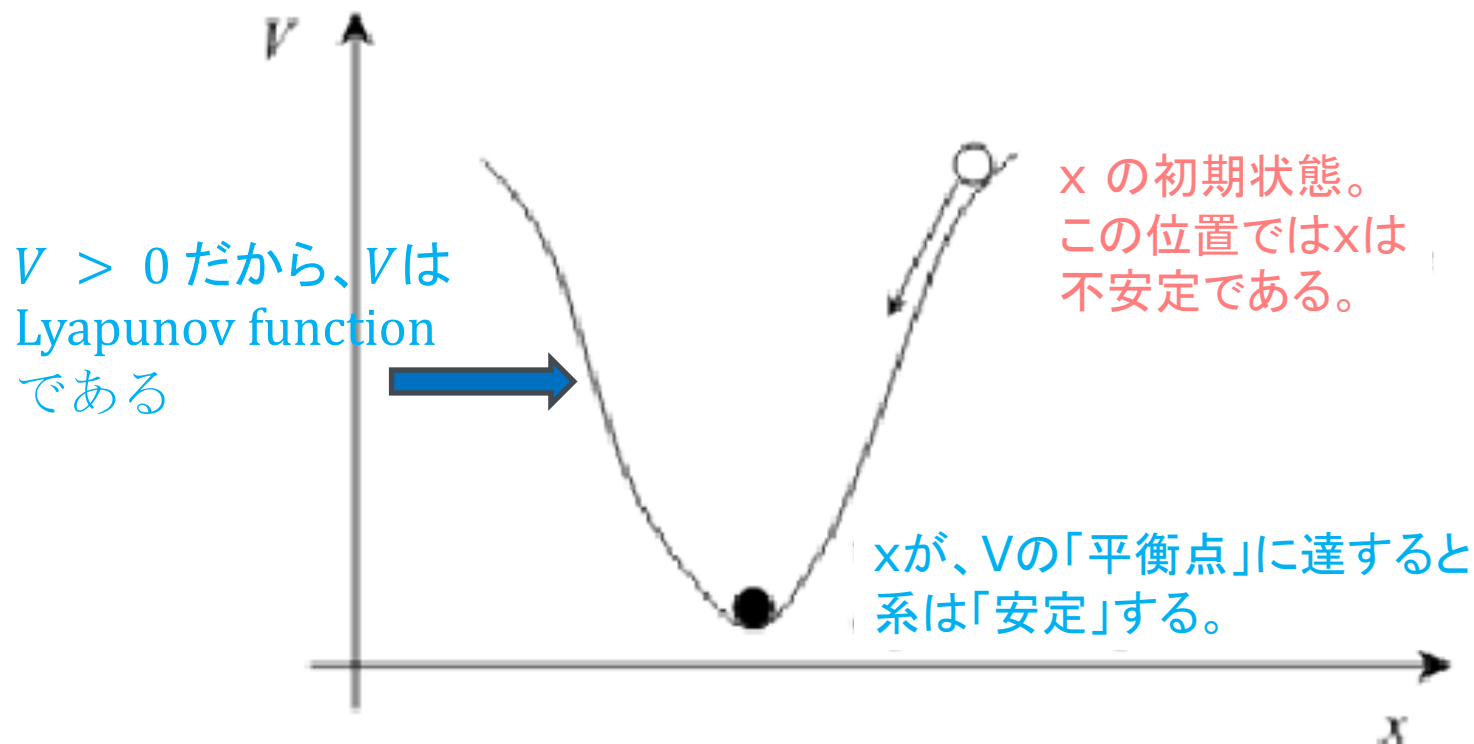
を満たす時、 $V$  を Lyapunov function と呼びます。



# ポテンシャルを表す Lyapunov functionと系の「安定性」

Lyapunov function の定義に出てきた、「連続なスカラー関数  $V: \mathbb{R}^n \rightarrow \mathbb{R}$ 」はポテンシャルの関数として解釈できます。

この図は、 $V =$  ポテンシャル = 位置エネルギーの例です。



## $d_i S \geq 0$ から得られる情報

しかし、熱力学ポテンシャルは例外的な状況にのみ存在します。  
関数の全微分を含まない不等式

$$d_i S \geq 0$$

から、Lyapunov関数を定義することは一般には、できません。

エントロピーの生成を表す式として  $d_i S \geq 0$  を考えていたのですが、この不等式から得られる情報は、平衡状態  $d_i S = 0$  が与える情報に比べても、とても少ないものです。

# エントロピーの生成を明示的に表す式

熱力学第二法則の不等式による定式化は、通常の意味で明確に定義された理論というよりは、(符号を除いて) エントロピーの生成については正確なことは何も語っていないことに注意する必要があります。この不等式の有効範囲さえも特定されていないのです。

これが、熱力学の応用が基本的に平衡過程に限られていた主な理由の一つです。熱力学を非平衡過程に拡張するには、エントロピー生成の明示的な式が必要になります

## 「局所平衡」の仮定の導入と 明示的なエントロピー生成の定式化

この問題は、平衡状態以外でも、エントロピーは平衡状態と同じ変数にのみ依存すると仮定することで解決されることとなります。  
これが「局所」平衡の仮定です。

この仮定が受け入れられると、単位時間当たりのエントロピー生成量 $P$ は次の式で表されることとなります。

$$P = \frac{d_i S}{dt} = \sum_p J_p X_p$$

ここで、 $J_p$ は関与する様々な不可逆過程（化学反応、熱流、拡散...）の流れであり、 $X_p$ は対応する一般化された力（親和力、温度勾配、化学ポテンシャル...）です。

これが不可逆過程のマクロな熱力学の基本式となります。

# 非可逆過程での熱力学平衡

ここで留意すべきことは、このエントロピー生成の明示的な式を導くために、いくつかの仮定を用いたということです。基本的には、この式は、平衡点の近傍においてのみ成立します。

熱力学的平衡では、すべての不可逆的過程 $p$ において、同時に次のようになることが求められます。

$$J_p = 0 \text{ かつ } X_p = 0$$

## 不可逆過程の線形熱力学

したがって、少なくとも平衡付近では、流れ $J_p$ と力 $X_p$ の間に $L_{pp'}$ を係数の行列として線形の均質な関係があると仮定するのはごく自然なことです。

$$J_p = \sum_{p'} L_{pp'} X_{p'}$$

このようなスキームには、熱の流れが温度勾配に比例することを表すフーリエの法則や、拡散の流れが濃度勾配に比例することを表すフィックの法則のような経験則が自動的に含まれています。

このようにして、この関係で特徴づけられる不可逆過程の線形熱力学が得られることとなります。



# 不可逆過程の線形熱力学の 二つの基本定理と「散逸構造」

セミナー「創発について考える」  
Part 1 : プリゴジンが考えたこと 4

# 不可逆過程の線形熱力学の 二つの基本定理と「散逸構造」

今回のセッションでは、不可逆過程の線形熱力学の発展の転換点になった二つの定理を紹介します。一つは「Onsagerの相互律」で、もう一つは Prigogineの「最小エントロピー生成原理」です。

また、セッションの最後では、身近なところに現れる「散逸構造」を紹介します。

# ふりかえり

- 単位時間当たりのエントロピー生成量  $P$

$J_p$ を関与する様々な不可逆過程の流れ(化学反応、熱流、拡散...)、 $X_p$ を対応する一般化された力(親和力、温度勾配、化学ポテンシャル...)とした時、

$$P = \frac{d_i S}{dt} = \sum_p J_p X_p$$

- 流れと力の線形性の条件

線形性を表す係数(行列)を $L_{pp'}$ とする時

$$J_p = \sum_{p'} L_{pp'} X_{p'}$$

- 平衡条件

$$J_p = 0 \text{ かつ } X_p = 0$$

## 二つの不可逆過程の同時進行

熱伝導だけ考えれば、熱は高いものから低いものに流れます。物質の拡散だけを考えても、その式は簡単に求められます。

問題は、二つの過程が同時に進行する場合です。

例えば、冷たい角砂糖を熱いお茶につけて溶かす時、熱伝導 $p$ と拡散 $q$ の二つの過程が同時に進行します。

# Onsagerの相互律

不可逆的な熱伝導を $p$ , 不可逆的な拡散を $q$  とすると、線形性の条件は次の二つの式で表されます。

$$J_p = L_{pp}X_p + L_{pq}X_q$$

$$J_q = L_{qp}X_p + L_{qq}X_q$$

Onsagerの相互律は、次のことを主張しています。

不可逆過程 $p$ に対応する流れ $J_p$ が、不可逆過程 $q$ の力 $X_q$ の影響を受けるとき、流れ $J_q$ も同じ係数を通して力 $X_p$ の影響を受ける。

この時

$$L_{pq} = L_{qp}$$

# Onsagerの相互律の重要性

この関係の重要性は、その一般性にあります。この関係は多くの実験でその妥当性が試されてきました。

この定理は、非平衡線形熱力学が平衡熱力学と同様に、特定の分子モデルに依存しない一般的な結果を導くことを初めて示したものです。

# 最小エントロピー生成原理

Prigoginの「最小エントロピー生成の原理」とは、不可逆的過程の定常状態、すなわち熱力学的変数が時間から独立した状態は、エントロピー生成率の最小値によって特徴づけられるということを主張します。

同じ境界条件に対応する、熱力学的変数が時間に依存した状態ではより高いエントロピーを生成するのに対して、時間から独立した、平衡に十分に近い定常状態では、エントロピー生産は最小になるということです。

## 「最小散逸」の状態での安定

最小エントロピー生成の定理は、非平衡系の一種の「慣性」的な特性を表現しています。与えられた境界条件によって系が熱力学的平衡（つまりエントロピー生成がゼロの場合）に達しない場合でも、系は「最小散逸」の状態で落ち着きます。

エントロピーの最小生成の定理によれば、平衡付近の厳密に線形な領域では、エントロピー生成はリアプノフ関数であらわされます。系に摂動が加わるとエントロピー生成は増大するのですが、系はエントロピー生成の最小値に戻ることによって反応します。

線形非平衡熱力学におけるエントロピー生成の最小値に対応する定常状態は自動的に安定なものになります。

## お味噌汁と「ベナール対流」

この定理は、厳密には平衡付近でのみ有効であることは明らかです。この定理を平衡から離れた状況に一般化するためにさまざまな努力が行われてきました。

そうした取り組みの中で、平衡から遠く離れたところでの、熱力学的挙動が全く異なる挙動を、不可逆過程の非平衡熱力学で説明することができるようになります。

その例が、「ベナール対流」です。ベナール対流は、昔からよく知られた現象で、上下で温度差のある容器に入れられた液体の対流です。この対流は、特徴的なパターンを形成します。熱いお味噌汁をそのままにしておくと、その表面にベナール対流の作るパターンが見えます。

# マクロなパターン形成

味噌汁の中の温度差による対流によって、まず、エントロピーの生成が増大します。

ただ、それは、味噌汁の状態を不安定なカオスにするわけではありません。その先には、静止状態に比べて組織化されたパターンが現れます。

こうしたパターンを実現するためには、巨視的な数の分子が巨視的な時間にわたってコヒーレントに動く必要があります。

「ベナール対流」が形成するパターン  
味噌汁の中の「散逸構造」



## 味噌汁の中の「散逸構造」

ベナール対流の場合、平均的な状態からの揺らぎとして常に小さな対流が現れていると想像することはできるのですが、温度勾配のある臨界値以下では、この揺らぎは減衰して消失します。

しかし、ある臨界値を超えると、ある種のゆらぎが増幅され、巨視的な流れが生じます。新しい超分子秩序が現れ、それは基本的に、外界とのエネルギー交換によって安定化された巨大なゆらぎに対応します。

これが味噌汁の中の「散逸構造」の発生によって特徴づけられる秩序です。



# 機械と散逸構造の違い

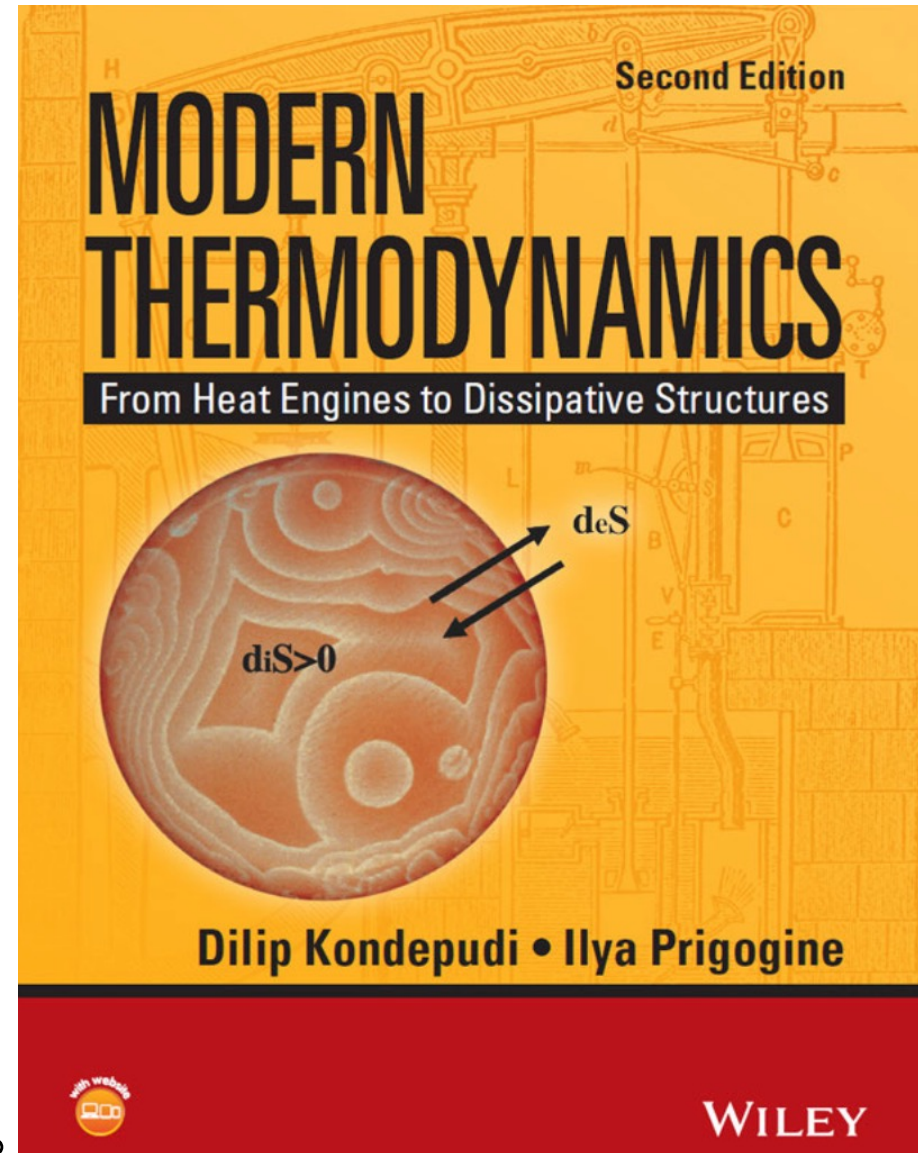
セミナー「創発について考える」  
Part 1：プリゴジンが考えたこと 5

# 現代熱力学

今回のセッションでは、プリゴジンの「機械と生物」との違いについての議論を紹介します。

依拠したのは、プリゴジンの死後(2003年)に第二版が出た、“Modern Thermodynamics -- From Heat Engines to Dissipative Structures”です。

第一版は、1998年出版されました。



# 生物の散逸構造と機械

プリゴジンの「散逸構造論」は、オープンな系での「自己組織」構造の自発的な発生を非平衡の熱力学で基礎付けたものです。彼の散逸構造論は、分子から宇宙の構造まで、非常に広い範囲の現象に適応可能です。

この本の、「生物の熱力学的理論に向けて」という節は、生物に焦点を合わせて、生物の生み出す散逸構造を論じたものです。この中で、彼は生物と機械との区別を論じています。

ここでの彼の議論は、「機械自身の運動からは、散逸構造のような新しい構造は生まれません」と言っているのだと思います。それは正しいのだと思います。

# 機械と生物

機械の組織、その構造と機能は、機械の外部のプロセスから生まれる。そのため、これまで遭遇したことのない機械に遭遇したとき、それを分解し、その構成部品がどのように作られ、配置されているかを分析することで、私たちは機械を作る方法を得ることができる。

生物はそうではない。私たちは100年以上にわたって、生きた細胞の構造と組織について分析を行ってきた。分子レベルに至るまで、その構成については膨大な知識を持っているにもかかわらず、生きた細胞の作り方や、それに漠然と似た細胞の作り方すら知らないのだ。

# 機械と生物になぜこのような違いがあるのか？

機械と生物になぜこのような違いがあるのか？ それは、生物を生み出すプロセスが、私たちのテクノロジーの一部ではないからである。

生物は、不可逆的な内部プロセスによってもたらされる自己組織化に基づいている。機械の組織は、機械の外部にあるプロセスからもたらされ、理想的な機械は可逆的である。

私たちはまだ、不可逆的過程とそこから生まれる自己組織化に基づく技術を開発していない。50年以上前に始まった散逸構造の研究は、生物を生み出すプロセスについての洞察をまだ得られていない。それは未来への挑戦である。私たちはどのようにしてこの課題に立ち向かうのだろうか？

# 機械と散逸構造の違い

概念的な枠組みを構築するために、まず散逸構造と機械の違いを深く調べることから始めよう。次の表に相違点の一覧を示す。

## 設計された構造 (機械／コンピューター)

- 力学の可逆法則に基づく
- システムの外部プロセスを通じて設計され、組み立てられる構造体
- 散逸的なプロセスは、システムの効率を制限する。
- 構造と機能が明確に区別された設計
- ある機能を果たすために設計された構造
- 一般に自己修復性はない

## 散逸構造 (非平衡系と生物)

- 不可逆過程と熱力学の法則に基づく
- 構造はエントロピーを発生させる散逸過程を通じて自発的に生じる
- 散逸過程は系にとって不可欠であり、それがなければ構造は消滅する。
- 構造と機能は常に明確に区別することはできない。
- 構造があるからこそ、文脈に依存した機能が生まれる。
- 自己修復

# 生物の「目的指向的進化」

これらの違いに加えて、生物では、構造と機能の両方が、餌を求めたり悪条件を避けたりといった、特定の目的に向かっていていると思われる行動を引き起こしている。生物は環境と相互作用し、特定の目的に到達できるように反応する。

熱力学的な観点からは、これを「目的指向的進化 end-directed evolution」と解釈することができる。これは、システムがいくつかの異なる経路をとりながらも、最終的には同じ状態に到達することを意味する。

# 生物の「目的指向的進化」

結局のところ、第二法則によれば、系はエントロピーが最大、あるいはギブスエネルギーやヘルムホルツエネルギーが最小の状態へと進化する。系が最終的な極限状態に到達するメカニズムや経路はわからないかもしれないが、その状態に到達することは確かである。この場合、最終状態とはエントロピーの生成がない平衡状態である。

目的指向的進化もまた、平衡から遠い系の性質である可能性がある。しかし、これまでの散逸構造の研究は、そのほとんどが構造と構造を生成し維持するプロセスについて行われてきた。散逸構造における目的指向的進化はあまり研究されていない。

# 進化が向かう状態

目的指向的進化論には、進化が向かう状態とその状態の安定性という2つの側面がある。

前者については、非平衡系はエントロピー生成率が最大となる状態へと進化するという仮説が立てられている。この仮説によると、生物に見られる多様な行動や生命の進化は、エントロピー生成率が増加し続ける状態へと自然が進化した結果であるということになる。

先に、系に新たな触媒が出現することによって生じる構造の不安定性について述べた。一般に、ある系に触媒が入り込むと、対応する親和力が著しく低下していない限り、反応速度を増加させることによってエントロピー生成速度を増加させる。このような観察結果は、この仮説と一致している。

## 進化が向かう状態

しかし、最大エントロピー生産仮説は多くの議論の対象となっている。多くの研究がその妥当性を示している一方で、そうでないことを示す研究もある。実際、平衡に近い線形領域では、エントロピー生成率は最大ではなく最小であることがわかっている。

しかし、この仮説は、平衡からかけ離れた定常状態を予測するのに有用なツールである。おそらく、この仮説が有効であるために満たすべき一連の条件がまだ特定されていないのであろう。また、研究される系によって異なる原理、あるいは複数の原理が存在するのかもしれない。今後の研究が、この考え方に光を当てることになるだろう。

# 最終状態の安定性

第二の側面、最終状態の安定性に目を向けよう。

これは、摂動が生じたときに、システムが最終状態に戻るための経路を見つけることを意味する。言い換えれば、環境の変化によって最終状態が乱された場合、システムは最終状態を回復する手段を見つける。システムが最終状態に到達するためにとりうる経路の多様性は、生物に見られるような興味深い行動につながるかもしれない。

目的指向的進化は、「餌」の供給源が変化したときの生物の反応に似ている。確かに、生物はもっと複雑である。それでも、散逸構造を研究することで、生物の行動の基礎となる熱力学的原理が得られるだろう。

## 機能的対称性の破れ

集団行動のような生物の特徴は他にもいくつかあるが、機能分化は、互いに作用し合うもっと単純な散逸構造にも現れる可能性がある。ここで、機能的対称性の破れという概念を導入しよう。

非平衡系において、互いに相互作用する $N$ 個の同一要素を考える。相互作用が弱い場合、例えば $N$ が小さい場合、 $N$ 個の要素の振る舞いは同一であるか、あるいは機能的に対称である、つまり統計的に同一の性質を持つと予想される。

相互作用の強さが増すと、例えば $N$ が増加すると、系は遷移点に達し、そこ以上では各要素はもはや同一の挙動を示さなくなり、機能的対称性が破れ、明らかに異なる挙動を示す協調的挙動を示すようになるかもしれない。このような遷移は、一般的な極限化原理にも合致するかもしれない。

# 生物の熱力学的理論

上記の概念や例は、生物の熱力学的理論の氷山の一角に過ぎないことは明らかである。

しかし、私たちが生物に見られる行動の起源について熱力学的理論を構築する上で、**目的指向的進化、熱力学的安定性、機能的対称性の破れ**といった概念の可能性をそこに見ることができる。





## Part 2 : ジェインズが考えたこと

## Part 2 : ジェインズが考えたこと

### 最大エントロピー原理とベイズ推論 入門

- 推計から推論へ
- Gibbsの方法とMAXENT
- 「統計的推論」の新しい見方
- 相対エントロピーとベイズ推論



「最大エントロピー原理」とは何か？  
-- 推計から推論へ

セミナー「創発について考える」  
Part 2 : ジェインズが考えたこと 1

# 統計学的「推計」と認識論的「推論」

このセッションでは、Jaynesの「最大エントロピー原理」を紹介し  
ます。

サブタイトルの「推計から推論へ」の「推計」も「推論」も、英語では  
“inference” ですので、サブこのタイトルを英語に翻訳すると、奇  
妙なことになります。

ただ、このサブタイトルで示したかったことは、Jaynesの関心は、  
統計学の「推定」の手法に限られていたわけではなく、もっと広い  
人間あるいは生物の認識の方法としての「推論」にあったのでは  
ということです。

## これらを結びつけるものは何か？

統計学の手法としての「推定」と、認識の機能としての「推論」は異なるものです。

また、熱力学から生まれた「エントロピー」概念は、統計的「推計」手法とも、認識の「推論」能力とも、直接には関係がないように見えます。

あとで述べるように、Jaynesの「最大エントロピー原理」は、一見するとバラバラに見えるこれらを一つに束ねる強力な原理なのですが、これらを結びつけるものは、何なのでしょう？。

# 熱力学的エントロピーと情報論的エントロピー

それは、熱力学的エントロピーと情報論的エントロピーとが、同一のものであるという認識です。

1940年、それまでの熱力学的エントロピー概念の系譜とは全く独立に、Shannonが発見した「不確かさの関数」=「情報量」の概念に、「エントロピー」という名前をつけることを勧めたのは、Von Neumannだと言われています。

<https://www.eoht.info/page/Neumann-Shannon%20anecdote>

[https://en.wikipedia.org/wiki/Talk%3AHistory\\_of\\_entropy](https://en.wikipedia.org/wiki/Talk%3AHistory_of_entropy)

# 熱力学的エントロピーと 情報論的エントロピーの同一性

その時、その理由として、Von NeumannはShannonに次のように言ったそうです。

「第一に、その関数は、すでに熱力学ではエントロピーという名前で使われている。第二に、どうせ誰もエントロピーがどんなものかわかっていないんだから。もしも、君がエントロピーという言葉を使えば、いつだって議論に勝てるよ。」

Jayensの、大きな仕事の一つは、当時、Von Neumannぐらいしか明確な認識を持てなかった、二つのエントロピー概念の同一性を明確に示したことです。

(先の'anecdote'の記事の中にもJaynesの名前が出ています。)

# “Information Theory and Statistical Mechanics” Jaynes「最大エントロピー原理」の提唱 1957年

Jaynesが、「最大エントロピー原理」を提唱したのは、1957年の二つの論文、「情報理論と統計力学」においてでした。簡単にその概要を述べてみたいと思います。

彼は、統計力学的エントロピーではなく、情報理論とそのエントロピーの役割に注目します。

「情報理論は、部分的な知識の上に、(新しい知識の)確率分布を設定する構成的な基準を提供する。そして、それは最大エントロピー評価という統計的推論の一つの型を導く。それは、与えられた情報についての最もバイアスの少ない評価である。すなわち、それは(部分的知識では)失われた情報に関しては、あいまいなものとして最大限コミットしないという評価である。」

# 「最大エントロピー原理」とは何か？

Jaynesの基本的な主張は、情報理論が、ある「部分的な知識」から「新しい知識」を「推論」する枠組みを提供するというものです。

この出発点としての「部分的な知識」は、「与えられた情報についての最もバイアスの少ない」情報として特徴づけられています。こうした特徴づけは、統計力学的特徴というよりは、情報理論的、もっと強く言えば、認識論的特徴です。

元の「部分的知識」の情報理論的エントロピーが最大である時、そのことから、「新しい知識」を推論する「統計的推論」のルールを導くことができると彼は言います。彼は、それを「最大エントロピー原理」と呼びます。

## まだ、ここでは説明されていないこと

統計力学的エントロピーと情報理論的エントロピーが、基本的に同一であることが議論の出発点でした。ただ、エントロピーを情報理論の言葉である「情報」あるいは「知識」で置き換えることで、先のような議論が可能になるのはイメージとしてはわかると思いますが、まだ釈然としないことがあると思います。

例えば、ここで導かれた「統計的推論」というのは、客観的な物理学としての統計力学に根拠を持つことになるのでしょうか？ あるいは、こうした議論の物理学的側面と統計学的側面を明確に区別することはできるのでしょうか？

Jaynesの「最大エントロピー原理」は、これらの疑問にも答を用意しています。それについては、次の機会に触れようと思います。



# Gibbsの方法とMAXENT

セミナー「創発について考える」  
Part 2 : ジェインズが考えたこと 2

## 「最大エントロピー原理」に先行したもの

Jaynesの「最大エントロピー原理」の提唱には、先行した理論的なモデルがあります。もっともそれは、「最大エントロピー原理」のように「推論の原理」として提案されたものではありません。

Jaynesにインスピレーションをあたえたのは、統計力学で行われたある巧妙な計算の方法でした。この計算を行ったのは、Gibbsです。

「統計力学的エントロピー = 情報理論的エントロピー」という同一性に基づき、彼はこの強力な統計力学的導出に、情報理論的推論という解釈を与えたのです。

ここでは、Gibbsが行った計算の方法を見てみましょう。

# Gibbsの方法の不思議なパワー

Gibbsの方法は、不思議なパワーを持っています。

たとえば、ある系のエネルギーについて考えてみましょう。我々は、ある系あるいはその部分系がどのようなエネルギーをもつかを、観測によって知ることができます。

しかし、そのように観測されたエネルギーから、その系がどのようなエネルギー分布に従っているかを知ることが、一見すると不可能のように見えます。

# 我々が知っていること

我々が知っていることは、系の平均的エネルギーが  $E$  であるということだけです。系の確率分布を  $p_i$  とすると、この条件は次の式で表すことができます。

$$\sum_i p_i E_i = E$$

もう一つ、我々が知っていることがあります。それは、

$$\sum_i p_i = 1$$

だという自明の関係です。

# 系のエントロピーSを考える

Gibbsは、欠けている情報を、利用できる他の情報で補おうとします。

彼は、系のエントロピーSを考えます。

$$S = - \sum_i p_i \log p_i$$

これと、先の式を合わせて、次のような式をたてます。

$$-S = \sum_i p_i \log p_i + \alpha \left[ \sum_i p_i - 1 \right] + \beta \left[ \sum_i p_i E_i - E \right]$$

知っている情報  
ただしこの項は  
ゼロである

知っている情報  
ただしこの項は  
ゼロである

# 系のエントロピーが最大になる という条件を追加する

Gibbsが追加した条件は、「系のエントロピーは最大になる」というものでした、

この条件は、次の式で表せます。

$$\frac{\partial(-S)}{\partial p_i} = 0$$

この式から、彼は系の確率分布 $p_i$ の式の形を求めることに成功します。

計算

## ラグランジェの未定乗数法

$-S$ は、 $p_i, \alpha, \beta$ についての関数である。 $-S$ が $p_i$ について最小値をとるのは、次の条件を満たす時である。

$$\frac{\partial(-S)}{\partial p_i} = 0$$

$$\frac{\partial(-S)}{\partial p_i} = \frac{\partial}{\partial p_i} \left\langle \sum_i p_i \log p_i + \alpha \left[ \sum_i p_i - 1 \right] + \beta \left[ \sum_i p_i E_i - E \right] \right\rangle$$

最初の項は、 $\frac{\partial x \log x}{\partial x} = \log x + 1$  を使うと  $\log p_i + 1$  に、

二番目の項は  $\alpha$  に、三番目の項は  $\beta E_i$  になる。

$$\frac{\partial(-S)}{\partial p_i} = \log p_i + 1 + \alpha + \beta E_i = 0$$

これから、 $\log p_i + 1 + \alpha + \beta E_i = 0$  がわかる。

## 分布 $p_i$ の式の形

$$\log p_i + 1 + \alpha + \beta E_i = 0$$

$$\log p_i = -(1 + \alpha + \beta E_i)$$

だから、

$$p_i = e^{-(1+\alpha)} e^{-\beta E_i}$$

$Z = e^{1+\alpha}$  とすると、

$$p_i = \frac{1}{Z} e^{-\beta E_i}$$

## Z を求める

これを

$$\sum_i p_i = 1$$

に代入すると、

$$\sum_i \frac{1}{Z} e^{-\beta E_i} = 1$$

よって、

$$\frac{1}{Z} \sum_i e^{-\beta E_i} = 1$$

これから、

$$Z = \sum_i e^{-\beta E_i}$$

# 分配関数 $Z$

$$Z(\beta) = \sum_i e^{-\beta E_i}$$

を、分配関数(partition function)と呼ぶ。

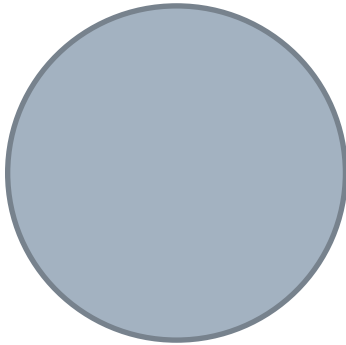
この時、エネルギーの分布 $p_i$ は、次の式で与えられる。

$$p_i = \frac{e^{-\beta E_i}}{\sum_i e^{-\beta E_i}}$$

# 「最大エントロピー原理」

Gibbsの導出を「推論」として解釈する

# Gibbsの方法



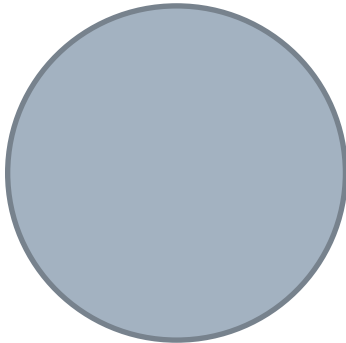
この系のエネルギー  
分布 $p_i$ の形を知りたい

# Gibbsの方法

知っていること

平均的 エネルギー  $\langle E \rangle$

確率  $\sum_i p_i = 1$



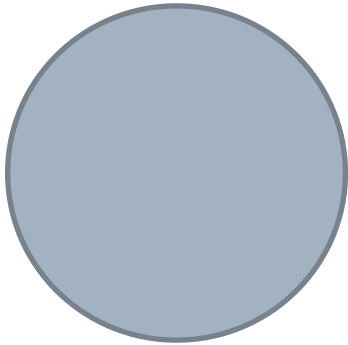
この系のエネルギー  
分布 $p_i$ の形を知りたい

# Gibbsの方法

知っていたこと

平均的 エネルギー  $\langle E \rangle$

確率  $\sum_i p_i = 1$



システムのエントロピー

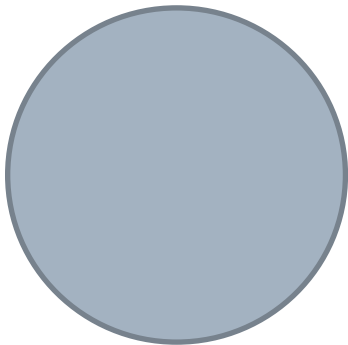
$$S = - \sum_i p_i \log p_i$$

# Gibbsの方法

知っていること

平均的 エネルギー  $\langle E \rangle$

確率  $\sum_i p_i = 1$



システムのエントロピー

$$S = - \sum_i p_i \log p_i$$

ラグランジェの未定乗数法

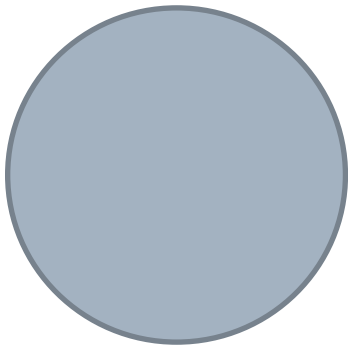
$$-S = \sum_i p_i \log p_i + \alpha \left[ \sum_i p_i - 1 \right] + \beta \left[ \sum_i p_i E_i - E \right]$$

# Gibbsの方法

知っていること

平均的 エネルギー  $\langle E \rangle$

確率  $\sum_i p_i = 1$



システムのエントロピー

$$S = - \sum_i p_i \log p_i$$

ラグランジェの未定乗数法

$$-S = \sum_i p_i \log p_i + \alpha \left[ \sum_i p_i - 1 \right] + \beta \left[ \sum_i p_i E_i - E \right]$$

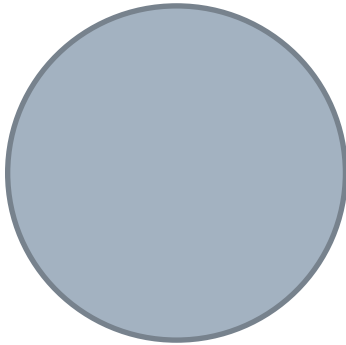
$$\frac{\partial(-S)}{\partial p_i} = 0$$

# Gibbsの方法

知っていたこと

平均的 エネルギー  $\langle E \rangle$

確率  $\sum_i p_i = 1$



システムのエントロピー

$$S = - \sum_i p_i \log p_i$$

$$\frac{\partial(-S)}{\partial p_i} = 0$$

新たに知ったこと

$$p_i = \frac{1}{Z} e^{-\beta E_i}$$

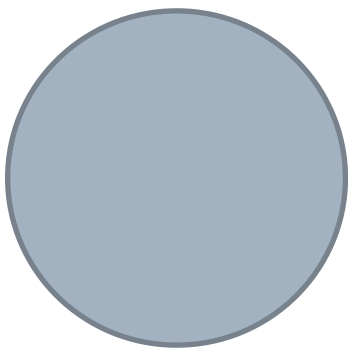
$$Z(\beta) = \sum_i e^{-\beta E_i}$$

# JaynesのMAXENT

事前を知っていたこと

平均的 エネルギー  $\langle E \rangle$

確率  $\sum_i p_i = 1$



システムのエントロピー

$$S = - \sum_i p_i \log p_i$$

認識の発展



$$\frac{\partial(-S)}{\partial p_i} = 0$$

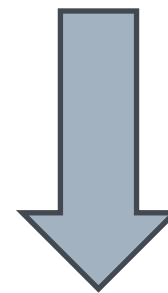
**MAXENT**

事前を知っていたこと

平均的 エネルギー  $\langle E \rangle$

確率  $\sum_i p_i = 1$

推論



事後に知ったこと

$$p_i = \frac{1}{Z} e^{-\beta E_i}$$

$$Z(\beta) = \sum_i e^{-\beta E_i}$$

# 「認識の発展」「推論」の原理として MAXENTを解釈する

MAXENT - 最大エントロピー原理は、ある系に対する現在の認識の状態を最もよく表現する確率分布は、もっともエントロピーの大きいもの、すなわち、もっともその系の情報が少ない確率分布だと考えようということを表しています。

ある系に対する認識が発展するのなら、その認識の発展は、現在の系が含んでいる最少の情報(それは最大のエントロピーを持ちます)からも得られるはずだということです。

それは、「推論」としても解釈できます。



# 「統計的推論」の新しい見方

セミナー「創発について考える」  
Part 2 : ジェインズが考えたこと 3

# 「推論する能力」と創発

僕の興味で言うと、人間は「推論する能力」を持つ存在なのですが、その「推論する能力」は、生物の進化の過程で、ある時点で人類が突然獲得したようにも見えます。

もちろん、人間の「推論する能力」は、さまざまな過程が複雑に入り組んだマクロな現象です。もしも、そのマクロな現象の基礎に、ミクロなレベルでの必然性が働いているとしたら、我々の「推論する能力」の捉え方は、どのように変わるのでしょうか？

Jaynesの議論は、その点ではとても示唆的です。後で見るように、Fristonも、こうした議論から大きな影響を受けています。

# Jaynesが「統計的推論」について考えたこと

このセッションでは、Jaynesが「最大エントロピー推論」という考えに至るまで、彼が「統計的推論」について考えたことを、もう少し追いかけてみようと思います。

## 時代背景：統計力学の到達点

「最近、古典力学や量子力学の観点から統計力学の方法を正当化しようとした過去の試みを非常に包括的に調査したものが出版され、この分野の未解決の問題を強調する上で、非常に好都合な時期に大いに役立っている。」

「このテーマの研究は、何年ものあいだ取り組まれてきたにもかかわらず、物理学者にあらゆる面で確信を与えるような、ミクロな力学の法則からマクロな現象に進む議論の道筋がないという意味で、我々はいまだ、完全で満足 of いく理論をもっていないのだ。」

# 統計的推論としての統計力学

「ギブスの統計力学には、古典力学の観点からは理解できないいくつかの困難があり、彼が構築したモデルを観測された事実に対応させる前に、古典力学の法則には含まれていない追加の制約を組み込む必要があった。

しかし、量子力学の発展により、当初は恣意的であった仮定が、現在では物理法則の必然的帰結とみなされるようになった。

このことは、統計力学がもはや物理的仮説に依存することなく、単なる統計的推論の一例となりうる状態に到達した可能性を示唆している。」

# 統計力学が取り組むべき課題

「現在がこれらの問題を再検討する好機であると考えられるのは、2つの最近の進展によるものである。

統計的手法は不可逆的過程を含む様々な具体的現象に適用されつつあるが、成功を収めた数学的手法はまだ統計力学の基本装置に組み込まれていない。

加えて、「情報理論」の発展が統計力学にとって大きな意味を持つと多くの人々が感じているが、その具体的な適用方法は曖昧なままである。」

# 熱力学的エントロピーと 情報理論的エントロピーの関係

「これに関連して、次のことに注意する必要がある。

$$-\sum p_i \log p_i$$

という同じ数式が統計力学と情報理論の両方に現れるという事実だけでは、これらの分野の間に何のつながりもない。

これは、熱力学的エントロピーと情報理論的エントロピーが同じ概念に見えるような新しい視点を見つけることによるのみ可能である。」

# 統計力学の正当化の問題に情報理論を適用する

「この論文では、統計力学の正当化の問題に情報理論を適用できるように、これを達成する統計力学の再解釈を提案する。

我々は、定常状態に割り当てられる確率のみを含む初歩的な取り扱いによって、平衡熱力学的性質の予測に取り組む。

密度行列の使用によって得られる改良や不可逆過程の議論は後の論文で取り上げる。」

## Gibbsの方法との関連

「第2節で示されるようなエントロピーの最大化に関する数学的事実は、はるか昔にギブスによって指摘されていた。

しかし、過去においては、これらの性質は理論にとって本質的なものではなく、それ自体統計力学の方法を正当化するものでもない、副次的なものであった。」

## 推論を逆転する

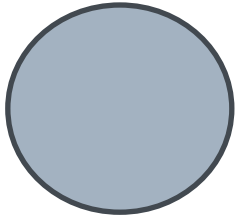
「最近になってシャノンが、エントロピーの表現が熱力学とは全く独立した深い意味を持っていることを証明した。これにより、統計力学における通常の推論を逆転させることが可能になった。

以前は、運動方程式に基づいて理論を構築し、エルゴード性、計量的他律性、あるいは等しい先験的確率といった付加的な仮説を補い、エントロピーの同定は最後に、得られた方程式を現象学的熱力学の法則と比較することによってのみ行われていた。

しかし現在では、エントロピーを出発概念とすることができ、ある確率分布がある制約条件下でエントロピーを最大化するという事実が、その分布を推論に用いることを正当化する本質的な事実となる。」

従来の方法での  
エントロピーの導出

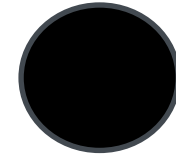
エントロピー



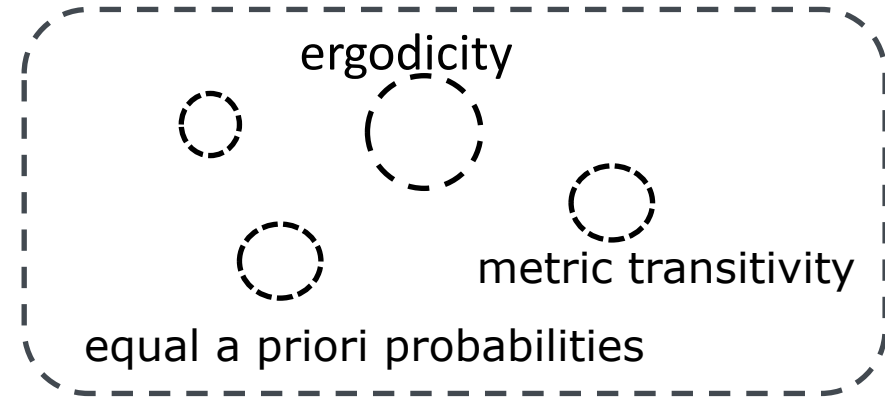
熱力学の法則との比較



運動方程式

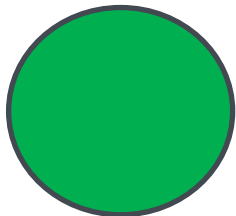


物理的仮説



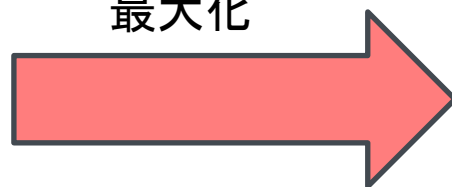
視点の逆転  
MAXENT

エントロピー



$$\langle f_r(x) \rangle = \sum_i p_i f_r(x_i)$$

最大化



$$p_i = \exp\{-[\lambda_0 + \lambda_1 f_1(x_i) + \dots + \lambda_m f_m(x_i)]\}$$

$$\langle f_r(x) \rangle = -\frac{\partial}{\partial \lambda_r} \ln Z,$$

$$\lambda_0 = \ln Z.$$

# 物理的仮説への依存から理論を解放する

「しかし、この視点の逆転がもたらす最も重要な結果は、概念的・数学的な単純化ではない。

上記のような物理的仮説への依存から理論を解放することで、統計力学をより一般的な観点から見ることができるようになる。その原理と数学的方法は、多くの新しい物理的問題を扱うために利用できるようになる。」

# 結論

「上記の議論において重要な点は、フォン・ノイマン・シャノンのエントロピーの式を、文字どおり、確率分布によって表される不確定性の大きさの尺度として受け入れることである。

さらに、統計力学の予測問題を主観的な意味で解釈し直せば、アンサンブルを考慮したり、エルゴード性や等しい先験的確率に関する通常の議論に訴えたりすることなく、非常に初歩的な方法で通常の間係を導くことができる。統計力学の原理と数学的方法は、従来の議論から想像されるよりもはるかに一般的な応用が可能であることがわかる。

予測の問題では、エントロピーの最大化は物理法則の適用ではなく、無意識のうちに恣意的な仮定が導入されていないことを保証する推論方法にすぎない。」

# 相対エントロピーとベイズ推論

セミナー「創発について考える」  
Part 2 : ジェインズが考えたこと 4



Thomas Bayes  
1701~1761



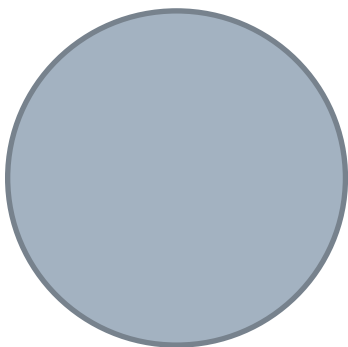
Edwin Thompson Jaynes  
1922~1998

# JaynesのMAXENT

事前を知っていたこと

平均的 エネルギー  $\langle E \rangle$

確率  $\sum_i p_i = 1$



システムのエントロピー

$$S = - \sum_i p_i \log p_i$$

認識の発展



$$\frac{\partial(-S)}{\partial p_i} = 0$$

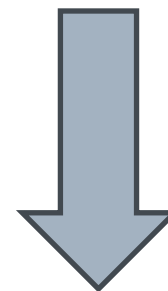
**MAXENT**

事前を知っていたこと

平均的 エネルギー  $\langle E \rangle$

確率  $\sum_i p_i = 1$

推論



事後に知ったこと

$$p_i = \frac{1}{Z} e^{-\beta E_i}$$

$$Z(\beta) = \sum_i e^{-\beta E_i}$$

# 「認識の発展」＝「推論」の 原理としてのMAXENT

MAXENT - 最大エントロピー原理は、ある系に対する現在の認識の状態を最もよく表現する確率分布は、もっともエントロピーの大きいもの、すなわち、もっともその系の情報が少ない確率分布だと考えようということを表しています。

ある系に対する認識が発展するのなら、その認識の発展は、現在の系が含んでいる最少の情報(それは最大のエントロピーを持ちます)からも得られるはずだということです。

それは、私たちの「推論」の原理としても理解できます。  
「推論」は「認識の発展」に他なりません。

# 確率分布とエントロピー

## 力学系のLyapunov関数としてのエントロピー

ある確率分布  $p_i$  が与えられた時、そのエントロピー  $S$  は次の式で与えられます。

$$S = - \sum_i p_i \log p_i$$

$$S > 0$$

エントロピーは、ある力学系のLyapunov関数としての性質を持ちます。

# 確率分布とエントロピー

ただ、どんな確率分布についても、アприオリに一つのエントロピーが先の公式で天下りの的に定まるということに、すこし違和感を持つ人がいるかもしれません。

そもそも、確率分布がアприオリに与えられるものかは、自明ではありません。

# 相対エントロピー

そういう人には、次の「相対的なエントロピー」という考えの方が、納得が行きやすいと思います。

「エントロピー」は、絶対的な確定したものではなく、事前に知っていたこととの関係で決まる、相対的なものだと考えるのです。

事前に知っていた確率分布を $p$ とし、実際に、観測して得られた新しい確率分布  $q$ とします。この時、次の式で確率分布 $p$ に対する確率分布 $q$ の「相対エントロピー」 $H_{\text{rel}}(q, p)$ を定義します。

# 相対エントロピー

確率分布 $p$ に対する確率分布 $q$ の「相対エントロピー」 $H_{rel}(q, p)$ を、次の式で定義します。

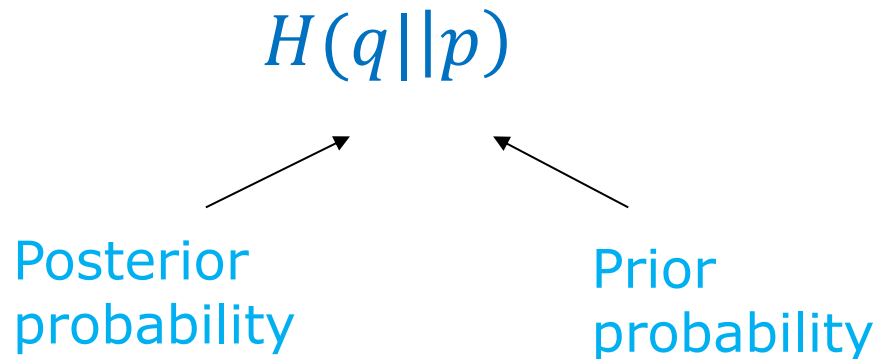
$$H_{rel}(q, p) = \sum_x q(x) \log \frac{q(x)}{p(x)}$$

確率分布 $p$ に対する確率分布 $q$ の「相対エントロピー」 $H_{rel}(q, p)$ を  $H(q(x) || p(x))$  と表すことがあります。

$$H(q || p) = H_{rel}(q, p) = \sum_x q(x) \log \frac{q(x)}{p(x)}$$

# 'Prior' と 'Posterior'

確率分布 $p$ に対する確率分布 $q$ の相対エントロピー  $H(q||p)$  で、  
  $p$  を、「事前確率」 = 'Prior'  
  $q$  を、「事後確率」 = 'Posterior'  
 と呼びます。



# 認識の発展のBayesian的解釈

こうした考え方は、Bayesianのものです。

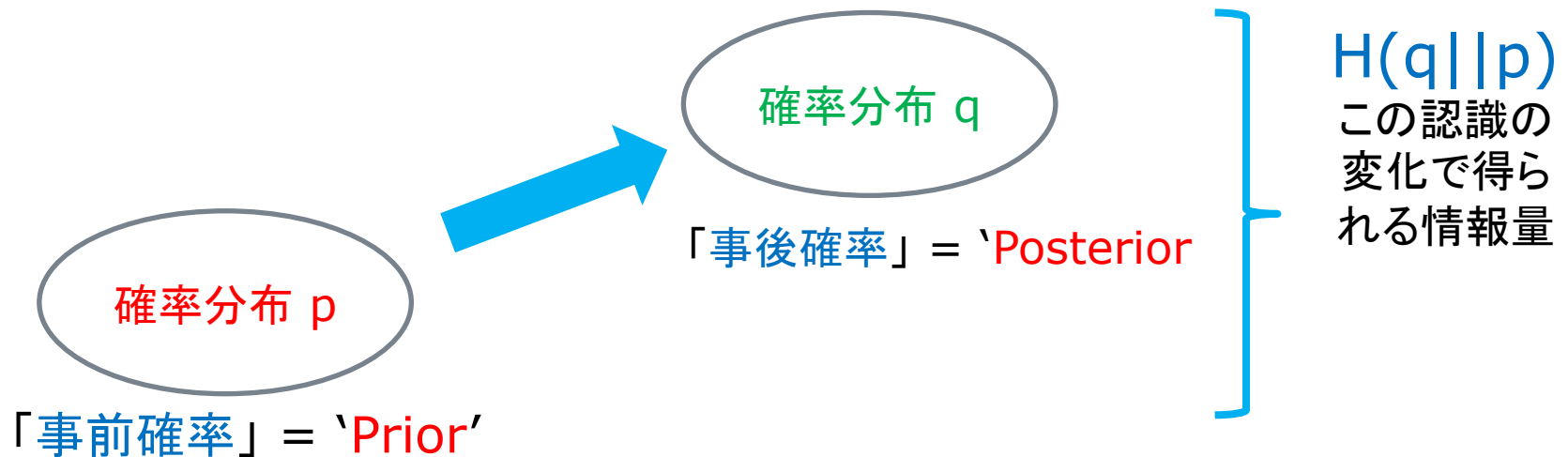
「相対エントロピー」というのは、アприオリな「シャノンのエントロピー」を、Bayesianの考え方で、相対化したエントロピーと考えることができます。

エントロピー＝情報量のこのBayesian的な解釈は、人間の認識で得られる情報量の解釈には、とても向いています。

Prior の  $p$  という仮説的な確率の認識は、Posterior の  $q$  という確率の認識に「発展」したと考えることができるからです。

# 相対エントロピーの直観的意味

直観的に言えば、相対エントロピー  $H(q||p)$  は、あるシステムが確率分布  $p$  に従っているという仮説的認識から出発して(これが 'Prior' です)、その後、そのシステムの「正しい」あるいは「実際」は、確率分布  $q$  に従っていることを学んだ(これが 'Posterior')時に、得られる情報量です。



# コインスを例に相対エントロピーを計算する

**Case 1:**例えば、コインスに使われるコインが、かたよりがなく公正なものだという仮定から出発して、実際に、コインの表がでたとすれば、その相対エントロピーは、 $\log 2$  となって、我々は1bitの情報を得たことになります。

**Case 2:**しかし、コインは常に表が出るという仮説から出発すれば、表が出たとしても、我々の得る情報、すなわち相対エントロピーはゼロになります。

$$H(q||p) = \sum_x q(x) \log \frac{q(x)}{p(x)}$$



$$H(q||p) = q(\text{表}) \log \frac{q(\text{表})}{p(\text{表})} + q(\text{裏}) \log \frac{q(\text{裏})}{p(\text{裏})}$$

# コインスを例に相対エントロピーを計算する

$$H(q||p) = \sum_x q(x) \log \frac{q(x)}{p(x)}$$
$$H(q||p) = q(\text{表}) \log \frac{q(\text{表})}{p(\text{表})} + q(\text{裏}) \log \frac{q(\text{裏})}{p(\text{裏})}$$

## Case 1:

仮定: コインは公正  $p(\text{表})=p(\text{裏})=1/2$ ; 結果:  $q(\text{表})=1, q(\text{裏})=0$

$$H(q||p) = 1 \log \frac{1}{1/2} + 0 \log \frac{0}{1/2} = \log 2 = 1 \text{ bit}$$

## Case 2:

仮定: コインはインチキ  $p(\text{表})=1, p(\text{裏})=0$ ; 結果:  $q(\text{表})=1, q(\text{裏})=0$

$$H(q||p) = 1 \log \frac{1}{1} + 0 \log \frac{0}{0} = 0 \text{ bit}$$

「認識の発展」＝「推論」＝「学習」

相対エントロピーのBayesian的解釈

# 認識の発展: Step $t = 0$

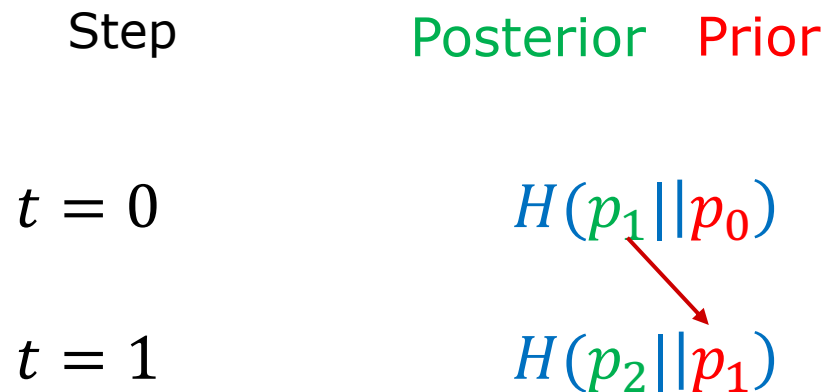
Step                      Posterior    Prior

$$t = 0 \qquad H(p_1 || p_0)$$

認識のある段階で、我々はあるシステムの  $p_0$  で与えられると考えていた確率分布が、本当は確率分布  $p_1$  に従うことを発見したとする。この発見によって得られる情報量は、相対エントロピー  $H(p_1 || p_0)$  によって与えられる。

# 認識の発展: Step $t = 1$

Step	Posterior	Prior
$t = 0$	$H(p_1    p_0)$	
$t = 1$	$H(p_2    p_1)$	



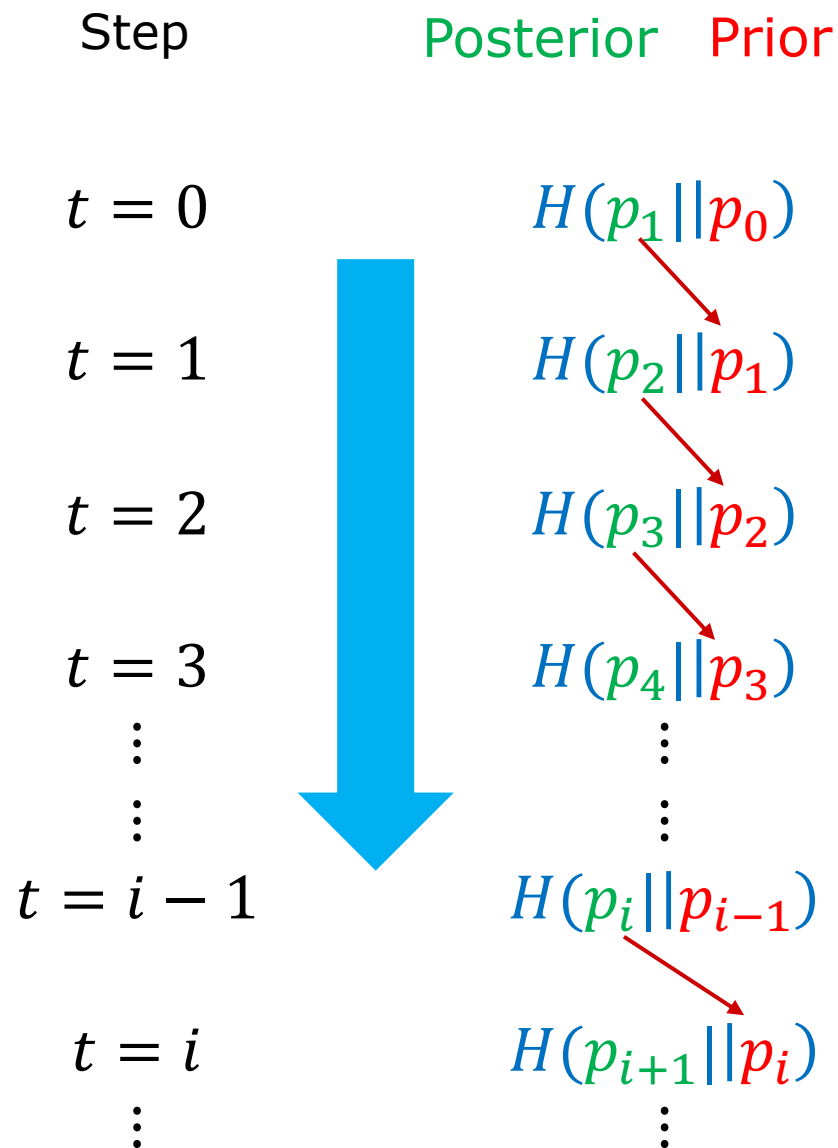
認識の次の段階で、あるシステムの  $p_1$  で与えられると考えていた確率分布が、本当は確率分布  $p_2$  に従うことを発見するかもしれない。この発見によって得られる情報量は、相対エントロピー  $H(p_2 || p_1)$  によって与えられる。

# 認識の発展: Step t = 2

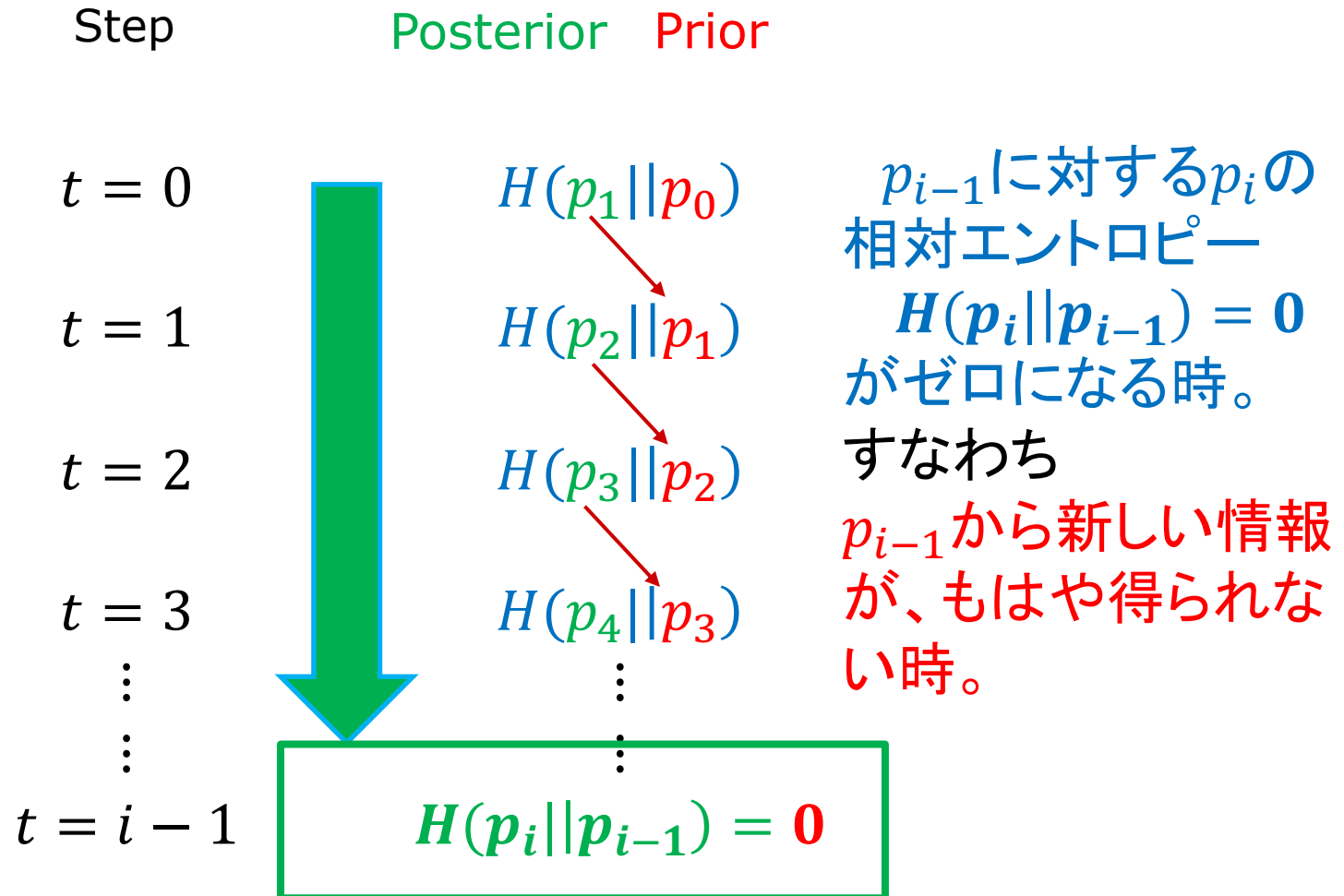
Step	Posterior	Prior
$t = 0$	$H(p_1    p_0)$	
$t = 1$	$H(p_2    p_1)$	
$t = 2$	$H(p_3    p_2)$	

認識の次の段階で、あるシステムの  $p_2$  で与えられると考えていた  
確率分布が、本当は確率分布  $p_3$  に従うことを発見するかもしれない。  
この発見によって得られる情報量は、相対エントロピー  $H(p_3 || p_2)$   
によって与えられる。

# PriorとPosteriorで「認識の発展」を記述する



# この「認識の発展」が終わるのは？



# 「学習」のBayesian的解釈

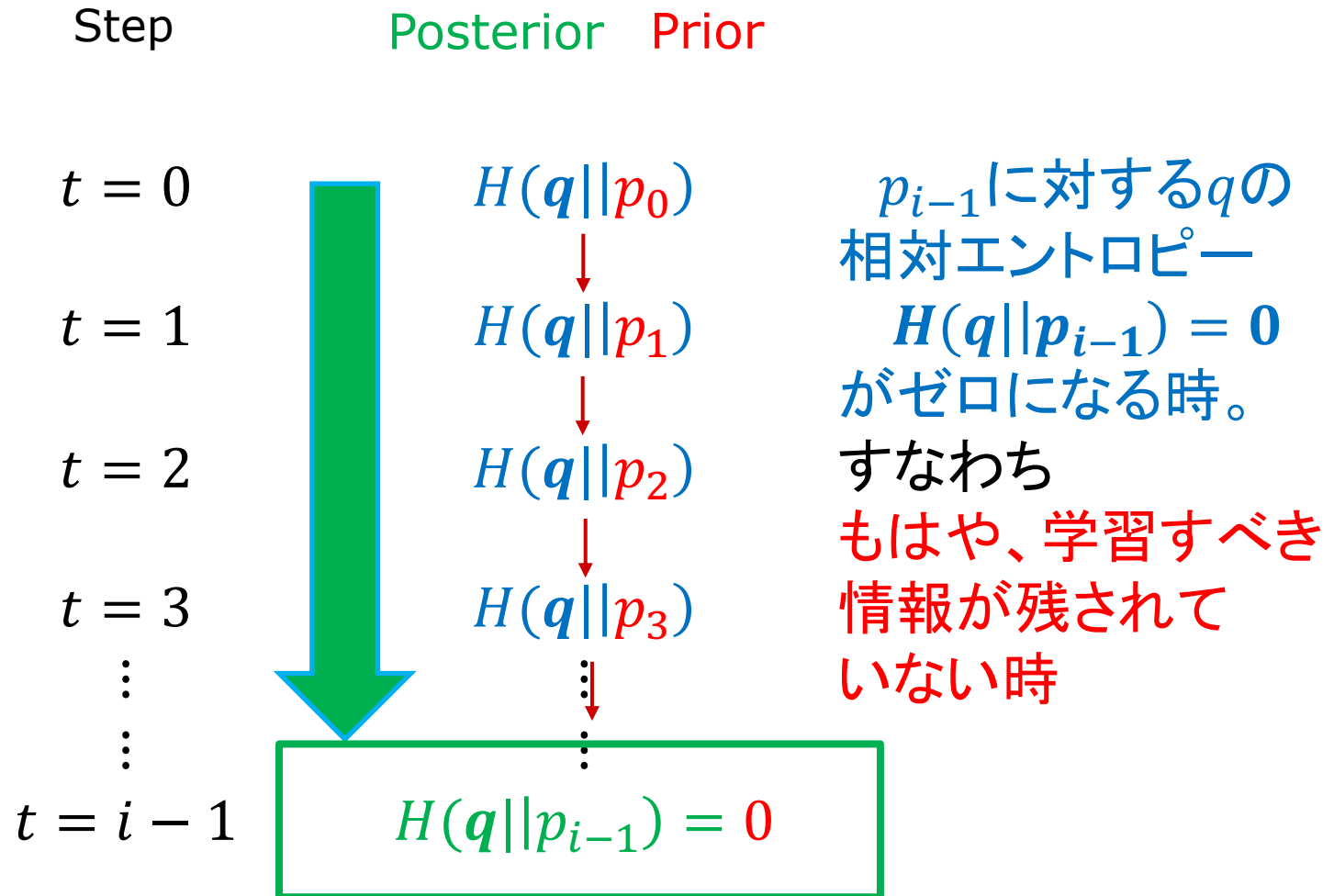
逆に、もしも、最初から正しい分布 $q$ を、何らかの方法で我々が知っていて、実験 $p(t)$ を繰り返すのなら、 $H(q||p(t))$ は、実測値 $p(t)$ から、正しい答え $q$ に至るために「学習しなければいけない情報量」を表すことになります。

$q$ は、「常に正しい」と仮定しているので、それは時間には依存しません。 $q$ は、 $t$ を含まないことに注意してください。

ここでは、 $H(q||p(t)) = 0$ は、「もはや、学習すべき情報が残されていない」ことを意味して、その状態で、学習は終わります。



# この「学習」が終わるのは？



# クロス・エントロピー

ディープラーニングでコスト関数として利用される「クロス・エントロピー」は、こうした「相対エントロピー」の一種です。

「正しい」分布を $q$ 、実測値を $p$ としたとき、クロス・エントロピー  $H_{cross}(q, p)$  は、次の式で定義されます。

$$H_{cross}(q, p) = \sum q_i \log p_i$$

## クロス・エントロピーと相対エントロピー

先の相対エントロピー  $H(q||p)$  の定義から、次のことがわかる。

$$\begin{aligned} H(q||p) &= \sum q_i \log \frac{q_i}{p_i} = \sum q_i \log q_i - \sum q_i \log p_i \\ &= -H(q) - H_{cross}(q, p) \end{aligned}$$

## クロス・エントロピー

個人的には、シャノンのエントロピーの定義も、ぶっきらぼうだと思いますが、クロス・エントロピーの定義の意味は、式の形からはわかりにくいように感じています。

ディープラーニングでの「学習」の直接の目的は、クロス・エントロピーを最小にすることなのですが、相対エントロピーの言葉で言えば、それは、「正しい」認識に至るために、「残された学習すべき情報を最小のものにすること」となります。

こちらの方がわかりやすいように思います。





# Part 3 : フリストンが考えたこと

## Part 3 : フリストンが考えたこと

### 最小自由エネルギー原理と脳のモデル 入門

- 脳研究の動向
- 自由エネルギーと脳
- 拡大された「推論」概念
- こころの時間の数学



# 脳研究の動向

セミナー「創発について考える」  
Part 3 : フリストンが考えたこと 1

# 脳研究の動向

セミナーのこのパートでは、Fristonらの「脳の理論モデル」を紹介しようと思います。

ただ、今回のセッションでは、本論の「脳の理論モデル」の話に入る前に、2014年にEUの“Human Brain Project”で起きた、脳研究の方法をめぐる大きなトラブルと“Human Brain Project”の一時的挫折を振り返ってみようと思います。

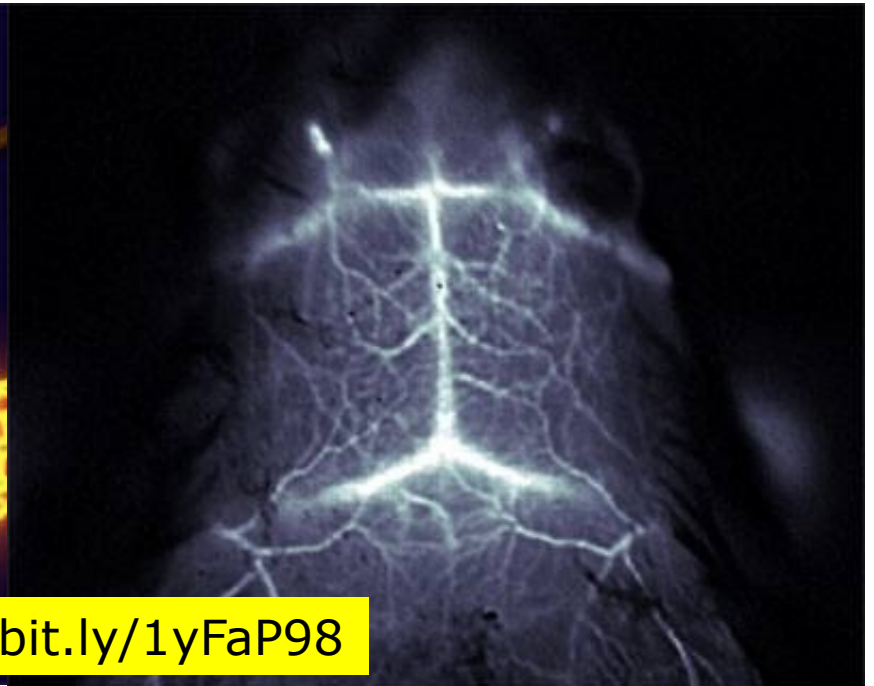
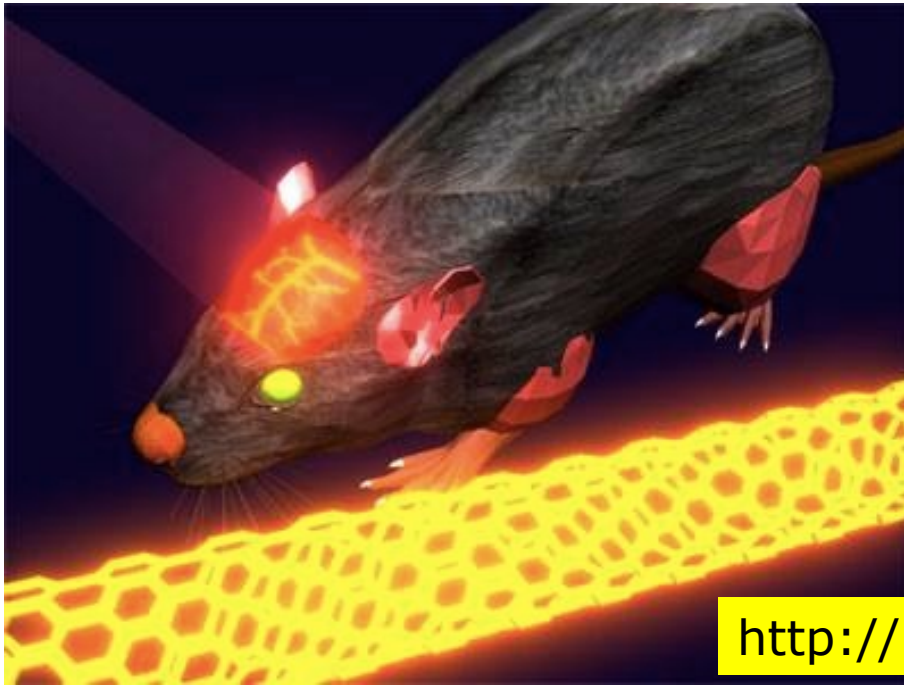
それは、今回取り上げる「脳の理論モデル」への関心の高まりの、背景の一因だと、僕は考えています。

10年前の脳研究を振り返る

# 新しい研究手法への期待 1

## 非侵襲的な脳のモニター技術の進歩

脳の血流を、MRIやCTよりもはるかに正確に、かつリアルタイム(200ms間隔)に、観察する新技術。カーボン・ナノチューブを血液に注入して、それが発する近赤外線領域の蛍光波長を見る。



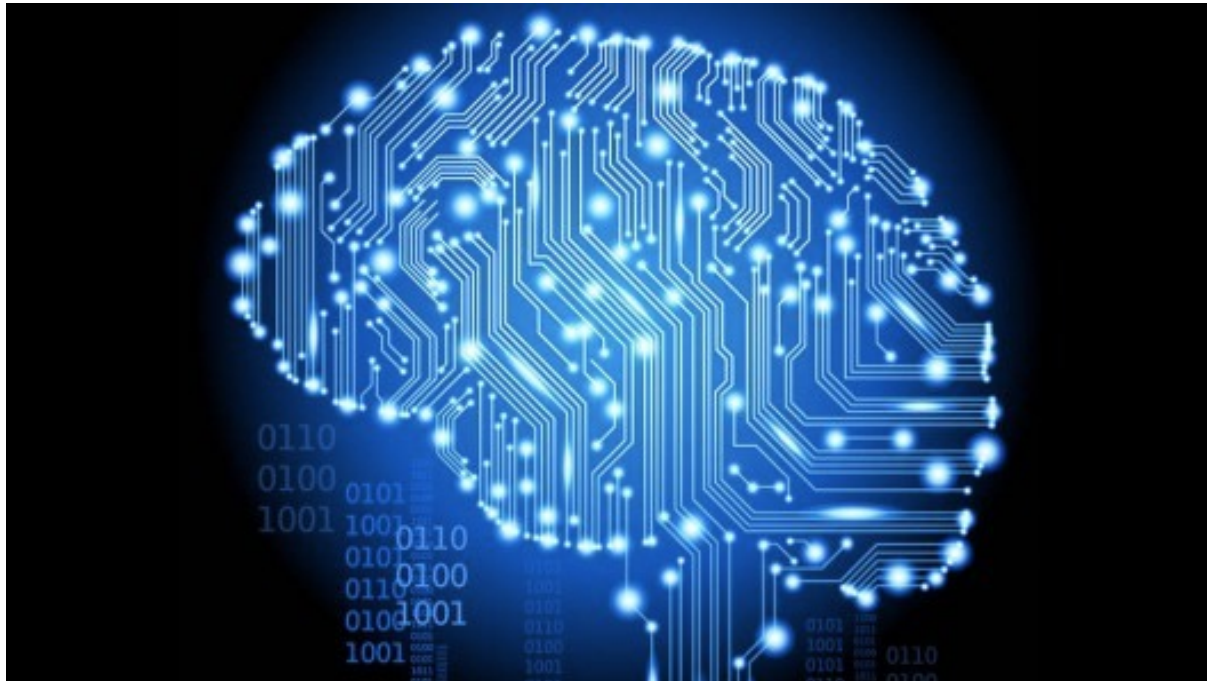
<http://bit.ly/1yFaP98>

## 新しい研究手法への期待 2 DNAシーケンサーの進歩

- 最初のヒトゲノムの解析には、10年と30億ドルかかった。
- 2010年までには、ヒトゲノムの解析は、一ヶ月以内で3万ドルで出来るようになった。
- 今では、ヒトゲノムの解析は、数日で4000ドル以下で出来る。
- DNAシーケンサーのコスト・パフォーマンスは、コンピュータでのMooreの法則よりも速いスピードで向上しているという。個人が、ゲノムを解析する時代が始まろうとしている。

# Blue Brain Project Henry Markram

- “I wanted to **model the brain** because we didn’t understand it.”
- “The best way to figure out how something works is to **try to build it from scratch.**”



# EU Human Brain Project のたちあげ



**2013年10月** 10年間で総額約12億ユーロ

# EU Human Brain Project ビジョン

人間の脳を理解することは、21世紀の科学が直面している最も偉大な挑戦の一つである。もしも、我々が、それに対して立ちあがることが出来るならば、我々は、我々を人間にしているものが何であるかについて深い洞察を得て、革命的なコンピュータ技術を構築し、脳の異常に対して新しい治療法を開発出来るだろう。今日、初めて、現代のICT技術が、こうした目標を到達可能なものになっている。

# EU Human Brain Project

## 中心的な戦略目標

### ● 未来のニューロサイエンス

遺伝子から行動に至る、生物学的器官のすべてのレベルを貫いて、健康な脳と病気の脳についてのデータと知識を統合した、人間の脳の統一された複数レベルでの理解を達成する。こうして、**脳を理解する基本的な方法論として、シリコン上での実験**を確立する。

### ● 未来のコンピューティング

脳の回路とコンピュータの原理に基づいて、**新しいニューロコンピュータとニューロロボットの技術**を開発する。脳のシミュレーション、ロボットと自律的なシステムコントロール、その他のデータを大量に利用するアプリケーションの為にスーパーコンピュータの技術を開発する。

# EU Human Brain Project

## 中心的な戦略目標

### ● 未来の医療

複数レベルの臨床データに基づいて、客観的で生物学的に基礎付けられた、神経症と精神病のマップを開発する。このマップを利用して、脳の病気の分類と診断を行い、これらの病気のモデルを構成する。シリコン上での実験を利用して、脳の病気の原因を理解し、新薬やその他の治療法を開発する。神経症と精神病に対して、個人化された医療を確立する。

# EU Human Brain Project への批判 ECへの公開質問状

## Open message to the European Commission concerning the Human Brain Project

[Sign the letter](#)

*Summary:* Neuroscience advances our understanding of normal and pathological brain function, offering potentially enormous benefits to society. It is, therefore, critical to Europe. The Human Brain Project (HBP), sponsored by the European Commission (EC), was meant to forward this mission. However, due in great part to its narrow focus, it has been highly controversial and divisive within the European neuroscience community and even within the consortium, resulting in on-going losses of members. The HBP is now scheduled for review and we wish to draw the attention of the EC to these problems. We believe the HBP is not a well conceived or implemented project and that it is ill suited to be the centerpiece of European neuroscience. We are particularly concerned about the plan to tie a substantial portion European member states' neuroscience funding to the HBP through so-called 'partnering projects'. We call for the EC to go beyond the strict requirements of the upcoming review, to demand transparency and accountability and, if necessary, change the structure of the HBP's governance and supervision to correct their shortcomings. Failing that we call for the EC to redirect the HBP funding to smaller investigator-driven neuroscience grants. We stand fully behind a strong and united European neuroscience strategy and we pledge not to seek funding through HBP partnering projects that would compromise that mission.

[Official Response by the EC - 18 July 2014](#) | [Press Releases](#)

[Read the full letter](#)

# 脳研究をめぐる「対立」の構図

認知科学 vs 脳科学、トップダウン vs ボトムアップ

- 発端は、来年度のプロジェクトの予算配分で、認知科学的なアプローチの予算が、ばつさりと切られたことにあるらしい。それに反発した研究者のグループが、150名の連名で、欧州委員会に公開質問状を提出した。
- 対立の根底にあるのは、脳研究でのアプローチの違い。Henry Markramら主流派は、ニューロンとシナプスの数学的モデルに基づいて脳全体のモデルを**ボトムアップ**に作り上げようというアプローチ。一方、反対派は、脳研究には、認知科学の知見に基づいた**トップダウン**のリバーズ・エンジニア的なアプローチが必要だという。
- 「鳥の羽の全てをシミュレーションしたとしても、鳥が空を飛べることを解明出来ないのと同じことだ」

# 脳研究の現在

# Human Brain Project (EU)

Learn about the research  
being done by the  
Human Brain Project in  
the brochures.



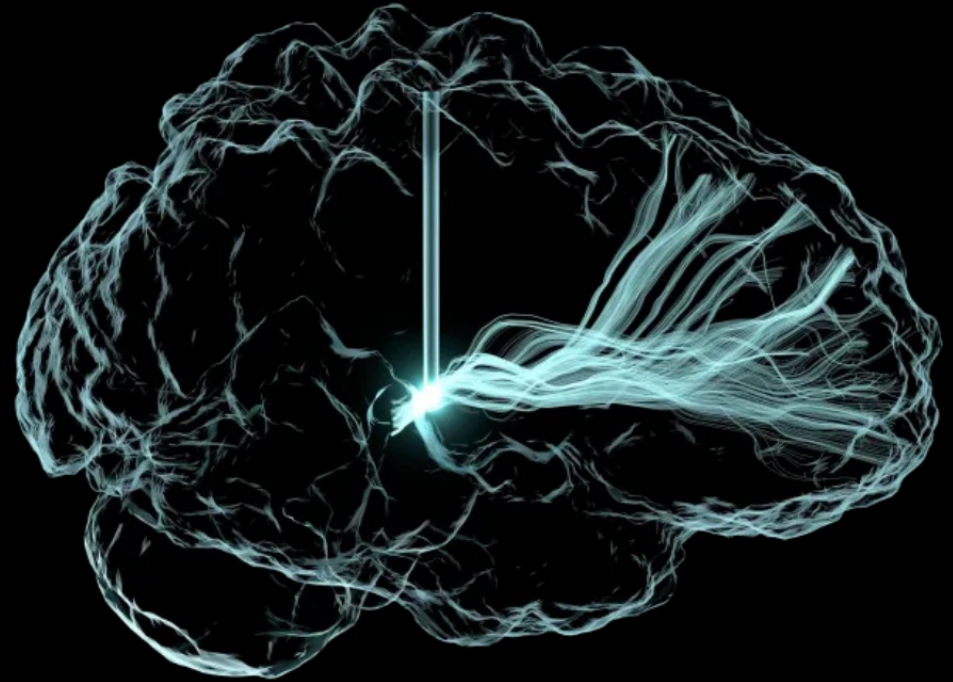
<https://www.humanbrainproject.eu/en/>

# Brain Initiative (US)

## The *Brain Research Through Advancing Innovative Neurotechnologies*<sup>®</sup> (BRAIN) Initiative

Revolutionizing our understanding  
of the human brain

---



<https://braininitiative.nih.gov/>

LEARN MORE

# Simian Symphony

音楽にリアルタイムに反応する海馬のニューロンの映像



[https://www.youtube.com/channel/UC\\_muVZ3XF1aCNtM\\_YfcCyyg/videos](https://www.youtube.com/channel/UC_muVZ3XF1aCNtM_YfcCyyg/videos)

# 脳の理論モデル

Fristonらの「脳の理論モデル」の研究は、Markramのようにコンピュータ上に脳のモデルを作ろうというものでも、それに反対した認知科学者たちのアプローチとも異なるものだ。

それは、「脳のモデル」を構築する上での原理的な方法論を明確にしようという問題意識を、彼らが持っていることだ。Fristonはそれを「最小自由エネルギー原理」と呼ぶ。

全てをそれで説明しようとするのは難しさもあるかもしれないのだが、その立場は、Prigogine や Jaynesといった、20世紀の偉大な達成を忠実に受け継ごうとしているように、僕には見える。



# 自由エネルギーと脳

セミナー「創発について考える」  
Part 3 : フリストンが考えたこと 2

# 自由エネルギーと脳

今回のセッションでは、現在、脳研究の新しいフレームとして注目を集めているFristonらの研究の概要を紹介します。

依拠したのは、彼らの初期の論文 “Free-Energy and the Brain” (2007年) です。

<https://link.springer.com/article/10.1007/s11229-007-9237-y>



1821- 1894  
Hermann von Helmholtz



1959-  
Karl Friston

# Helmholtzの知覚論

## 無意識的推論

ヘルムホルツは、視覚的印象の形成は主に無意識の判断によって達成され、その判断結果は「一度も意識的判断の平面に昇格することができない」ため、「意識的思考の浄化と精査の作業を欠いている」。

にもかかわらず、無意識の判断結果は意識的な制御を受けにくく、矛盾に強いため、「取り除くことは不可能」であり、「その影響を克服することはできない」とした。

# Helmholtzの知覚論

## 無意識的推論

その理由は、視覚的な感覚的印象が神経学的に処理される方法にある、とヘルムホルツは示唆した。意識的な熟考を担う高次皮質中枢は、視覚的印象の形成には関与していない。

しかし、このプロセスは自発的かつ自動的であるため、私たちはどのようにして判断に至ったかを説明することができない。なぜなら、無意識的な結論の結果は、「あたかも外部の力がわれわれを拘束するかのよう、いわばわれわれの意識に促されて、われわれの意志ではどうすることもできない」解釈だからである。

# Helmholtzの知覚論の影響

細胞や脳のような自己組織化する生物学的システムは、変分自由エネルギーを最小化するものとして理解できるという考え方は、無意識の推論に関するヘルムホルツの観察と、それに続く心理学や機械学習における研究に基づいている。

「ヘルムホルツに倣い、我々は人間の知覚システムを、感覚入力の原因を推論する統計的推論エンジンとみなす。この種の装置は、教師が各感覚入力ベクトルにその根本的な原因をラベル付けしなくても、これらの推論を実行する方法を学習できることを示す。」

*Dayan, P., Hinton, G. E., & Neal, R. (1995). The Helmholtz machine. Neural Computation, 7, 889–904.*

# 知覚の推論と学習のモデル

「知覚に関するヘルムホルツのアイデアを現代の理論で定式化すると、知覚の推論のモデルと学習のモデルに到達する。注目すべきことは、そのモデルが、非常に広い範囲の神経生物学的事実を説明することができるということである。

統計物理学で構成された概念を用いることで、感覚入力の原因を推論する問題と、感覚器における因果的規則性を学習する問題が、まったく同じ原理を用いて解決できることが示すことができる。さらに、この推論と学習を生物学的に妥当な方法で進めることができる。」

# ベイズ理論と階層的モデル

「我々の理論の次のスキームは、経験的ベイズ理論と、感覚情報がどのように生成されるかについての階層的モデルに基づいている。

階層的モデルを用いることで、脳は動的かつ状況に応じて事前予想を構築することができる。このスキームは、脳の組織と反応の多くの側面を理解するための原理的な方法を提供する。」

# 知覚プロセスの「創発性」

「この論文では、このような知覚プロセスは、自由エネルギー原理に適合するシステムの創発的特性のひとつに過ぎないことを示唆する。

ここでいう自由エネルギーとは、その状態や構成によってコード化された期待のもとで、環境とのやりとりすべてに内在する「驚き」を制限するものである。」

# 脳と行動と知覚の変化

「システムは、その構成を変更することで自由エネルギーを最小化することができる。それらは、環境をサンプリングする方法を変更するために、あるいは、その期待を変更するために行われる。これらの変化はそれぞれ行動と知覚に対応し、生物システムに特徴的な、環境との適応的交換を可能にする。

こうした問題の扱いは、システムの状態と構造が、環境に関する暗黙的で確率的なモデルをエンコードしていることを意味する。

この論文では、脳が必要とするモデルに注目し、自由エネルギーの最小化が脳のダイナミクスと構造をどのように説明できるかを考察する。」



# 拡大された「推論」概念

セミナー「創発について考える」  
Part 3 : フリストンが考えたこと 3

# Fristonは「推論」をどのように考えているか

前回は、Fristonの「推論する脳」というアイデアが、Helmholtzの知覚論の影響を受けているという話をしました。

今回のセッションでは、Fristonが「推論」について、どのような考えを持っているのか、もう少しレンジを広げて、紹介したいと思います。取り上げるのは、彼の“The mathematics of mind-time”というエッセイです。

<https://aeon.co/essays/consciousness-is-not-a-thing-but-a-process-of-inference>

ここで彼は、「意識はモノではなく、推論のプロセスだ」という議論を展開しています。こうした立場から、Fristonは「推論」の概念を大幅に拡大します。

## Dennettの「こころの進化」

このエッセイの中で、Fristonは「こころの進化」というサブタイトルを持つ、Dennettの本“From Bacteria to Bach and Back”を引用しています。

タイトルは、「ゲーデル、エッシャー、バッハ」(いわゆる「GEB本」)への当てこすりなのだと思うのですが、この本面白いです。

「創発」を考える上では、とても大事なことを言っていると思います。時間があったら、ぜひ、この「BBB本」お読みください。

# 「推論の逆転」

この本に、「二つの奇妙な推論の逆転」という章があります。二つというのは、DarwinとTuringが「奇妙な推論の逆転」を行ったという議論です。

Dennettは、Darwinを批判したBeverleyという人の進化論批判を紹介します。19世紀の議論ですが。

# From Bacteria The Evolution of Minds to Bach and Back



DANIEL C. DENNETT

## 第4章 二つの奇妙な推論の逆転

4

### Two Strange Inversions of Reasoning

How Darwin and Turing broke a spell

DarwinとTuringは、  
いかにして呪縛を破ったのか

# ベヴァリーによるダーウィン批判 「奇妙な推論の逆転」

「ダーウィンの理論では、「絶対的無知」が製作者である。したがって、このシステム全体の基本原理として、

「完璧で美しい機械を作るためには、  
作り方を知る必要はない」

と宣言することができる。この命題は、注意深く検討すれば、この理論の本質的な趣旨を凝縮した形で表現しており、ダーウィン氏の意をすべて一言で表現していることがわかるだろう。

彼は、奇妙な推論の逆転によって、絶対的無知が、創造的技術のすべての成果において絶対的知恵の代わりに務める資格が十分にあると考えているようだ。」(ベヴァリー1868)

## 創発論としての「創造説」

今回のセミナーのテーマは「創発」なのですが、多分、現代でも影響力のある代表的な「創発論」は「創造説」です。全知万能の存在が、「絶対的知恵」の持ち主が、全てを「創造」したという考えです。Beverley は「創造説」の立場から、「絶対的知恵」の代わりに「絶対的無知」を「創造者」の立場に置いたと、Darwinを批判しているのです。

Bennettの議論を離れませんが、「絶対的無知」を出発点とするという考えは、Jaynesの「最大エントロピー原理」と共通するところがあります。彼は、それによって、従来の統計力学の「推論の逆転」を行ったのです。

# チューリングの「奇妙な推論の逆転」

TuringについてBennettは、こう語ります。

「チューリングが行った奇妙な推論の逆転は、ベヴァリーの言葉を借りれば、次のようになる。

「完璧で美しいコンピュータであるためには、  
数学とは何かを知る必要はない。」

彼が指摘したように、「どの瞬間におけるコンピューターの動作も、彼が観察しているシンボルと、その瞬間の彼の『心の状態』によって決まる」(チューリング 1936)。」

こうした議論については、「機械は推論する」という丸山の議論を参照されたい。例えば、「人工知能と数学」

<https://www.marulabo.net/docs/aimath>

# 「機械の上に構築された構造の中」 で起きた「創発」

僕の議論は、機械に「創発」として新たに「推論の能力」が生まれるというものではありません。コンピュータには、最初から「数学的・論理的推論能力」が備わっているというもので、それは、Turingの議論と、基本的には同じです。

ただ、今回のセミナー「創発について考える」で僕が関心を持った「創発的現象」は、機械そのものではなく、「機械の上に構築された構造の中」で起きたと思われる「創発」です。

この区別は、意味があるものだと思います。

MENU / Q



aeon

DONATE / NEWSLETTER / SIGN IN

# The mathematics of mind-time

The special trick of consciousness is being able to project action and time into a range of possible futures

<https://aeon.co/essays/consciousness-is-not-a-thing-but-a-process-of-inference>

## 「ダーウィンの危険な考え」

私の意識に関する見解は、心の起源を理解することにそのキャリアを費やしてきた哲学者ダニエル・デネットのそれと共鳴する。

デネットは、心のない単なる「原因」(AはBにつながる)が、いかにして我々が知っているような心ある「理由」(Aが起こるからBが起こる)を生み出すことができるのかに関心を抱いている。

デネットの解決策は、彼が「ダーウィンの危険な考え」と呼ぶもので、設計者がいなくても設計が可能であり、理解がなくても能力があり、推論者がいなくても理由(あるいは「自由に浮動する理由」)があるという洞察である。

# 物事は、推論するために存在する

私は、自然が自ら理由を持たずとも、その理由を作り出すことができることを示したいのだ。

以下では、物事は理由のために存在するのではなく、あるプロセスを推論するために存在するのだと主張したい。

ここで言う「推論」とは、推論や仮説設定から生じる説明、つまり潜在的な原因や規則、原理に基づいて観察を説明しようとすることを意味する。

## 推論は仮説の設定である

プロセスに対するこの視点は、心が存在する理由について、むしろ抑制的ではあるが、エレガントな物語へと導いてくれる。

推論は、進化、意識、そして生命そのものを含むすべての理論に近い。それはすべて仮説の設定なのだ。

私たちはすでに動いているプロセスとして世界に投げ込まれ、プロセスは(注意深く選択されたとしても)世界のまばらなサンプルに基づいて、「そこにあるもの」に向かって推論することしかできない。

しかし、推論を行う推論者が存在する前は、推論はどのように行われていたのだろうか？不活性な物質が意識に至るプロセスをどのように始めたのだろうか？

## 時間と共に、混沌に進むはずではないのか？

物理学者によれば、複雑な系はその状態によって特徴づけられる。例えば量子系では、粒子の状態は、その位置、運動量、エネルギー、スピンを含む波動関数によって記述することができる。

私たち自身のような大きな系では、私たちの状態は、身体部分の位置や運動、脳の電気化学的状态、臓器の生理学的変化などすべてを含む。形式的に言えば、システムの状態は、可能な状態の空間における座標に対応する。

時間が進むにつれて、すべてがよりランダムに、分散し、混沌としてくるはずだ。では何が起きているのか？

# リアプノフ関数が系の振る舞いを決める

この空間を何かがどのように動くかは、そのリアプノフ関数に依存する。リアプノフ関数は、特定の条件下でシステムがどのような振る舞いをするかを記述する数学的量である。これは、特定の状態にある確率を、その状態の関数として返す。

システムの各状態のリアプノフ関数がわかれば、ある状態から次の状態への流れを書き出すことができる。これは、ある山の地形の高さをすべての場所で知ることで、その地表を水の流れがどのように流れるかを記述できるようなものだ。山の地形はリアプノフ関数を表し、水の動きはシステムが時間とともにどのように進化するかを表す。

## 関数が返す数はどんどん小さくなっていく それが意味すること

さて、複雑系の重要な特徴は、リアプノフ関数を用いてより確率の高い状態に向かっているように見えることである。

つまり、関数が返す数はどんどん小さくなっていく。このことは、このような系が少数の状態しか占めない傾向があること、さらにそれらの状態が何度も繰り返される傾向があることを意味する。

溪流に例えるなら、水は海に向かって流れ落ち、蒸発した後、雨雲によって山腹に戻ってくる。体温は一定の範囲内で推移し、心臓はリズムカルに鼓動し、呼吸は吸ったり吐いたりする。

# 反復的で自己組織的振る舞いと 熱力学の第二法則

このような反復的で自己組織的な振る舞いが注目に値するのは、宇宙が通常どのように振る舞うかに反しているからだ。

すべてのものは、時間の経過とともに、よりランダムになり、分散し、カオスになるはずである。それが熱力学の第二法則である。すべてのものはカオスに向かう傾向があり、エントロピーは一般的に増大する。

では何が起きているのか？

# 自己組織化とアトラクター

複雑系が自己組織化するのには、アトラクターを持っているからである。

アトラクターとは、相互に補強し合う状態のサイクルのことで、エネルギーが失われて停止するのではなく、動的平衡と呼ばれる状態を通じて、プロセスが安定点を達成することを可能にする。

アトラクターからの逸脱があると、思考、感情、動作の流れが誘発され、最終的にアトラクターのサイクルに戻り、慣れ親しんだ状態に戻る。人間の場合、身体と脳のすべての興奮は、アトラクターに向かって、つまり最も可能性の高い状態に向かって動いているとすることができる。

## アトラクターと推論のイメージ

私たちは皆、多様な可能性を秘めた巨大で高次元の状態空間を流れているが、アトラクターによって限られた円の中を動き回ることが余儀なくされている。

私たちは紅葉の葉のようなもので、小川の乱流の中で終わることのない軌跡をたどりながら、自分の小さな軌跡が世界全体だと思っているのだ。私たち自身を遊び心に満ちたループと表現するのは、目的論的に不毛に聞こえるかもしれない。しかし、これは、あなたや私のような、引きつけ合う状態を持つ複雑系の本質にとって、深い意味を持つ。

新しい経験をするたびに、あなたの体の器官は、起きていることを慣れ親しんだパターンに当てはめようと推論する。

# 何が我々を突き動かしているのか？

要約すると、われわれを含む複雑系は、リアプノフ関数がわれわれ自身のプロセスを正確に記述している限りにおいて存在する。

さらに、もし私たちが存在するのであれば、私たちのすべてのプロセス、私たちのすべての思考と行動は、私たちをより確率の高い状態へと押しやりながら、私たちのリアプノフ関数からの出力の数を減少させているに違いない。

では、実際にはどうだろうか？ここでの要点は、リアプノフ関数の性質を理解することだ。この関数を理解すれば、何が我々を突き動かしているのかがわかる。

# リアプノフ関数の二つの解釈

リアプノフ関数には2つの解釈があることがわかる。

1つ目は情報理論に由来するもので、リアプノフ関数は驚き、つまり特定の状態にある可能性の低さであるという解釈である。

2つ目は統計学に由来するもので、リアプノフ関数は(負の)証拠、つまりその状態を説明する与えられた説明やモデルが正しい確率であるという解釈である。

端的に言えば、私たちが存在するのであれば、驚きを最小化することで、モデルの証拠や自己証明力を高めているに違いないということだ。

このような解釈によって、実在する運動の力学に目的と目的論を与えることができる。

# 推論は世界のモデルに対する証拠を最大化する

推論とは、我々が「世界」と呼ぶシステムの観測された状態を説明する最良の原理や仮説を見つけ出すプロセスである。技術的には、推論は世界のモデルに対する証拠を最大化することを意味する。

私たちは証拠を最大化する義務があるため、事実上、私たち自身をモデルとして世界について推論を行うことになる。そのため、新しい経験をするたびに、起きていることを慣れ親しんだパターンに当てはめようとしたり、この新しい事実を考慮するように自分の内部状態を修正したりするために、何らかの推論を行うのである。

これはちょうど、統計学者が病気の蔓延を説明するために新しいルールが必要かどうか、あるいは銀行の破綻が経済のモデル化に影響を与えるべきかどうかを判断しようとするときに経験するようなプロセスである。

# アトラクターの重要性

これで、なぜアトラクターが非常に重要なのがわかる。アトラクター状態は、驚きが少なく、証拠能力が高い。したがって複雑系は、慣れ親しんだ信頼できるサイクルに陥る。

なぜなら、これらのプロセスは必然的に、自らの存在を支える原理を検証することに関わるからである。アトラクターはシステムを予測可能な状態に押しやり、それによってシステムがその世界について生成したモデルを強化する。

この驚きを最小化し、自己検証し、推論する振る舞いが失敗すると、システムは驚きのある、なじみのない状態へと衰退していく。アトラクターは、それ自身を呼び起こすために推論を行うプロセスの産物である。言い換えれば、アトラクターは「生きている」ということの基礎なのである。

# 我々は、推論を行う存在である

上記の定式化を受け入れる限り、生物を含むあらゆる種類の複雑系について、究極の抑制的な説明ができることになる。

特定の状態を繰り返し占めるプロセス(あなたや私のような)は、その存在そのものによって、推論を行っているに違いない。

(後半に続く)



# こころの時間の数学

セミナー「創発について考える」  
Part 3 : フリストンが考えたこと 4

# こころの時間の数学

このセッションでは、前回に引き続き、Fristonの「こころの時間の数学」の後半の紹介をします。

タイトルには「数学」という言葉が入っているのですが、ここには全く数学の話は出てきません。とても平易な文章です。

少し長くなりましたが、ぜひ、スライドのpdfで、ゆっくりお読みください。

MENU / Q



aeon

DONATE / NEWSLETTER / SIGN IN

# The mathematics of mind-time

The special trick of consciousness is being able to project action and time into a range of possible futures

<https://aeon.co/essays/consciousness-is-not-a-thing-but-a-process-of-inference>

# 「淘汰」は推論である

進化の過程や自然淘汰を推論という観点から考えることはほとんどないだろう。

理論的神経生物学では、現在まさにそのような解釈がなされている。例えば、自然界が生物を生存能力や繁殖能力で「選択」する方法は、推論に基づいていることが判明している。

多様な表現型は、何が「機能」しうるかについての複数の仮説に相当する。各個体は、この生態学的ニッチを占めるべきものの仮説またはモデルであり、環境からの圧力のもとで淘汰を競わなければならない。

進化は複雑なシステムであるため、自己決定的でなければならない。つまり、進化は常に、その生態学的ニッチを占める可能性が高い生物を「選択」する。

大きな鉤爪は獲物を捕らえるのに適しているため存続するかもしれないし、硬い殻は捕食者に対抗するのに役立つかもしれない。

つまり、適応適性とは、ある表現型がその環境に存在する限界的な可能性以上でも以下でもないのだ。言い換えれば、その生存は、その表現型がニッチにとって良いモデルであるという証拠以上でも以下でもない。

# 意識は推論のプロセス

同じ考え方を意識に当てはめると、意識は推論のプロセスでもあるはずだ。

意識的な処理とは、感覚状態の原因を推論することであり、それによって驚きを回避するために世界をナビゲートすることである。

自然淘汰が異なる生物の中から選択することで推論を行うのに対し、意識は同じ生物(特にその脳)の異なる状態の中から選択することで推論を行う。

脳を推論の自己証明器官と見なせば、その解剖学的・生理学的側面のほとんどすべてが、驚きを最小限に抑えるように調整されているように思える。

このように「外の世界」の因果構造を内面化することは、自分自身の状態を予測するためには、そのような感覚がどのように生成されるのかという内的モデルが必要だという事実を反映している。

# 複雑な推論プロセスはすべて意識的か？

しかし、もし意識が推論であるならば、進化から経済、原子に至るまで、複雑な推論プロセスはすべて意識的であるということになるのだろうか？

おそらくそうではないだろう。

ウイルスには、推論プロセスの自己組織化ダイナミクスがすべて備わっている。

しかし、明らかにウイルスにはベジタリアンと同じ性質はない。では、何が違うのだろうか？

# 意識的な生き物とそうでない生き物の区別

意識的な生き物とそうでない生き物を区別するのは、彼らが行動と時間について推論する方法である。

私の議論のこの部分は、システムと世界の相互関係にかかっている。世界はシステムに働きかけ、推論の基礎となる感覚的印象を与える。

一方、システムは世界に働きかけて、感覚の流れを変化させ、自分が識別した世界のモデルに適合させる。例えば、私たちは見て、次にどこを見るべきかを推論する。

# 未来についてのモデル

行動が推論に依存するのであれば、システムは行動の結果について推論できなければならない。予想される結果について推測することができなければ、何をすべきかを選ぶことはできない。

しかし、ここで重要なひねりがある。生き物は、その未来についてのモデルを持っていない限り、行動の結果を推測することはできない。こうしたらどうなるか、ああしたらどうなるかを知る必要がある。

しかし、ある行動がもたらす結果についての感覚的な証拠は、時間の容赦ない前進のおかげで、それが実行されるまで手に入らない。

# 意識時間の厚み

時間の矢の結果として、将来の行動の影響を把握できるシステムは、必然的に時間的な厚みを持たなければならない。実際に起こっていないこと、起こらないかもしれないことを予測できるような、自分自身と世界に関する内部モデルを持たなければならない。

このようなモデルは、どれだけ前方まで予測するかによって、またどれだけ後方まで予測するかによって、より厚くなったり薄くなったり、より深くなったり浅くなったりする。

より深い時間構造を持つシステムは、自分の行動の事実にも反した仮想的な結果を推論するのが得意である。神経科学者のアニル・セスはこれを「反事実的深さ」と呼んでいる。

## 分厚い時間モデルと「目的」

では、システムが分厚い時間モデルを持つ場合、どのような行動を推測し、選択するのだろうか？

答えは簡単で、ある行動の後に予想される驚きを最小化する。この証明は、私たちがすでに知っていることから、帰謬法によって導かれる。

実際のところ、システムはどのようにして予想される驚きを最小化するのだろうか？まず、システムは不確実性を減らすために行動する。つまり、将来起こりうる驚き(寒い、空腹、死など)を避けるために行動するのだ。私たちの行動はほとんどすべて、このような不確実性を最小化する衝動の観点から理解することができる。

このようなシステムが世界に及ぼす行動には目的が備わっているように見える。それは、まだ現実には起こっていないが、起こりうる驚きを最小限に抑えるという目的である。

# ウィルスと人間の違い

このようなシステムを、エージェントあるいは自己と呼ぶことができるかもしれない。

ウィルスは外的環境の変化に推論的に反応するとしても、過去についての深い理解や、まだ起こっていない驚きを最小限に抑えることができるような未来についての長期的な展望を体現しているわけではない。

一方、人間はサプライズを最小化し、前向きで目的意識的な方法で自己検証を行う。そこでは、エージェントの将来の見通しが行動選択の本質的な部分となる。

# 自然淘汰のプロセス

同じように、私たちは進化を意識的なものと呼ぶことの問題を回避している。

自然淘汰のプロセスは、驚きを最小化する(つまり適応適性を最大化する)が、不確実性やシステム全体に期待される驚き(つまり、ダーウィンの進化ではない別の進化操作のもとで期待される適応適性)は最小化しない。

# 意識と自己組織化プロセスとの違い

意識と、より普遍的な自己組織化プロセスとの重要な違いは、淘汰の必要性にあるようだ。

例えば、競合するシステム間の選択(進化における表現型など)や、反射の誘発(単純な生物における化学走性など、化学物質の濃度が高い方に向かったり、低い方に向かったりする)などである。

逆に、私たちが意識と結びつけて考えている淘汰は、同じシステムの中で並行して行われる。つまり、異なる状況下で複数の未来をシミュレートし、最も意外性の少ない結果をもたらす行動を選択できるシステムである。意識的な自己とは、このような反事実的な未来を、能動的な推論を容易にする方法でとらえる方法にすぎない。

進化から意識的な処理に至るまで、すべての生物学的プロセスは、推論を実行していると見なすことができる。

# 能動的推論としての意識 意識レベル

能動的推論としての意識は、現実的に意味があるのだろうか？  
私はそう主張したい。

精神医学の観点からすると、意識の変容状態には2つの種類がある。例えば、睡眠時、麻酔時、昏睡時などである。あるいは、精神医学的症候群や向精神薬、サイケデリック・ドラッグに関連した意識状態の変化である。

意識レベルの違いは、行動への影響と絡み合っている。簡単に言えば、意識レベルの低下の特徴は、反応性の欠如である。意識はないが刺激に反応して行動する人を想像してほしい。人が引き出せる唯一の反応は、「今、ここ」での驚きの最小化を反映した反射である。

対照的に、この人物はいったん目覚めると、過去と未来に関する予測機械を起動させることができる。

このことは、私たちの日常生活において、睡眠と覚醒のサイクルによって時間的な厚みや深さが変化すること、つまり、意識レベルと私たちが行っている推論の厚みの間に対応関係があることを示唆している。

この見解では、私たちのモデルが「厚み」を失い、ウィルスのように「薄く」なるたびに、意識の喪失が起こる。

# 誤った推論としての精神疾患

精神科医として、私はいくつかの理由から、意識の変容を推論の変容として捉える考え方に惹かれている。その中でも重要なのは、精神疾患を誤った推論として理解できることである。

例えば、統計学では、偽陽性と偽陰性の2種類の偽推論がある。

偽陽性とは、幻覚や妄想のように、そこに何かがないのに、そこに何かがあると推論してしまうことである。

逆に、偽陰性とは、何かを認識できなかつたり、不可能な曖昧さを受け入れることができなかつたりするなど、そこに何かがあるにもかかわらず、何かを推論できなかつた場合のことである（例えば、患者がよく投げかける質問：私は誰ですか？）

これは臨床的には、見当識障害や、認知症やその他の心の器質的症候群を特徴づけるさまざまな形の失認症となる。実用的な観点からは、能動的推論の背後にある神経細胞の仕組みが次第に解明されつつあるため、これは有用な視点である。

# すべての生物学的プロセスは推論を実行している

私たちはかなり迅速に議論を進めてきた。まず、複雑系（生命系を含む）について語りたいのであれば、そのプロセスが示す必要な振る舞いを特定しなければならない。

これは、生きるということは、何度も何度も繰り返される一連の引きつけられた状態に存在することを意味する、ということに注目すれば、かなり簡単にできる。

このことは、情報理論における（負の）自己証明やサプライズと同じリアプノフ関数の存在を意味する。これは、進化から意識的な処理に至るまで、すべての生物学的プロセスが何らかの推論を実行していると解釈できることを意味する。


# いつ意識が呼び起こされるのか

もしそうだとしたら、私たちはどの時点で意識を呼び起こすのだろうか？

ここで提案されているのは、心は、将来の行動の結果について推論する根拠となる、時間的な厚み、あるいは反事実的な深みを自己認知が持つときに生まれるというものだ。

心が存在する本当の理由はない。心が存在するように見えるのは、存在そのものが推論のプロセスの終着点だからにほかならない。

意識とは、私の未来についての推論に過ぎないのだ。



おわりに  
今後の展望

セミナー「創発について考える」  
おわりに

# 今回のセミナーのトピックについて 留意すべきこと

今回のセミナーは、人工知能上で創発が起きるのかという問題意識に触発されたものですが、ここで取り上げたトピックは、散逸構造の遍在や新しい推論能力の生成や人間の脳の進化を対象としたもので、人工知能を対象にしたものではありません。

取り上げたトピックは、主にエントロピー論との関連で、新しい構造・新しい機能の発生について論じたものです。ただ、これらのトピックスは、新しい創発について考える上での参考になると僕は考えています。

## 新しい創発の舞台について

すでにプリゴジンが論じたように、こうしたエントロピーによる新しい構造の生成は、不可逆な過程が生み出すもので、決定論的で可逆な法則に支配される「機械」の中では起こりません。

どんな大規模なコンピュータシステムでも、どんなに長いこと電源を入れていたとしても、100万年後にその中から新しいコンピュータシステムの赤ちゃんが自発的に産まれることはありません。

ただ、コンピュータの中で生成される情報の世界は、コンピュータそのものとは異なるものだと思います。もしも、それが明確な「存在する過程」として形を取りうるなら、新しい構造・新しい機能が生まれる可能性はあるかもしれないと考えています。

# 現代の熱力学としての量子情報理論

熱力学やエントロピーというと、何か古い学問だというイメージがあるかもしれませんが、でも、それは誤解だと思います。

19世紀の熱力学は統計力学に姿を変え、20世紀の量子論の登場を準備しました。ただ、量子論によって、ミクロの世界とマクロな世界のリンクが完全に解明されたわけではありません。宇宙のようなマクロな世界を支配する重力理論と量子論との統合は、まだ未解決の課題です。

量子論と相対論の統一の研究の中で注目されているのが、量子情報理論です。それは、シャノンの情報理論の洗礼を受けた、熱力学の直系の子孫に他なりません。

# 創発と還元

僕は、創発という現象に興味を持っています。ただ、そうした現象に興味を持つことと、そうした現象の根拠を問うこととは、別のことだと考えています。

創発という現象への関心は、それだけだと我々の認識の発展に大きく寄与することはないと感じています。我々の認識が飛躍的に発展するのは、現象の背後にある根拠が明確になった時です。

それは科学では一般には、上位の階層の世界の現象を、下位の世界の運動法則と結びつける「還元」によって行われます。

ボルツマンの分子運動論による熱力学の基礎づけも、プリゴジンの非平衡の熱力学の導入による散逸構造論も、基本的には還元による現象の根拠づけです。

「人工知能上の創発現象」も、いずれその根拠を問う問いに向かうはずです。



Boltzmann  
1844~1906

# 今後の関心

## 数学的抽象化と分子生物学的還元

「人工知能上の創発」はさておき、脳のモデルについては、次のような関心を持っています。

一つは、脳のモデルの数学的抽象化としての、カテゴリー論的脳モデルの定式化です。

もう一つは、脳までは辿り着いていないのですが、感覚器官の成立と進化の分子生物学的還元です。

# カテゴリー論の応用としての Toby Smitheのベイジアン脳モデル

Mathematical Foundations for a  
Compositional Account of the  
Bayesian Brain



Toby St Clere Smithe  
St Edmund Hall  
University of Oxford

A thesis submitted for the degree of  
*Doctor of Philosophy*

Trinity 2023

カテゴリー論的な脳モデルでは、  
Toby St Clere Smithe の  
“Mathematical Foundations  
for a Compositional Account  
of the Bayesian Brain” が面白  
いです。もちろん、Fristonのことは  
よく研究しています。

いつかセミナーで紹介できればと  
思っています。

# もっとも原始的な感覚としての「嗅覚」の 分子生物学的進化

感覚器官の分子生物学的進化については、A. S. Barwichの「匂いの哲学 - 鼻がこころに語るもの」  
“Smellosophy -- What the Nose Tells the Mind” がとても面白かったです。

もう少し勉強して、紹介できればと思っています。

