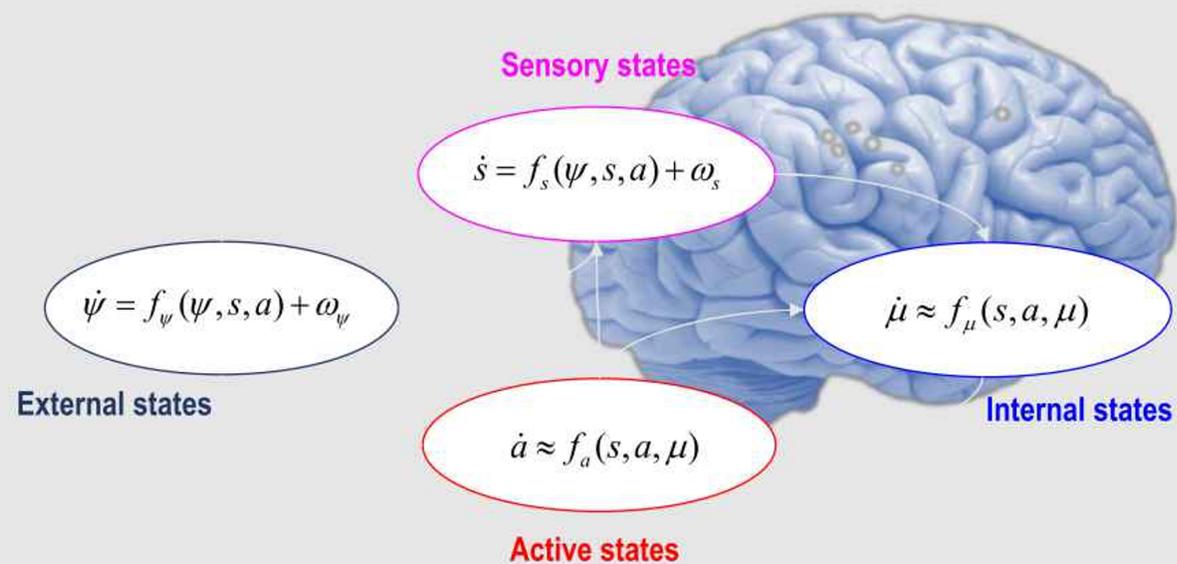


自由エネルギー—原理勉強会

1章 & 2章前半

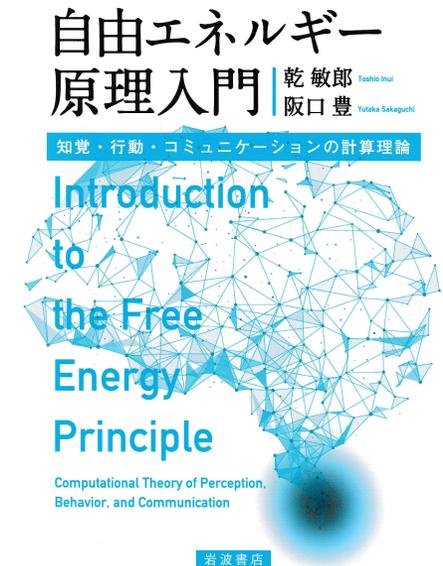


2022/3/7 紺野大地 (東京大学 池谷研究室)

勉強会について

- 目的：自由エネルギー原理について、数学的な部分も含めて理解する
- 教科書：『自由エネルギー原理入門: 知覚・行動・コミュニケーションの計算理論』
- 形式：1回1.5-2時間、毎週水曜日19時開始、基本的に1回1章
- 発表者やその他情報は以下のnotionページを参照

<https://daichikonno.notion.site/daichikonno/abe96c17997745d9aad947bc2ecd9411>



自由エネルギー原理(Free energy principle, FEP)とは

REVIEWS

The free-energy principle: a unified brain theory?

Karl Friston

Abstract | A free-energy principle has been proposed recently that accounts for action, perception and learning. This Review looks at some key brain theories in the biological (for example, neural Darwinism) and physical (for example, information theory and optimal control theory) sciences from the free-energy perspective. Crucially, one key theme runs through each of these theories — optimization. Furthermore, if we look closely at what is optimized, the same quantity keeps emerging, namely value (expected reward, expected utility) or its complement, surprise (prediction error, expected cost). This is the quantity that is optimized under the free-energy principle, which suggests that several global brain theories might be unified within a free-energy framework.

Free energy
An information theory measure that bounds or limits (by being greater than) the surprise on sampling some data, given a generative model.

Homeostasis
The process whereby an open or closed system regulates its internal environment to maintain its states within bounds.

Entropy
The average surprise of outcomes sampled from a probability distribution or density. A density with low entropy means that, on average, the outcome is relatively predictable. Entropy is therefore a measure of uncertainty.

The free-energy principle
The free-energy principle (BOX 1) says that any self-organizing system that is at equilibrium with its environment must minimize its free energy¹. The principle is essentially a mathematical formulation of how adaptive systems (that is, biological agents, like animals or brains) resist a natural tendency to disorder²⁻⁴. What follows is a non-mathematical treatment of the motivation and implications of the principle. We will see that although the motivation is quite straightforward, the implications are complicated and diverse. This diversity allows the principle to account for many aspects of brain structure and function and lends it the potential to unify different perspectives on how the brain works. In subsequent sections, I discuss how the principle can be applied to neuronal systems as viewed from these perspectives. This Review starts in a rather abstract and technical way but then tries to unpack the basic idea in more familiar terms.

Motivation: resisting a tendency to disorder.
The defining characteristic of biological systems is that they maintain their states and form in the face of a constantly changing environment¹⁻⁴. From the point of view of the brain, the environment includes both the external and the internal milieu. This maintenance of order is seen at many levels and distinguishes biological from other self-organizing systems; indeed, the physiology of biological systems can be reduced almost entirely to their homeostasis⁵. More precisely, the repertoire of physiological and sensory states in which an organism can be is limited, and these states define the organism's phenotype. Mathematically, this means that the probability of these (interoceptive and exteroceptive) sensory states must have low entropy; in other words, there is a high probability that a system will be in any of a small number of states, and a low probability that it will be in the remaining states. Entropy is also the average self information or 'surprise'⁶ (more formally, it is the negative log-probability of an outcome). Here, 'a fish out of water' would be in a surprising state (both emotionally and mathematically). A fish that frequently forsook water would have high entropy. Note that both surprise and entropy depend on the agent: what is surprising for one agent (for example, being out of water) may not be surprising for another. Biological agents must therefore minimize the long-term average of surprise to ensure that their sensory entropy remains low. In other words, biological systems somehow manage to violate the fluctuation theorem, which generalizes the second law of thermodynamics⁸.

The Wellcome Trust Centre for Neuroimaging, University College London, Queen Square, London, WC1N 3BG, UK.
e-mail: k.friston@fil.ion.ucl.ac.uk
doi:10.1038/nrn2787
Published online 15 January 2010

NATURE REVIEWS | NEUROSCIENCE

VOLUME 11 | FEBRUARY 2010 | 127

© 2010 Macmillan Publishers Limited. All rights reserved.

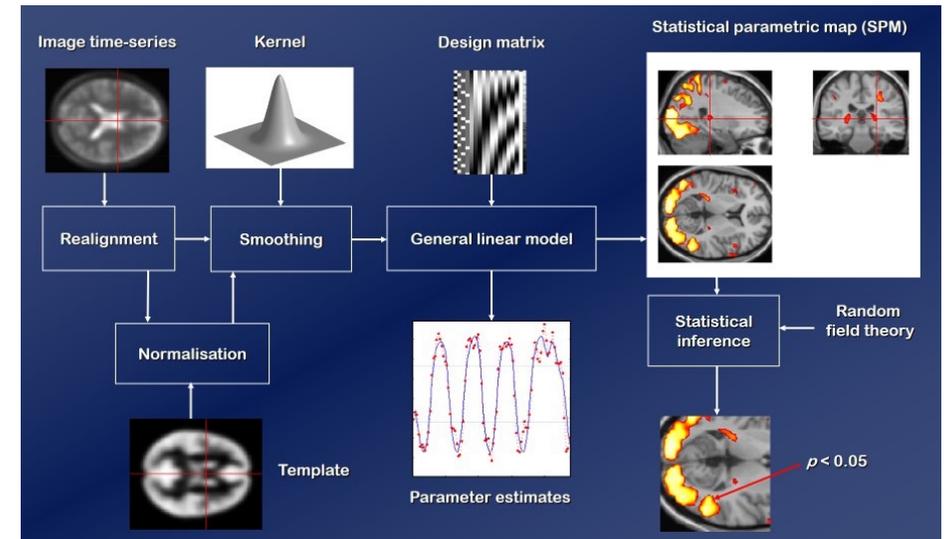
- “Any self-organizing system that is at equilibrium with its environment **must minimize its free energy.**”
- “**脳は自由エネルギーを最小化するように設計されている**”
(『自由エネルギー原理入門』より引用)
- “**生物の知覚や学習、行動は自由エネルギーと呼ばれるコスト関数を最小化するように決まり、その結果生物は外界に適応できる**”
(神経回路は潜在的な統計学者 | 理化学研究所)
- “Several global brain theories **might be unified** within a free-energy framework.”

“**脳の統一理論**”としての期待

[Friston, 2010]

Karl Fristonとは

- 1959年 イギリス生まれ
- 初期は統合失調症の研究に従事
- その後、脳イメージング研究を開始
- 1994年 統計的パラメトリックマッピング
(Statistical parametric mapping, SPM) 開発
- 2006年 自由エネルギー原理の提唱



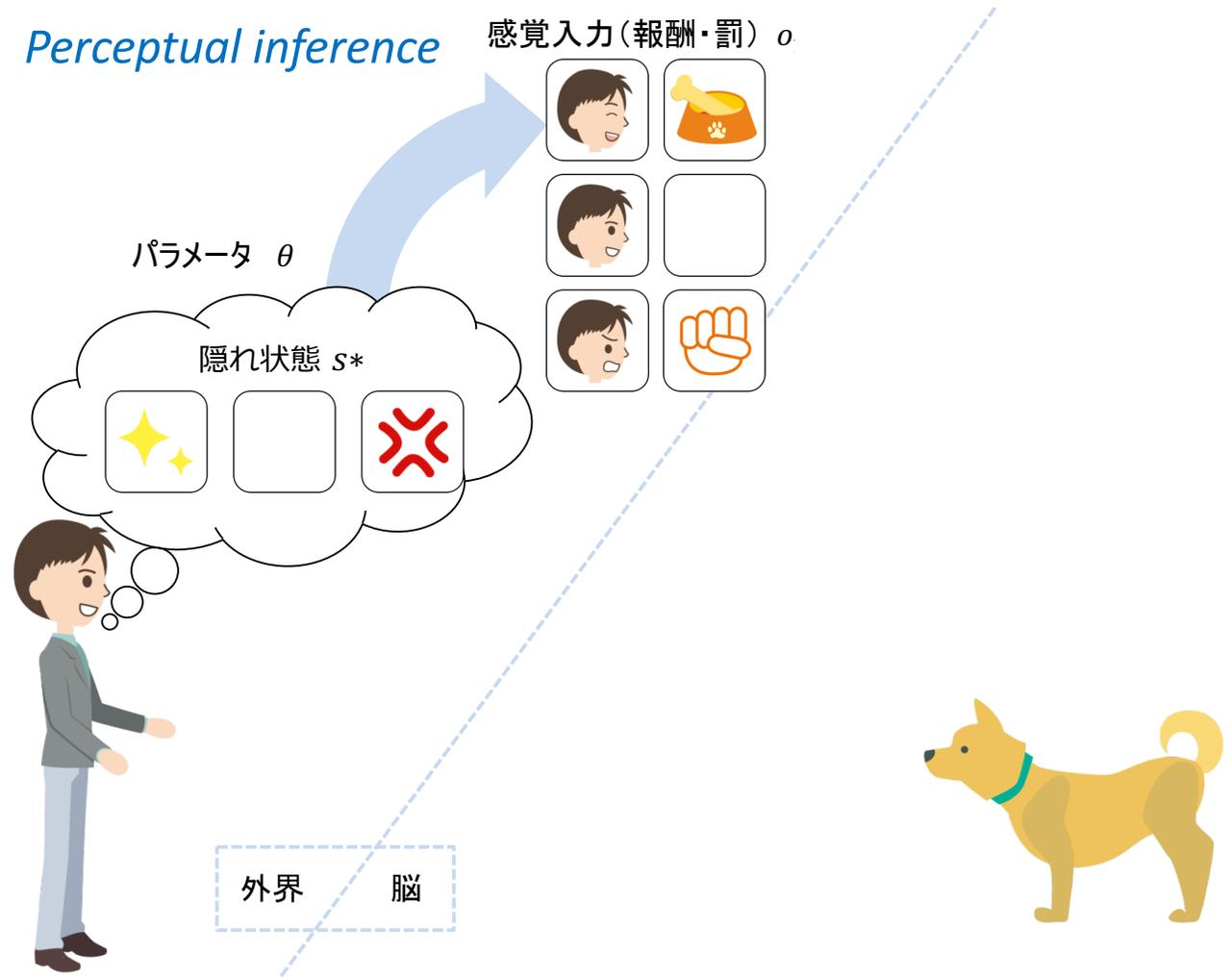
<https://www.fil.ion.ucl.ac.uk/spm/>

自由エネルギー原理のイメージをつかむ

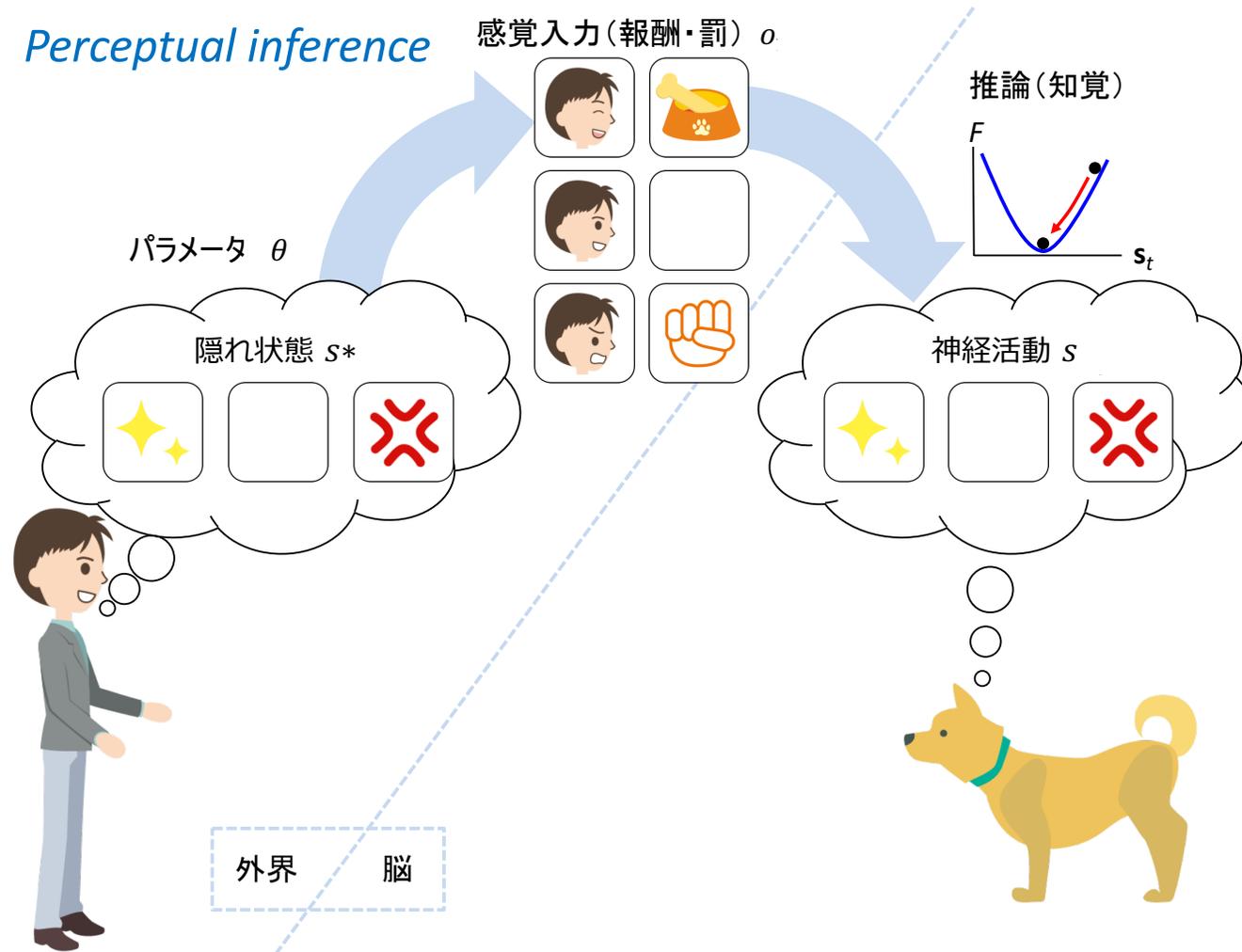
磯村拓也先生(理化学研究所)のスライドを使わせていただきます(了承済)



Perceptual inference



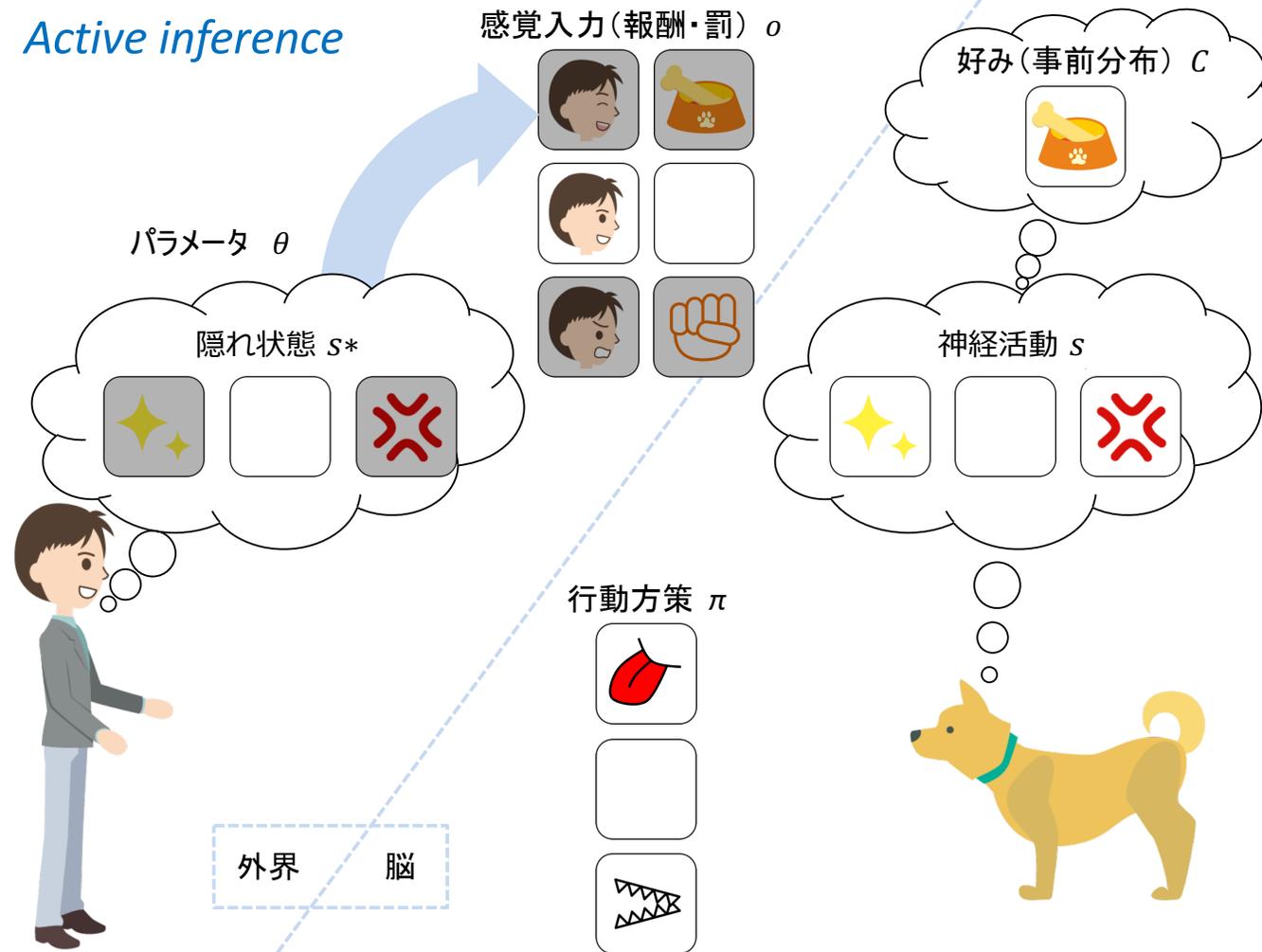
Perceptual inference



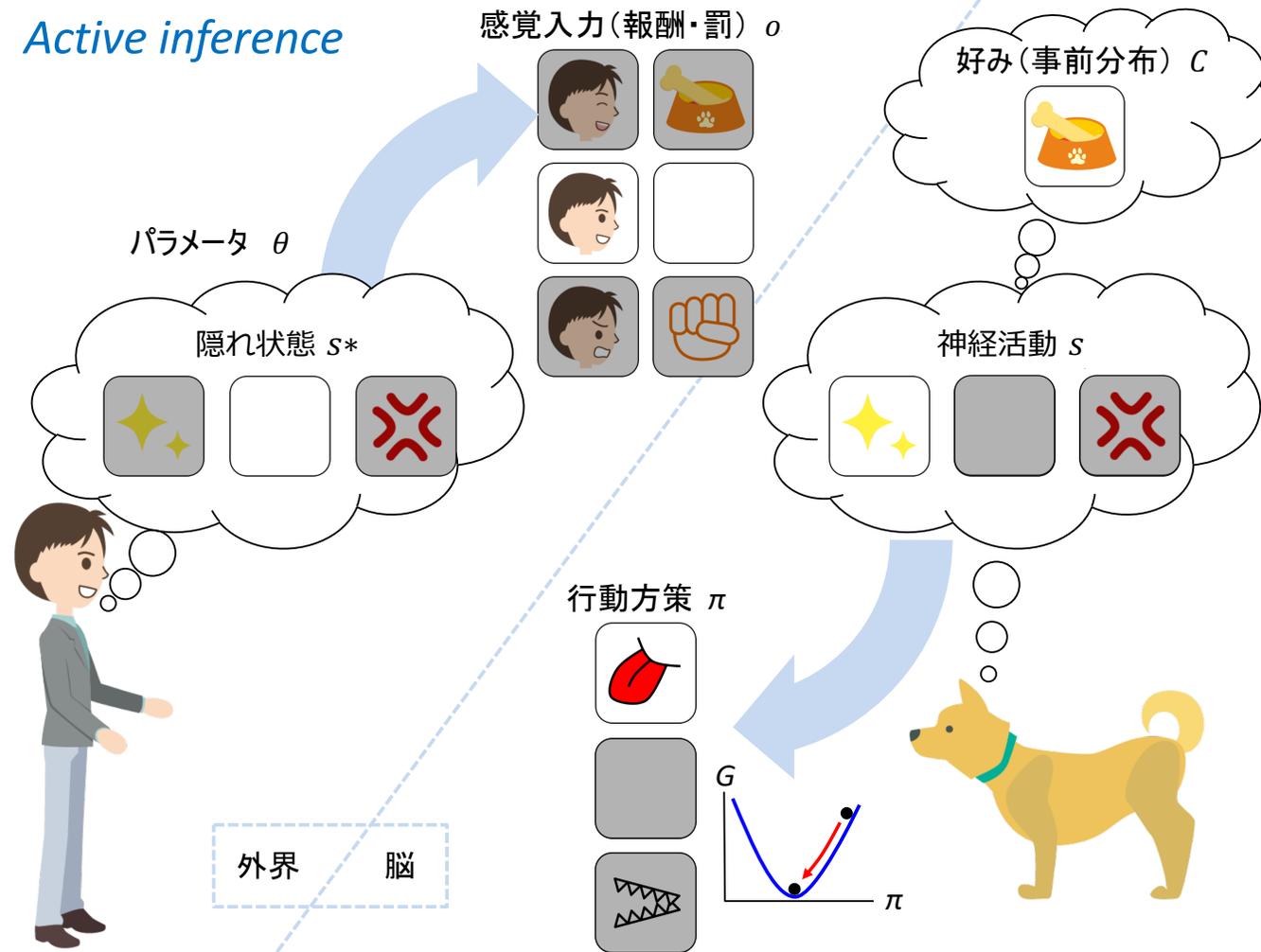
知覚：感覚入力から生成モデルに基づいて外界の隠れ状態を推論する。

この時、自由エネルギーを最小化させる知覚が選択される。

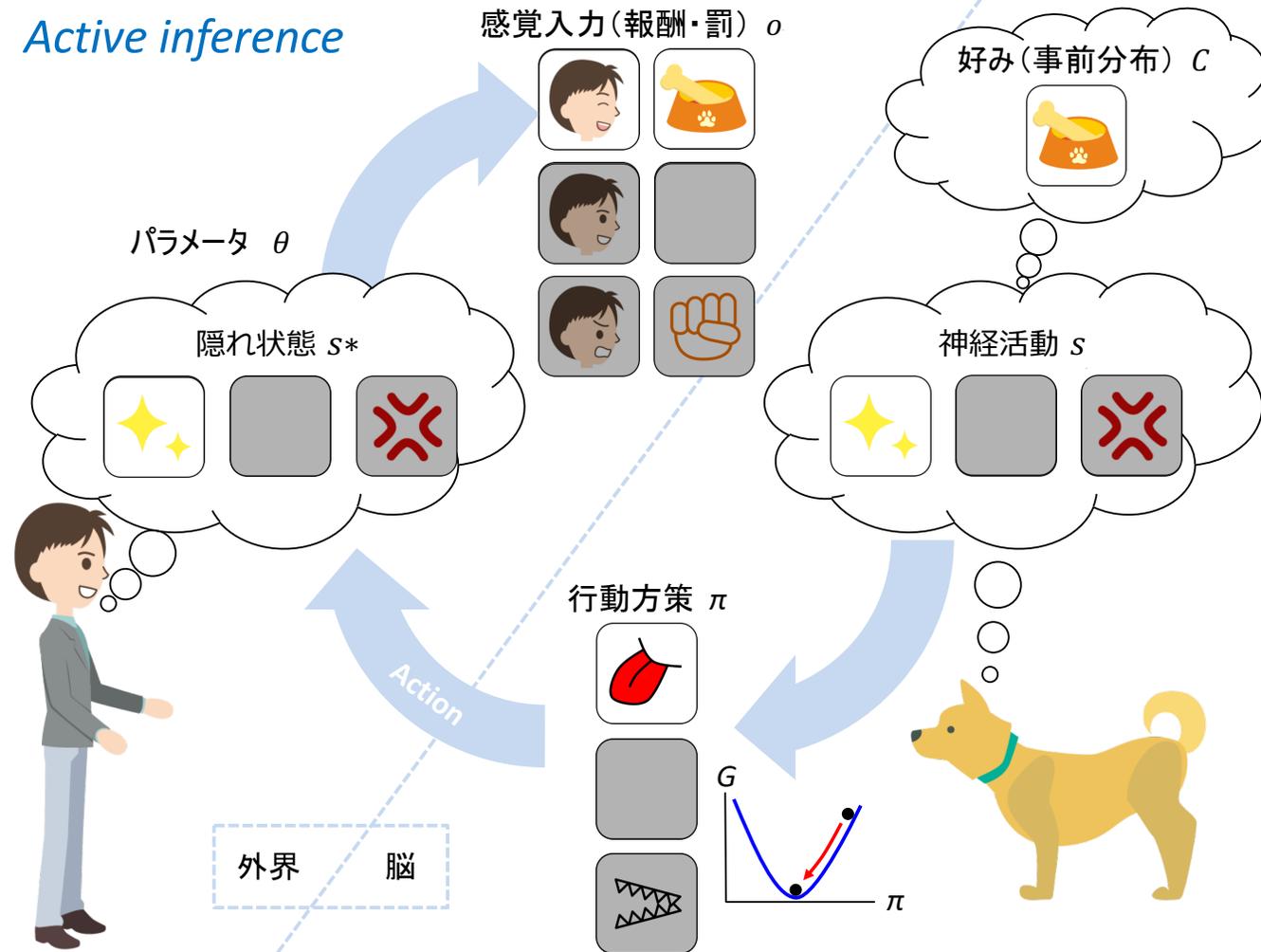
Active inference



Active inference



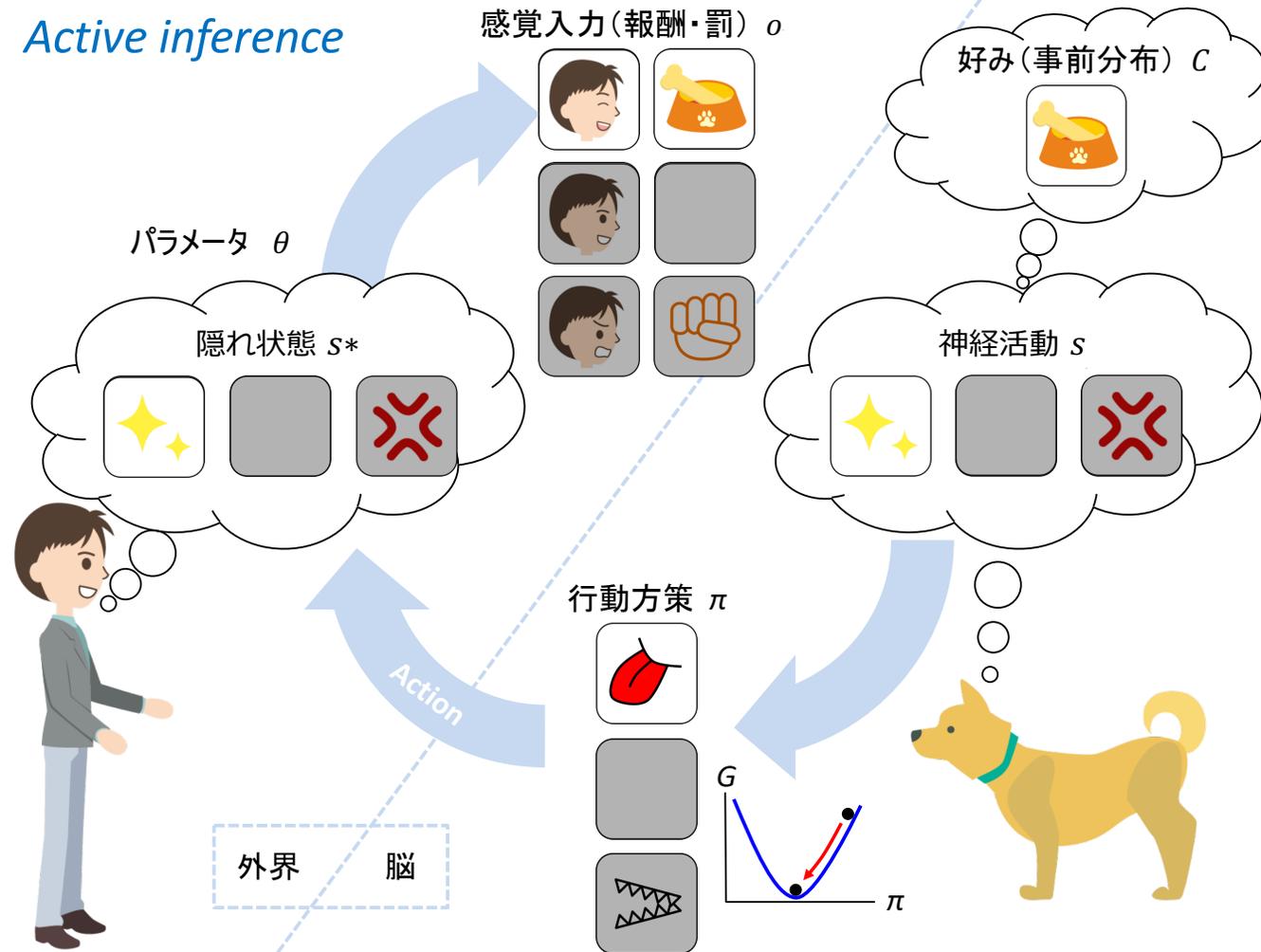
Active inference



行動：行動により、自分が望む感覚入力を得られるように外界に介入することができる。

これを**能動的推論**と呼ぶ。この時、自由エネルギーを最小化させる行動が選択される。

Active inference



学習：知覚・行動と並行して、生成モデルそのものもアップデートされる。

このアップデートも、自由エネルギーを最小化させるように決定される。

ここまでのまとめ

1. **知覚**：感覚入力から生成モデルに基づいて外界を推論し、知覚が得られる。

この時、自由エネルギーを最小化させる知覚が選択される。

2. **行動**：行動により、自分が望む感覚入力を得られるように外界に介入することができる。

これを**能動的推論**と呼ぶ。この時、自由エネルギーを最小化させる行動が選択される。

3. **学習**：知覚・学習と並行して、生成モデルそのものもアップデートされる。

このアップデートも、自由エネルギーを最小化させるように決定される。

このように、知覚・行動・学習は全て自由エネルギーを最小化させるように実現される。

この過程を経て、世界は生物にとって「より予測可能な場所」になる。

自由エネルギー原理のイメージがつかめたところで、
いよいよ教科書の内容に入っていきます！

本の構成

はじめに

変数記号一覧

第1章 自由エネルギー原理を理解するために

- 1.1 神経回路の数理理論の発展
- 1.2 環境の状態の推論機能からHelmholtz マシンまで
- 1.3 神経細胞の基本特性と学習アルゴリズム
- 1.4 脳の多層構造と双方向性結合の機能
- 1.5 自由エネルギー原理

第2章 知覚過程のモデル化

- 2.1 諸行無常の環境を記述する
- 2.2 推論に必要な環境の生成モデル
- 2.3 生成モデルの表現
- 2.4 推論過程をモデル化する
- 2.5 変分Bayes推定とHelmholtzの自由エネルギー
- 2.6 自由エネルギーを具体的に書く
- 2.7 自由エネルギーの最小解の求め方
- 2.8 階層的動的モデルを階層的神経回路で解く
- 2.9 階層的なメッセージパッシング

第3章 パラメータの学習

- 3.1 パラメータの学習と平均場近似
- 3.2 平均場近似の生理学的意味
- 3.3 パラメータの学習とHebbの学習則

第4章 能動的推論と行動決定

- 4.1 能動的推論とは
- 4.2 生成過程と生成モデル
- 4.3 腕の運動
- 4.4 制御状態と行為選択
- 4.5 能動的推論としてのホメオスタシス
- 4.6 感覚減衰
- 4.7 離散システムにおける能動的推論
- 4.8 期待自由エネルギーの導出と行動決定
- 4.9 行動の価値と主体感

第5章 離散系と連続系の統合

- 5.1 読書過程の離散系意思決定モデル
- 5.2 離散系における単語・文の処理
- 5.3 離散系と連続系のインターフェースの詳細
- 5.4 視線制御の神経回路による実装

第6章 運動制御と運動認識

- 6.1 経時的処理切り替えの神経機構
- 6.2 運動制御モデルの例1：ランプへの到達運動
- 6.3 運動制御モデルの例2：系列的な運動
- 6.4 運動制御モデルの例3：手書き文字の生成と認識

第7章 音声コミュニケーション機能

- 7.1 環境のダイナミクスを予測できる生成モデル
- 7.2 階層構造で異なる時間スケールのダイナミクスを作る
- 7.3 コミュニケーション機能への応用
- 7.4 鳥の歌の学習モデルとコミュニケーション
- 7.5 Markovブランケットの意味と役割

参考文献

索引

本の構成

はじめに

変数記号一覧

今回扱う部分

第1章 自由エネルギー原理を理解するために

- 1.1 神経回路の数理理論の発展
- 1.2 環境の状態の推論機能からHelmholtz マシンまで
- 1.3 神経細胞の基本特性と学習アルゴリズム
- 1.4 脳の多層構造と双方向性結合の機能
- 1.5 自由エネルギー原理

第2章 知覚過程のモデル化

- 2.1 諸行無常の環境を記述する
- 2.2 推論に必要な環境の生成モデル
- 2.3 生成モデルの表現
- 2.4 推論過程をモデル化する
- 2.5 変分Bayes推定とHelmholtzの自由エネルギー
- 2.6 自由エネルギーを具体的に書く
- 2.7 自由エネルギーの最小解の求め方
- 2.8 階層的動的モデルを階層的神経回路で解く
- 2.9 階層的なメッセージパッシング

第3章 パラメータの学習

- 3.1 パラメータの学習と平均場近似
- 3.2 平均場近似の生理学的意味
- 3.3 パラメータの学習とHebbの学習則

第4章 能動的推論と行動決定

- 4.1 能動的推論とは
- 4.2 生成過程と生成モデル
- 4.3 腕の運動
- 4.4 制御状態と行為選択
- 4.5 能動的推論としてのホメオスタシス
- 4.6 感覚減衰
- 4.7 離散システムにおける能動的推論
- 4.8 期待自由エネルギーの導出と行動決定
- 4.9 行動の価値と主体感

第5章 離散系と連続系の統合

- 5.1 読書過程の離散系意思決定モデル
- 5.2 離散系における単語・文の処理
- 5.3 離散系と連続系のインターフェースの詳細
- 5.4 視線制御の神経回路による実装

第6章 運動制御と運動認識

- 6.1 経時的処理切り替えの神経機構
- 6.2 運動制御モデルの例1：ランプへの到達運動
- 6.3 運動制御モデルの例2：系列的な運動
- 6.4 運動制御モデルの例3：手書き文字の生成と認識

第7章 音声コミュニケーション機能

- 7.1 環境のダイナミクスを予測できる生成モデル
- 7.2 階層構造で異なる時間スケールのダイナミクスを作る
- 7.3 コミュニケーション機能への応用
- 7.4 鳥の歌の学習モデルとコミュニケーション
- 7.5 Markovブランケットの意味と役割

参考文献

索引

目次 (1章)

- 1.1 神経回路の数理理論の発展
- 1.2 環境の状態の推論機能からHelmholtzマシンまで
- 1.3 神経細胞の基本特性と学習アルゴリズム
- 1.4 脳の多層構造と双方向性結合の機能
- 1.5 自由エネルギー原理

目次 (1章)

1.1 神経回路の数理理論の発展

~~1.2 環境の状態の推論機能からHelmholtzマシンまで~~ → 割愛

1.3 神経細胞の基本特性と学習アルゴリズム

~~1.4 脳の多層構造と双方向性結合の機能~~ → 割愛

1.5 自由エネルギー原理

目次 (1章)

1.1 神経回路の数理理論の発展

~~1.2 環境の状態の推論機能からHelmholtzマシンまで → 割愛~~

1.3 神経細胞の基本特性と学習アルゴリズム

~~1.4 脳の多層構造と双方向性結合の機能 → 割愛~~

1.5 自由エネルギー原理

計算論的神経科学のごく簡単な歴史

1943年 McCullochとPittsが神経細胞のモデル（**神経素子モデル**）を提唱

- 神経細胞の入出力を初めてモデル化

1949年 Hebbがシナプスの学習則（**Hebb則**）を提唱

- シナプス前細胞と後細胞が同時に発火すると、両者間のシナプス結合強度が上昇する

1958年 Rosenblattが**パーセプトロン**を提唱

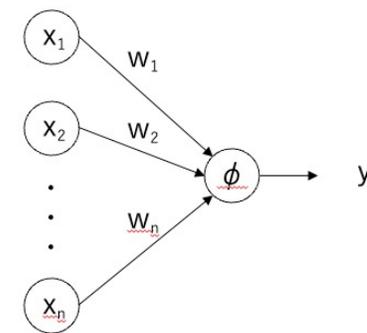
- McCulloch-Pittsの神経素子モデルを用いた、ネットワークとしての脳のモデル

1986年 Rumelhartが**誤差逆伝播法**（**Backpropagation**）を提唱

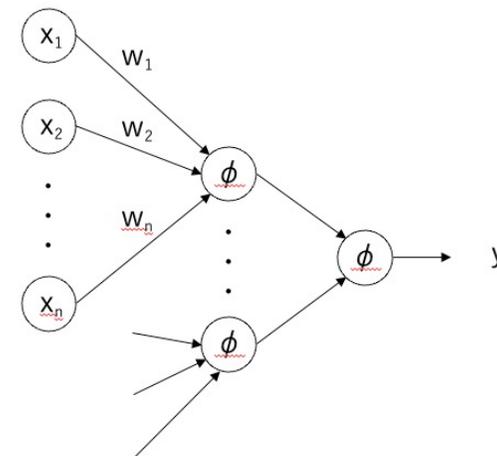
- 多層パーセプトロンの学習が可能に

2006年 Hintonが**オートエンコーダ**を提唱

- 現在の深層学習につながる



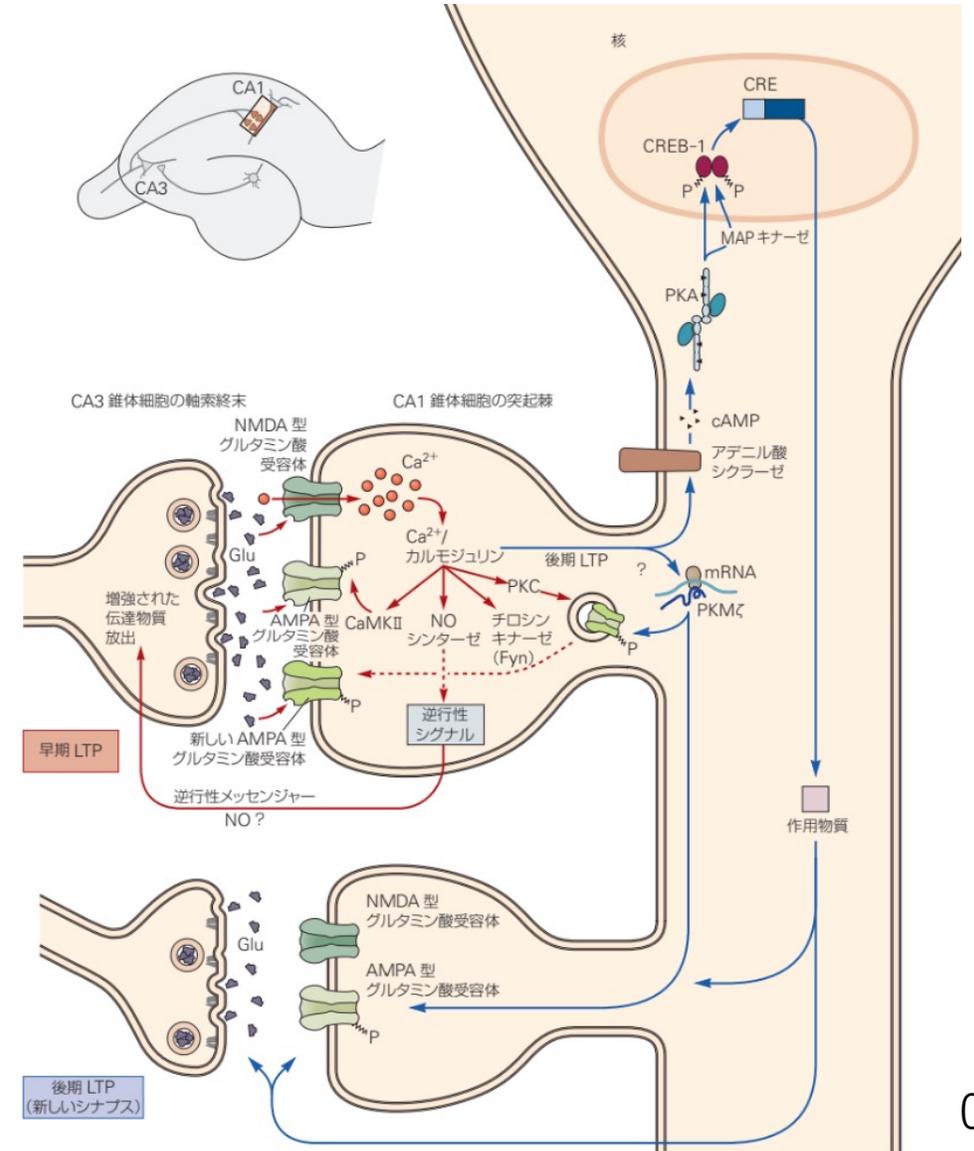
McCulloch-Pittsモデル



パーセプトロン

シナプス可塑性の生物学的メカニズム

- 神経伝達物質：イオンチャネルを直接開閉
 - グルタミン酸、GABA、...
- 神経修飾物質：イオンチャネルの開閉を修飾
 - ドーパミン、セロトニン、アセチルコリン、...
- NMDA受容体による長期増強（LTP）



目次 (1章)

1.1 神経回路の数理理論の発展

~~1.2 環境の状態の推論機能からHelmholtzマシンまで → 割愛~~

1.3 神経細胞の基本特性と学習アルゴリズム

~~1.4 脳の多層構造と双方向性結合の機能 → 割愛~~

1.5 自由エネルギー原理

Hermann von Helmholtzについて



Hermann von Helmholtz (1821-1894)

- Helmholtzの無意識的推論

“人間の視覚は不完全であるため、無意識的に推論を行い、不足した情報を補っているはずである”

ex. 網膜に投影された2次元データから、3次元の世界を推論

- 熱力学、神経科学、流体力学、...

- 熱力学におけるヘルムホルツの自由エネルギー

$$F = U - TS$$

F : 自由エネルギー、 U : 内部エネルギー、 T : 温度、 S : エントロピー

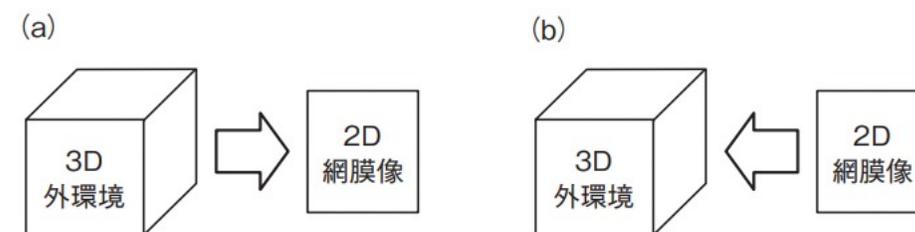


図1 (a)網膜像は外環境の3次元構造に従って決まる。
(b)脳は(a)とは逆に、網膜像から外環境の3次元構造を推論する。

『脳の大統一理論』より引用

自由エネルギー原理

- Fristonは、無意識的推論が自由エネルギー最小化により実現できると提唱した（初期の自由エネルギー原理）
- 2009年に能動的推論という概念が提唱され、自由エネルギー原理は知覚だけでなく行動にも拡張された
- さらに学習なども含めて、様々な脳のメカニズムを自由エネルギー最小化という同一の枠組みで説明できるとFristonは主張しており、「脳の統一理論」として期待されている

ここまでのまとめ（再掲）

1. 知覚：感覚入力から生成モデルに基づいて外界を推論し、知覚が得られる。

この時、自由エネルギーを最小化させる知覚が選択される。

2. 行動：行動により、自分が望む感覚入力を得られるように外界に介入することができる。

これを能動的推論と呼ぶ。この時、自由エネルギーを最小化させる行動が選択される。

3. 学習：知覚・学習と並行して、生成モデルそのものもアップデートされる。

このアップデートも、自由エネルギーを最小化させるように決定される。

このように、知覚・行動・学習は全て自由エネルギーを最小化させるように実現される。

この過程を経て、世界は生物にとって「より予測可能な場所」になる。

本の構成

はじめに

変数記号一覧

今回扱う部分

第1章 自由エネルギー原理を理解するために

- 1.1 神経回路の数理理論の発展
- 1.2 環境の状態の推論機能からHelmholtz マシンまで
- 1.3 神経細胞の基本特性と学習アルゴリズム
- 1.4 脳の多層構造と双方向性結合の機能
- 1.5 自由エネルギー原理

第2章 知覚過程のモデル化

- 2.1 諸行無常の環境を記述する
- 2.2 推論に必要な環境の生成モデル
- 2.3 生成モデルの表現
- 2.4 推論過程をモデル化する
- 2.5 変分Bayes推定とHelmholtzの自由エネルギー
- 2.6 自由エネルギーを具体的に書く
- 2.7 自由エネルギーの最小解の求め方
- 2.8 階層的動的モデルを階層的神経回路で解く
- 2.9 階層的なメッセージパッシング

第3章 パラメータの学習

- 3.1 パラメータの学習と平均場近似
- 3.2 平均場近似の生理学的意味
- 3.3 パラメータの学習とHebbの学習則

第4章 能動的推論と行動決定

- 4.1 能動的推論とは
- 4.2 生成過程と生成モデル
- 4.3 腕の運動
- 4.4 制御状態と行為選択
- 4.5 能動的推論としてのホメオスタシス
- 4.6 感覚減衰
- 4.7 離散システムにおける能動的推論
- 4.8 期待自由エネルギーの導出と行動決定
- 4.9 行動の価値と主体感

第5章 離散系と連続系の統合

- 5.1 読書過程の離散系意思決定モデル
- 5.2 離散系における単語・文の処理
- 5.3 離散系と連続系のインターフェースの詳細
- 5.4 視線制御の神経回路による実装

第6章 運動制御と運動認識

- 6.1 経時的処理切り替えの神経機構
- 6.2 運動制御モデルの例1：ランプへの到達運動
- 6.3 運動制御モデルの例2：系列的な運動
- 6.4 運動制御モデルの例3：手書き文字の生成と認識

第7章 音声コミュニケーション機能

- 7.1 環境のダイナミクスを予測できる生成モデル
- 7.2 階層構造で異なる時間スケールのダイナミクスを作る
- 7.3 コミュニケーション機能への応用
- 7.4 鳥の歌の学習モデルとコミュニケーション
- 7.5 Markovブランケットの意味と役割

参考文献

索引

目次 (2章前半)

- 2.1 諸行無常の環境を記述する
- 2.2 推論に必要な環境の生成モデル
- 2.3 生成モデルの表現
- 2.4 推論過程をモデル化する
- 2.5 変分ベイズ推定とHelmholtzの自由エネルギー

目次 (2章前半)

- 2.1 諸行無常の環境を記述する
- 2.2 推論に必要な環境の生成モデル
- 2.3 生成モデルの表現
- 2.4 推論過程をモデル化する
- 2.5 変分ベイズ推定とHelmholtzの自由エネルギー

一般化座標の導入

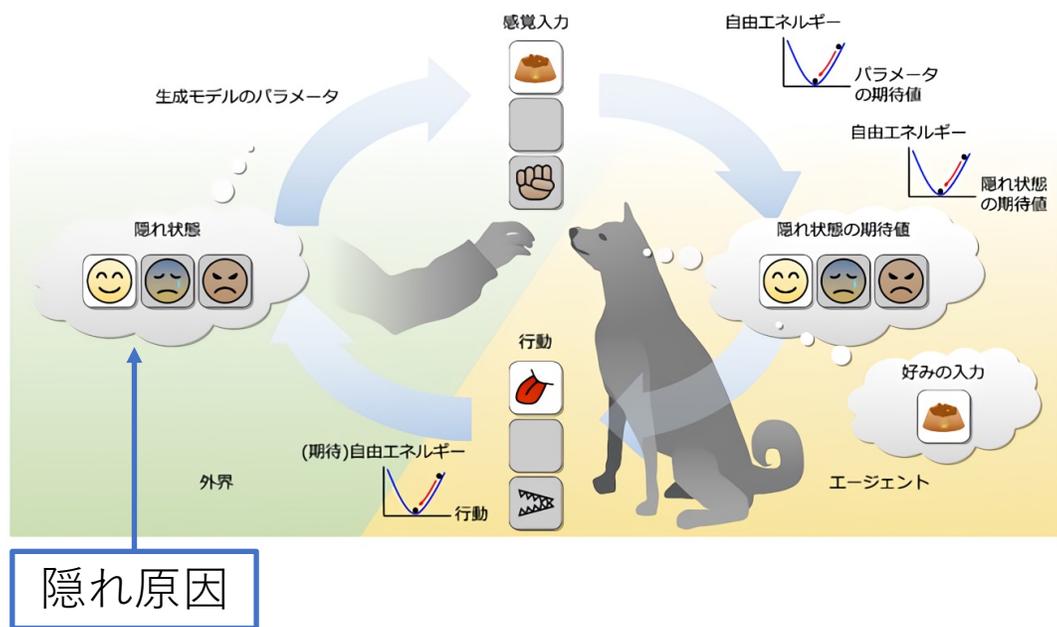
- 一般化座標 $\tilde{x}(t)$ を右のように定義する。

$$\tilde{x}(t) = \begin{bmatrix} x(t) \\ x'(t) \\ x''(t) \\ \vdots \\ \cdot \end{bmatrix} \quad \begin{array}{l} x : \text{隠れ状態} \\ t : \text{時間} \end{array}$$

- 自由エネルギー原理では、一般化座標で表現された状態変数が神経細胞の活動として符号化されていると考える。

ex. 視覚皮質では、光の強度($x(t)$)、その時間変化($x'(t)$)、…が神経活動として符号化されていることが知られている。

隠れ状態、隠れ原因、生成モデルの定義



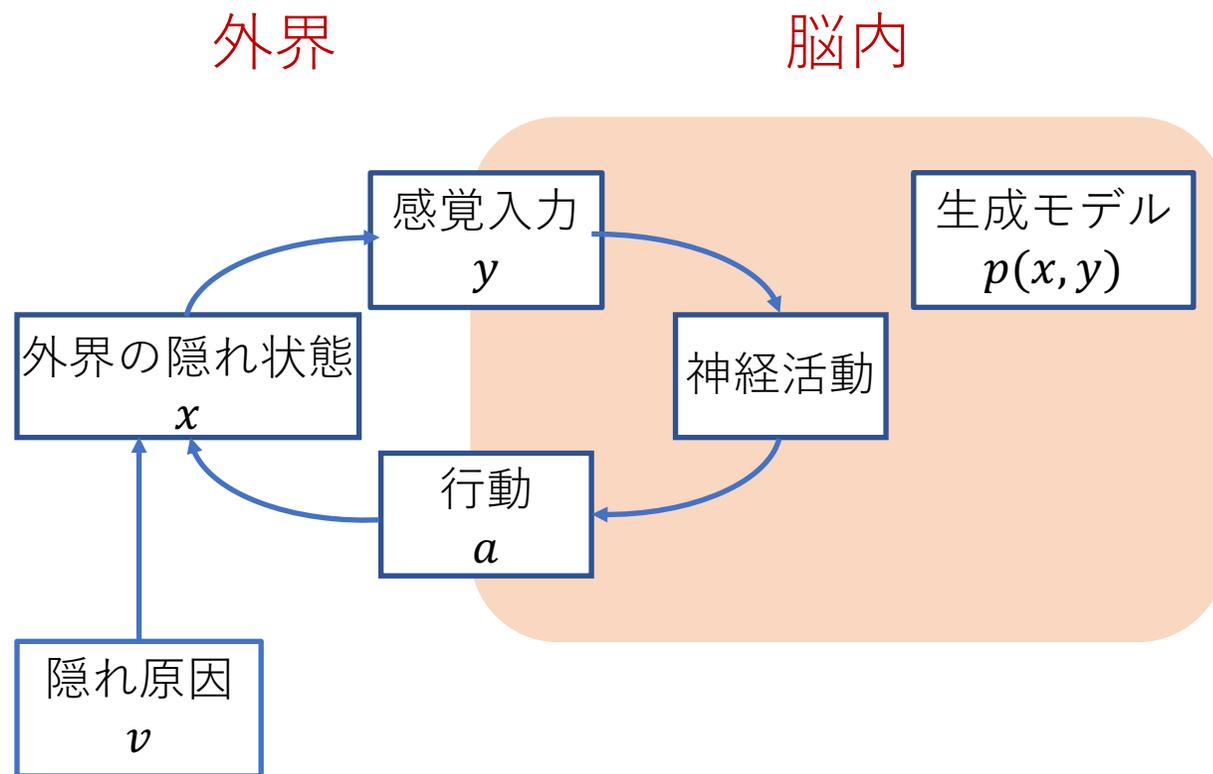
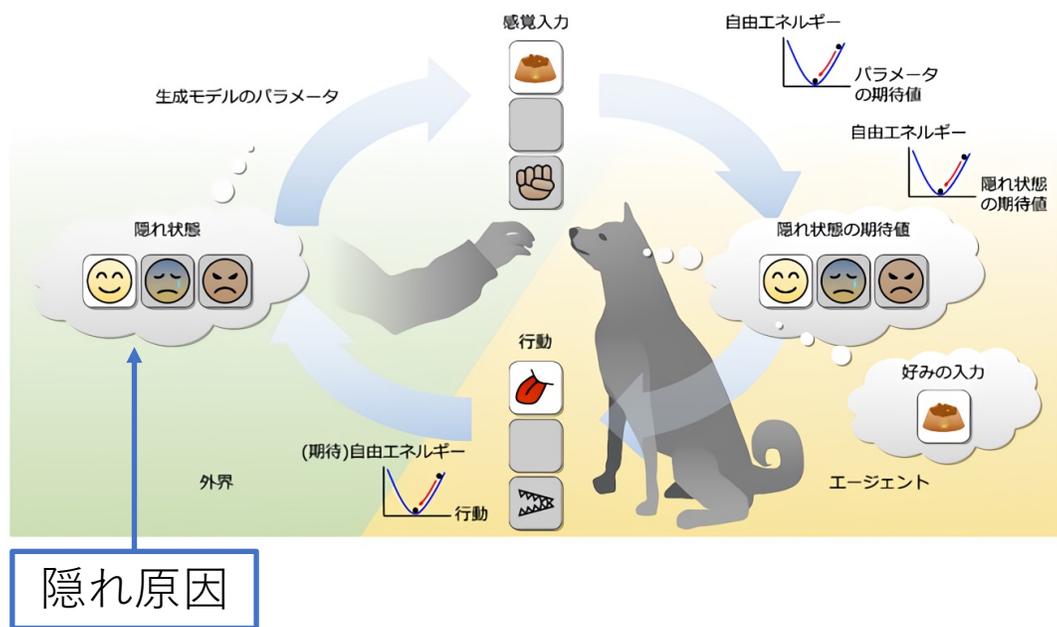
隠れ状態 : 直接アクセスできない外界の状態

隠れ原因 : 隠れ状態に変化を引き起こす要因

生成モデル : 隠れ状態と感覚入力の関係性を記述した

脳内のモデル

隠れ状態、隠れ原因、生成モデルの定義



神経回路は潜在的な統計学者 | 理化学研究所

観測方程式と状態方程式の定義

- 自由エネルギー原理では、以下の2つの式で生成モデルを表す
 - 観測方程式：隠れ状態 x から感覚信号 y が生成される過程を記述した式

$$y(t) = g(x(t), v(t)) + \epsilon_y(t)$$

- 状態方程式：隠れ状態 x の時間変化を記述した式

$$Dx(t) = \begin{bmatrix} x'(t) \\ x''(t) \\ x'''(t) \\ \vdots \end{bmatrix} = f(x(t), v(t)) + \epsilon_x(t)$$

$y(t)$: 感覚信号
 g : 非線形関数
 $x(t)$: 隠れ状態
 $v(t)$: 隠れ原因
 $\epsilon_y(t)$: 観測ノイズ
 D : 微分演算子
 f : 非線形関数
 $\epsilon_x(t)$: システムノイズ

目次 (2章前半)

- 2.1 諸行無常の環境を記述する
- 2.2 推論に必要な環境の生成モデル
- 2.3 生成モデルの表現
- 2.4 推論過程をモデル化する
- 2.5 変分ベイズ推定とHelmholtzの自由エネルギー

自由エネルギー原理とベイズの定理

- 自由エネルギー原理では脳の目的を、「真の事後分布 $p(x|y)$ を知ること」(感覚入力 y のもとで、隠れ状態 x を推論すること)と考える。
- 真の事後分布 $p(x|y)$ は、ベイズの定理を用いて求めることができる。

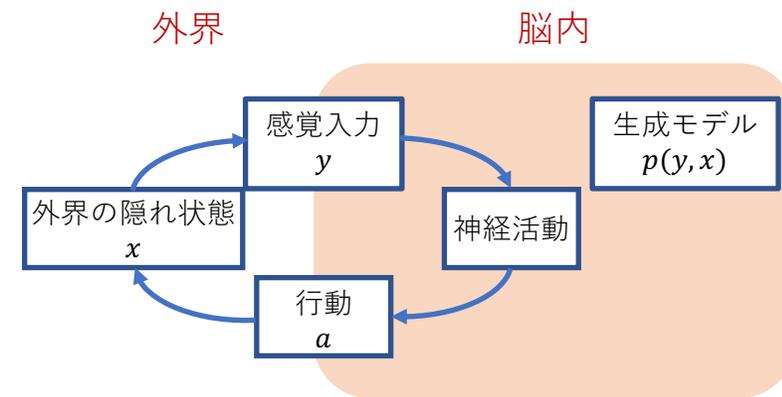
$$p(y, x) = p(y|x)p(x) = p(x|y)p(y)$$

$$p(x|y) = \frac{p(y, x)}{p(y)} = \frac{p(y|x)}{p(y)} p(x)$$

- ここで、 $p(y)$ の計算にはありうる全ての隠れ状態 x_i について総和を取る必要がある。

$$p(y) = p(y|x_1)p(x_1) + p(y|x_2)p(x_2) + \dots + p(y|x_n)p(x_n)$$

- しかしながらこれは現実的に困難だし、脳はそのような計算をしていないとFristonは(おそらく)考えている。



$p(y, x)$: 隠れ状態 x と感覚入力 y の同時分布 (生成モデル)

$p(x)$: 隠れ状態 x についての確率分布 (世界についての事前知識)

$p(y)$: 感覚入力 y についての確率分布 (周辺尤度)

$p(x|y)$: 真の事後分布 (求めたいもの)

$p(y|x)$: 隠れ状態 x のもとで感覚入力 y が生じる確率 (尤度)

変分ベイズ推定とKLダイバージェンス

- そこで新たな分布 $q(x)$ を導入し、これが真の事後分布 $p(x|y)$ をなるべく近似するようになれば良い。

→ 変分ベイズ推定

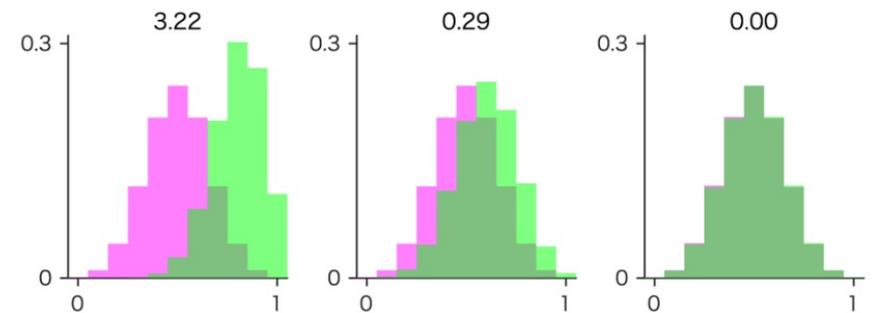
- 2つの分布の類似度を測る指標

→ Kullback-Leibler divergence (D_{KL})

$$D_{KL}(q(x)||p(x)) = \int_{-\infty}^{\infty} q(x) \log \frac{q(x)}{p(x)} dx$$

- $D_{KL} = 0$ は、2つの分布 $p(x)$ と $q(x)$ の一致を意味する。
- 今回求めたいのは $D_{KL}(q(x)||p(x|y))$ だが、直接は計算できない。
(そもそも知りたいのが事後分布 $p(x|y)$)

KLダイバージェンスのイメージ



『自由エネルギー原理入門 改め
自由エネルギー原理の基礎徹底解説』より引用

(変分)自由エネルギー

$$\begin{aligned} \boxed{D_{KL}(q(x)||p(x|y))} &= \int q(x) \log \left(\frac{q(x)}{p(x|y)} \right) dx \\ \text{KLダイバージェンス} & \\ &= \int q(x) \log \left(\frac{q(x)p(y)}{p(y,x)} \right) dx && \text{ベイズの定理} \\ &= \int q(x) \left\{ \log \left(\frac{q(x)}{p(y,x)} \right) + \log p(y) \right\} dx \\ &= \int q(x) \log \left(\frac{q(x)}{p(y,x)} \right) dx + \log p(y) \int q(x) dx && q(x) \text{ は確率密度関数なので} \\ &= \boxed{\int q(x) \log \left(\frac{q(x)}{p(y,x)} \right) dx} - \boxed{(-\log p(y))} && \text{積分すると1} \\ & \quad \text{(変分)自由エネルギー} \quad \text{サプライズ} \end{aligned}$$

(変分)自由エネルギー = KLダイバージェンス + サプライズ

次回は、自由エネルギーの値や最小化の
具体的な計算方法について扱います！

自由エネルギー原理を学ぶ上で参考にした日本語資料

磯村拓哉先生

- 『わかりやすい自由エネルギー原理』（スライド）
- 『[自由エネルギー原理の解説：知覚・行動・他者の思考の推論](#)』（論文）

吉田正俊先生

- 『[自由エネルギー原理入門 改め自由エネルギー原理の基礎徹底解説](#)』（スライド）
- 『[シンギュラリティサロン講演記録「自由エネルギー原理と視覚的意識」](#)』（YouTube）

乾敏郎先生

- 『[脳の大統一理論 自由エネルギー原理とはなにか](#)』（書籍）
- 『[自由エネルギー原理—環境との相即不離の主観理論—](#)』（論文）

おまけ

よくある疑問・質問

- 変分自由エネルギーは、もともと変分ベイズで定義されている値である
(自由エネルギー原理のために生み出された値ではない)
- 式変形により、変分自由エネルギーとヘルムホルツの自由エネルギーは「似た形」になる
→ 熱力学や統計力学との関係性の示唆 (次スライド)
- 脳は自由エネルギーを下げるように設計されているという「解釈ができる」

(変分)自由エネルギーとHelmholtzの自由エネルギー

- 内部エネルギー $U(x; y)$ を以下のように定義する

$$U(x; y) = -\log p(x, y) = -\log p(y|x) - \log p(x)$$

- このとき、自由エネルギー F は以下のように記述される

$$F(q, p; y) = \int q(x) \log \left(\frac{q(x)}{p(y, x)} \right) dx$$
$$= \boxed{\int q(x) U(x; y) dx} - \boxed{\left(- \int q(x) \log q(x) dx \right)}$$

認識分布に基づく
内部エネルギーの事後期待値

エントロピー

- これは、Helmholtzの自由エネルギー $F = U - TS$ と「似た表現」と考えることができる

→ 熱力学や統計力学との関係性の示唆