解 説

自由エネルギー原理と視覚的意識

吉 田 正 俊^{*1}, 田 口 茂^{*2}生理学研究所認知行動発達機構^{*1}, 北海道大学大学院文学研究科^{*2}

Free Energy Principle and Visual Consciousness

Masatoshi Yoshida^{*1} and Shigeru Taguchi^{*2}National Institute for Physiological Sciences^{*1}
Graduate School of Letters, Hokkaido University^{*2}

概要

フリストンの自由エネルギー原理では、外界に関する生成モデルと現在の認識から計算される変分自由エネルギーを最小化するために、1) 脳状態を変えることによって正しい認識に至る過程 (perceptual inference) と 2) 行動によって感覚入力を変えることによって曖昧さの低い認識に至る過程 (active inference) の二つを組み合わせていると考える。本解説の前半では自由エネルギー原理について、我々が視線を移動させながら視覚像を構築してゆく過程を例にとって、簡単な説明を試みた。本解説の後半では、このようにして理解した自由エネルギー原理を元にして「自由エネルギー原理と現象学に基づいた意識理論」を提唱した。この理論において、意識とは自由エネルギー原理における推測と生成モデルとを照合するプロセスそのものであり、イマココでの外界についての推測と非明示的な前提条件の集合である生成モデルとが一体になって意識を作り上げている。この考えはフッサール現象学における意識の構造についての知見と整合的である。

1. はじめに：視覚的経験の構造

1.1 視覚的経験の射映の構造

我々の視覚経験そのものとは正確にはどのようなものだろうか？ 外から与えられた感覚的刺激が受動的に処理されて脳内に像を結ぶ（表象を形作る）という見方は、手近な考え方としてしばしば援用されるものである。これに対し、我々自身のはたらき（広い意味での行動）を通してのみ、我々にとって視覚経験が成立すると考えるのがフッサール（1859-1938）によって創始された現象学の考え方である。我々が見ているものは、つねに自分の身体の運動可能性と結びつけられる仕方で経験されている。たとえば私が身体を右に動かせば、眼前の箱のこれまで見えていなかった右側の側面が見えてくる。別の仕方で身体を動かせば、同じ箱のまた別の側面が見えてくる。我々が視覚的に捉えている「物」とは、実質上、このような様々な現われを

システマティックに統合したものと言える。

物の知覚に限らず、我々にとっての視覚経験は一般に、身体の運動可能性のシステムに相関する仕方であり、「こうすればこうなる」という現われのシステマティックな変化によって成立している。視覚経験とは、外界の単なる写しでもないし、単に受動的な過程でもない。身体をもった主体が環境に働きかけ（介入し）、そこに予測→確認→修正→予測→……というループが形作られるとき、視覚経験に対応するものが立ち現われるのである¹⁾。後で論じる感覚運動随伴説や自由エネルギー原理を意識論として考え直す際には、このような現象学的な経験の見方が考え方の枠組みとして決定的に重要となる。

現象学のこうした視覚観は、次のような仕方でも敷衍することができる。ある視覚シーンを見ているときのことをエルンスト・マッハの「感覚の分析」では、我々の視覚が自分の眼や鼻によって輪郭付けられた一人称的な像であり、自分に近いところ、遠いところというように遠近法的描像をもっていることを図 1a のよう

^{*1} 〒 444-8585 愛知県岡崎市明大寺町字西郷中 38

^{*2} 〒 060-0810 北海道札幌市北区北 10 条西 7 丁目

図1 Active vision: 我々の視覚像は視線の移動によって構成されている。文献²⁾の図を改変して作成。

に表現した²⁾。たしかに我々の視覚像は写真のようなものでないことをこの図は的確に表現している。しかし、この図には「我々の視覚経験そのもの」というには妥当でないことがある。それは、我々の網膜の視細胞は中心窩以外では密度が低いため、我々の視力は視線の中心以外では低いということが考慮されていないということだ。つまり我々に与えられている視覚経験は正確には図 1b のようになっているはずだ (片眼をつぶって一点を見ながら確認してみしてほしい)。しかし我々は自分の視覚経験は図 1a のようになっているようにいわば錯覚している。これはなぜか。それは我々が眼を急速眼球運動 (サッカード) によって素早く動かしながら、中心視野の情報をサンプルすることによってひとつの視覚シーンを構成しているからだ (図 1c)。このような視覚観を active vision と呼ぶ³⁾。

1.2 知覚の感覚運動随伴性説

このように視覚がアクティブに構成される特性を持つこと自体を知覚成立の条件だと提案したのが Alva Noë による感覚運動随伴性 (Sensorimotor contingency, 略して SMC) 説⁴⁾。

標準的な視知覚のモデル (表象説) では、「見る」ということは脳内に外界に対応した表象を作ることであると捉える (図 2 左)⁴⁾。しかし SMC 説においては「見る」とは視覚対象にどのように働きかけることができるかを知ることであると主張する (図 2 右)⁴⁾。「見る」とはある主体が感覚運動随伴性に習熟することによって可能となる探索的な活動である。定義するならば、SMC 説とは、「知覚する能力は部分的には、知覚者の感覚運動随伴性に関する知識 (知覚者の動きによって感覚刺激がどのように変化するかを実際的な能力として把握していること) によって構成される」⁴⁾ということになる。

感覚運動随伴性の具体例を挙げてみよう。例えば、私たちが対象に向かって近づくと対象の網膜像は大きくなるが、対象が大きくなったわけではないことを知っ

図2 表象説と SMC 説。O'Regan の発表スライド⁵⁾を元に作成。

ている。このような感覚と運動の関係が感覚運動随伴性だ。いま繰り返し「知識」「知る」という表現をしたが、これは命題的な知識 (know-that) ではなくて技能的な知識 (know-how) として捉える必要がある。これは記憶にも命題的な記憶 (宣言的記憶) と技能的な記憶 (手続き記憶) があるのに対応している。感覚運動随伴性の習熟とは、喩えれば自転車の動かし方の知識ではなくて、自転車に乗れるようになることだ。

SMC 説は、表象説では説明できないいくつかの現象についてうまく説明をすることができる。たとえば白内障などによる生まれつき視覚障害を持っている患者は、開眼手術を行った直後でも、視覚入力があるにもかかわらず、視覚経験はぼやけたものしかない (経験盲)。これは開眼手術の直後にはまだ新たな感覚運動随伴性に習熟していないからと説明できる⁴⁾。

SMC 説に対する批判には複数あるがここでは 3 つ挙げたい。

(1) 「夢」議論: 「我々は夢を見ているときはある種の経験を持っているが行動は伴っていない。ゆえに知覚経験に行動は必須ではない。」この議論は感覚運動随伴性説がオンラインでの行動を必要とすると主張しているという誤解に基づいている。夢を見る我々はそれまでの経験で、歩いたら視野はどのように変わるか、手を伸ばしたら物に手が届くかといった感覚運動随伴性の技能知をすでに獲得している。ゆえに夢を見るときにはこの感覚運動随伴性の獲得によってある種の意識経験が可能となるといえる。

(2) 「必要条件と充分条件」議論: 「感覚運動随伴性は知覚経験にとっての必要条件であって充分条件ではない。」つまり、感覚運動随伴性はたしかに必要条件ではあるかもしれないが、イマココで知覚経験を因果的に充分に成立させるメカニズムではないのではないかと批判がある。この批判については本解説の後半で反駁を加える。

(3) 「実験の検証に耐えうる定式化された理論になっ

ていない。」この批判は確かに当たっている。本解説の後半では、自由エネルギー原理が SMC 説の定式化となっているのではないかということを議論する。

SMC 説を鍛え上げて視覚的意識の理論にするためには、自由エネルギー原理は重要な役目を果たすのではないか、このような観点から以下では自由エネルギー原理について簡単な例を挙げながら紹介をしてゆく。この目的のため本稿では、自由エネルギー原理の説明には大幅に簡略化した例を用いることで、数理的な詳細には深入りせずにそのエッセンスをざっとつかめるようにするというを目的としたい。本説明は Friston et al. (2012)⁶⁾ に準拠し、McGregor らの論文⁷⁾ を元にした表記を採用している。自由エネルギー原理についてのより正確な説明、数理的な詳細については本特集号の他の記事、とくに乾 (2018)⁸⁾、磯村 (2018)⁹⁾、大羽 (2018)¹⁰⁾ を参照していただきたい。

2. 自由エネルギー原理

2.1 自由エネルギー原理とは

自由エネルギー原理 (free-energy principle, 以下 FEP と呼ぶ) とはなにか。MRI の解析ソフト SPM の開発者として有名な UCL の Karl Friston が提唱した以下の考え方だ。

FEP とは「いかなる自己組織化されたシステムでも環境内で平衡状態であるためには、そのシステムの (情報的) 自由エネルギーを最小化しなくてはならない」という規範的な原理であり、「適応的なシステムが無秩序へ向かう自然的傾向に抗して持続的に存在しつづけるために必要な条件」¹¹⁾ のことである。よって FEP は生物が適応的に存在し続けるための条件であり、視覚意識についての理論そのものではない。

FEP が「原理」であると Friston が主張する意味は何か。脳に関わる理論には、ベイズ脳理論、予測符号化、知覚における情報量最大化、運動制御における最適化理論、シナプス可塑性におけるヘブ則などそれぞれのドメインに特化した理論があるが、FEP は脳についてのこれらの諸理論を統一的に説明する「原理」であり、これまでの理論を置き換えるためのものではない、と Friston は書く¹¹⁾。

自由エネルギー原理の定式化では図 3 の図式を想定している。主体 agent は外部世界と感覚入力 s および行動 a によって相互作用している。感覚入力 s は agent が直接アクセスできない感覚入力の原因 $\mathbf{X}$ によって引き起こされる。この $\mathbf{X}$ と s の関係は生成過程 $g(\mathbf{X}, s)$ として確率的に表現される。行動 a は感覚入力の原因 $\mathbf{X}$ を変えることによって結果として感覚

図3 FEP が想定する主体と外部世界の図式

入力 s を変化させる。たとえば「目の前のコップに手を伸ばして位置を動かす」といったことだ。また、行動 a は感覚入力 s に直接影響をおよぼすこともできる。たとえば「眼球を動かすことによって網膜を活動させる視覚情報が変化する」ことが可能になる。

Agent は生成過程 $s = g(\mathbf{X})$ そのものを知ることとはできないが、生成過程をモデル化した生成モデル $s = g(x)$ を獲得している (生成過程の方は太文字で別ものであることに注意)。ここでの x は脳内に表現された外界の原因 $\mathbf{X}$ についてのモデルである。外界の原因 $\mathbf{X}$ そのものには agent がアクセスすることはできない。生成モデルとはいわば外界にある物事についての知識であり、物理法則で規定された時間的変化も含んでいる。たとえば「左から飛んできたボールは慣性の法則でまっすぐ進み、視野の右に向かう」といったものだ。このセクションの説明では説明の単純化のため、 g の関数形は考慮せずに¹¹⁾ 離散的な同時確率分布 $p(x, s)$ を扱うことにする。以降 $p(x, s)$ を「生成モデル」と呼ぶことにする¹²⁾。

また agent はいまここで自分のいる外部世界の状態 x についての推測 $q(x)$ を常に作り上げている¹³⁾。たとえば「今眼前にあるのは蛾ではなく蝶だ」という認識だ。現在の推測 $q(x)$ および生成モデル $p(x, s)$ は確率的に表現されており、それぞれが不確定性を持っている。

現在の推測 $q(x)$ と生成モデル $p(x, s)$ は具体的には脳内では脳活動 b によって表現されている。つまり脳活動 b は $q(x)$ および $p(x, s)$ の媒介変数であり、 $q(x)$ および $p(x, s)$ はどちらも実態としては脳活動 b による関数として表現される。しかしこの解説記事では脳

¹¹⁾ 正確な表現では、この関数の形を決めるためのパラメータ θ が生成モデルには入っている。

¹²⁾ Friston 論文の表現では “generative density”。

¹³⁾ Friston 論文の表現では “recognition density”。

活動のことは考えずに, $q(x)$ および $p(x, s)$ という抽象的な表現が agent の中に推測の世界 (図 3 の緑部分) として存在するように表記したうえで, それが認識の過程でどのように推移するかを見てゆく。

このような設定において情報理論的な意味での変分自由エネルギーは直感的には「現在の推測 $q(x)$ 」から「生成モデル $p(x, s)$ 」までの距離, つまり予測誤差の大きさを表現する。

$$F(q, s) = \sum_x \{q(x) * \log \frac{q(x)}{p(x, s)}\} \quad (1)$$

自由エネルギー原理ではこの $F(q, s)$ が小さくなる方向に agent が変化する。実際に agent にとって可能なことは q および s を変化させることだ^{†4}。 $F(q, s)$ を下げる過程を理解するためには q または s のどちらかだけを変化させてみるのが有効だ。そこでベイズの定理 $p(x, s) = p(x|s)p(s) = p(s|x)p(x)$ を用いて式 (1) を二通りの式 (2), (3) のように変形してみる。

$$F(q) = \underbrace{D_{KL}(q(x)||p(x|s))}_{\text{Relative entropy}} + \underbrace{-\log(p(s))}_{\text{Sensory surprisal}} \quad (2)$$

$$F(s) = \underbrace{\sum_x \{-q(x) * \log(p(s|x))\}}_{\text{Uncertainty}} + \underbrace{D_{KL}(q(x)||p(x))}_{\text{Bayesian Surprise}} \quad (3)$$

ここで $D_{KL}(A||B)$ は確率分布 A から確率分布 B までの KL divergence を表す。この値は 0 以上のスカラー値で, 確率分布 A と確率分布 B が完全に一致するときのみ 0 となる。

すると自由エネルギー F を下げるためには, 以下の (a), (b) の 2 つの方策が可能であることがわかる。

(a) q を変える (s は固定) = 認識 (Perceptual inference): 式 2 において, 「現在の推測 $q(x)$ 」を変えて「生成モデル $p(x|s)$ 」と一致させたとき (=正しい認識), $Relative\ entropy = 0$ となり, F は最小になる。 $Sensory\ surprisal$ は s が固定されているので定数となる。

(b) s を変える (q は固定) = 行動に基づく認識 (Active inference): 式 3 において, 行動 a によって感覚

入力 s を変える (例: 対象に眼を向ける)。これによって「現在の推測 q 」をより選択的に生み出す感覚入力 s をサンプルしたとき, $Uncertainty$ が最小化し, F が最小になる^{†5}。 $Bayesian\ Surprise$ は q が固定されているので定数となる。

FEP によれば, 我々を含む agent はこの二つの方策を組み合わせる自由エネルギーを最小化していることになる^{†6}。ここからは具体例を用いてステップごとに説明してゆくことにしよう。

2.2 レベル 1: 認識 (perceptual inference)

まずは図 3 の図式の上半分だけ, 行動を考慮しないケースを考えてみよう。暗闇から照明がついたら図 4a(i) にあるようにあなたの眼前にある視覚シーンが写った。あなたの視線は赤十字のところにあって, すぐに見えているものが蝶であることに気づいた。いま見えたのはたまたま蝶だったけれども, 蛾がいる可能性もあった。つまり図 4a(ii) のようなケースだ。つまり「蝶だ」という認識とは (蛾という他の可能性を排除した) 「ほかでもない蝶だ」という認識のことであり, 「蝶だ」という表象を持つということはほかでもなく蝶だという選択を前提としている。

この世界についての知識はこれまでの経験からあなたの頭のなかに生成モデル $p(x, s)$ として獲得されている (図 4b)。この簡略化された世界では, 世界の原因は蝶 (x_1) か蛾 (x_2) しかいないものとする。それによって与えられる感覚入力は真っ暗闇の s_1 , 蝶っぽいものがある二次元画像 s_2 , 蛾っぽいものがある二次元画像 s_3 の三通りだけとなっている。このような確率的関係は同時確率として生成モデル $p(x, s)$ に表現されている。生成モデル $p(x, s)$ はすでに学習済みで固定とみなす。それではこのときの認識過程を順を追って見てみよう (図 4c)。

時点 t_1 : まず最初は真っ暗闇の感覚入力 s_1 が与えられている状態で, このとき蝶と蛾のどちらがいるのかまったくわからないので, 現在の推測は $q(x_1) = 0.5$, $q(x_2) = 0.5$ となる。生成モデルから事後分布

^{†5} Friston の元論文では $Accuracy (= -Uncertainty)$ が使われるが, $Accuracy$ は今回の説明で使う例では常に負値となるので, 明快さを追求して本記事では $Uncertainty$ の語で統一することにする。

^{†6} この記事で挙げる例では生成モデル $p(x, s)$ は学習済みで変化しないものとしている。じっさいには環境の変化に合わせて生成モデル $p(x, s)$ じたいは変化しうる。このような可塑的变化も FEP においては F を下げるものとして貢献する。このあたりについては本特集号の他の記事を参照していただきたい。

^{†4} ここで q を変えるとは, $q(x)$ という関数の形を変えるという意味であり, x が変わるわけではない。 $q(x)$ は確率分布なので, 和が 1 である拘束条件のもとでその形を変える。

図4 (a) 認識 (perceptual inference) の具体例. (b) 生成モデル $p(x, s)$. (c) Perceptual inference 中の推測および自由エネルギーの動態.

$p(x|s)^{†7}$ を計算することができて, $p(x_1|s_1) = 0.5$, $p(x_2|s_1) = 0.5$ となる. よって $Relative\ entropy = 0$ であり, 自由エネルギー $F = 2.32$ (単位は bit) と計算することができる.

時点 t_2 : 照明がついて感覚入力 s_2 が提示された直後はまだ推測 $q(x)$ がアップデートされていない直前の状態を引きずっている. このとき推測 $q(x)$ と生成モデルから計算された事後分布 $p(x|s)$ は一致していない, つまり $Relative\ entropy > 0$ となっている.

時点 t_3 : よくよく見るあいだに見ているものは蝶だと気づいた. このとき推測 $q(x)$ をアップデートして

生成モデルの事後分布 $p(x|s)$ と一致させることができている. つまり $Relative\ entropy = 0$ となっている. このとき正しい認識に至ったと説明することができる. これが Perceptual inference だ.

ひとつ注意すべきは, このとき認識 $q(x)$ は, $q(x_1) = 0.9$, $q(x_2) = 0.1$ となっており, 不確定性を残している. それは現在の設定での生成モデルから計算される事後分布 $p(x|s)$ が $p(x_1|s_2) = 0.9$, $p(x_2|s_2) = 0.1$ となっているからで, もしそれにもかかわらず $q(x_1) = 1.0$ となったとしたら, それは FEP を最小化できていない.

ここで見た例はヘルムホルツの視覚観¹²⁾の計算論的な定式化と捉えることができる. ヘルムホルツ的視覚

^{†7} Friston 論文の表現では “posterior density”.

図5 (a) 行動を伴う認識 (active inference) の具体例. (b) 生成モデル $p(x, s)$. (c) Active inference 中の推測および自由エネルギーの動態.

観によれば, agent は直接アクセス不可能な世界を知るために感覚入力から「無意識的推論」を行う. ちなみにここでの「無意識的」とは対象を意識していないという意味ではなく, 推論の過程を我々は意識していないということを意味している.

このような認識の捉え方は FEP が最初ではない. Mumford の予測符号化¹³⁾, 川人, 乾の順逆光学モデル^{14, 15)}から始まる多くの研究がある. 受動的な認識だけを考える場合, 予測符号化だけで十分な説明が可能だ. FEP が重要であるのはさらに行動を伴う認識過程 (Active inference) も説明可能な枠組みを提供した点にある.

2.3 レベル 2: + 行動 (active inference)

今度は図 3 の図式の上下両方, 行動を考慮したケースを考えてみよう. 図 1 でも説明したように我々の視力

は中心窩でのみ高いため, 眼前の視覚シーンは図 5a(i) のようにはくっきりとは見えない. 図 5a(ii) にあるように, 左下に視線 (赤十字) を向けている状況を考えてみよう. 右上になにかがあることは分かるが, 蝶か蛾かは十分に判断がつかない. 右上に視線移動をしてみたところ, 蝶であることが判明した, そういう状況だ.

この世界での生成モデル $p(x, s)$ はさきほどと同様だが, 感覚入力 s_1 のときに事後分布 $p(x_1|s_1) = 0.6$, $p(x_2|s_1) = 0.4$ となっており, 多少は情報がある点で異なっている (図 5b). それではこのときの認識過程を順を追って見てみよう (図 5c).

時点 t_1 : まず最初は左下を見ているときの感覚入力 s_1 が与えられた直後で, このとき蝶と蛾のどちらがいるのかまったくわからないので, 現在の推測は $q(x_1) = 0.5$, $q(x_2) = 0.5$ となる. 生成モデルから事後分布 $p(x|s)$ を

計算することができて、 $p(x_1|s_1) = 0.6$, $p(x_2|s_1) = 0.4$ となる。よって *Relative entropy* > 0 であり、自由エネルギー $F = 2.35$ と計算することができる。

時点 t_2 ：感覚入力 s_1 に基づく推測をアップデートすることによって $q(x_1) = 0.6$, $q(x_2) = 0.4$ となった。このとき推測 q と生成モデルの事後分布 $p(x|s)$ は一致して *Relative entropy* $= 0$ となる。ここまでは *perceptual inference* のときとまったく同じ。

時点 t_3 ：これ以上自由エネルギーを下げるためには、推測 $q(x)$ だけではなく感覚入力 s を変化させる必要がある。これは行動 a を選択することによって可能となる。いまの単純な世界では行動の選択肢は「視線を動かさない」 a_1 と「視線を右上に動かす」 a_2 だけにしておこう。ここで行動 a_2 を選択することによって *Uncertainty* を下げることができる。具体的には目を動かす前の生成モデルから計算された $p(s|x)^{18}$ が $p(s_1|x_1) = 0.23$, $p(s_1|x_2) = 0.16$ であったところを、現在の $q(x)$ から予測される候補 x_1 をより選択的にサンプルするような感覚入力 s_2 を選択することによって $p(s_2|x_2) = 0.69$, $p(s_2|x_1) = 0.08$ であるように変える。これによって *Uncertainty* は 2.30 から 1.75 に下がり、自由エネルギー F は 2.32 から 1.77 まで下がった。

時点 t_4 ：ふたたび *perceptual inference* で *relative entropy* を下げる余地ができた。推測 $q(x)$ をアップデートすることによって生成モデルの事後分布 $p(x|s_2)$ と一致させた。つまり *Relative entropy* $= 0$ となっている。これによって自由エネルギー F は 1.32 まで下がった。

ここまですべてをまとめよう。*Perceptual inference* では推測 $q(x)$ を生成モデルからの事後分布 $p(x|s)$ と一致させることによって自由エネルギー F を下げた。*Active inference* では行動 a によって感覚入力 s を変化させて *Uncertainty* を下げることによって自由エネルギー F を下げた。FEP は視線を動かしながら認識をアップデートさせてゆく過程、つまりは *active vision* を「*perceptual inference* と *active inference* を組み合わせる感覚入力 s の原因 x を推測する過程」として整合的に説明することに成功している。

2.4 レベル 3：+ 反実仮想 (counterfactual processing)

先のセクションの説明ではひとつごまかしがあった。それはなにかというと、行動 a の結果が必ずしも蝶らしきものが映った二次元画像 s_2 だとは限らない(図 6a(ii))

¹⁸ $p(s|x)$ は生成モデル $s = g(x)$ の確率的表現なので、図の中では「生成モデル」と呼んでいる。

ということだ。視線を右上に向けてみたらじつは蛾だった(図 6a(ii))ということも起こる。行動は賭けだ。時点 t_2 での事後分布が $p(x_1|s_1) = 0.6$, $p(x_2|s_1) = 0.4$ であることから、蛾である可能性も 40% あったことがわかる。

蝶の場合と蛾の場合の両方について自由エネルギー F を計算してみよう(図 6b)。行動 a によって感覚入力 s_2 をサンプルしたときは行動の直後に F が下がった(1.77)。一方で行動 a によって感覚入力 s_3 をサンプルした場合、予想していた原因 x_1 (蝶)を反証する感覚入力 s_3 をサンプルしたため、*Uncertainty* は上昇し、 $F = 2.40$ とかえって大きくなった。このことからわかるように、ある行動が自由エネルギーを下げるかどうかは結果を見るまでは保証されていない。

しかしこの二つのケースから F の期待値を計算してみると、 $E(F(q, s)) = 2.02$ と下がっていることがわかる。つまり行動 a_2 が選択されたのは F の期待値が下がるから、と説明することができる。つまり agent は将来平均的に F を下げることが期待される行動 a を選択している。

ここで重要なことは、この図式は行動の結果の予測を含んでいるという点だ。改めて強調すると、図 3 の図式の右中にあるような「現在の推測 $q(x)$ 」だけでなく、右下の「将来の予測 $q(x')$ 」を生成している。このことは、まだ実現していない行動 a_2 によって蝶という原因 x_1 を予測しているという意味で反実仮想的な推論を行っていることを示している。

反実仮想的な将来の予測 $q(x')$ を考慮した上で自由エネルギーの時間推移を見てみよう(図 6c)。行動 a_2 を選択した時点で将来の予測 $q(x')$ を生成している。すると行動選択の段階で自由エネルギーの期待値 $E(F(q, s))$ は 2.02 まで下がることになる。

反実仮想的な将来の予測の生成は神経生理で知られている *pre-saccadic remapping* の現象¹⁶⁾とよく似ている。後ほど議論するが、脳内での将来の予測の生成は意識的な情報処理と関わっているかもしれない。

2.5 このセクションのまとめと補足

以上のようにして FEP は *perceptual inference* と *active inference* とを組み合わせた自由エネルギーの最小化という観点から、我々が視線移動をしながら認識をアップデートしてゆく *active vision* の過程を整合的に説明することに成功している。また、反実仮想的な推論を考慮すると、自由エネルギーの最小化は行動の不確定性を考慮した期待値でのみ成立することを示した。

本セクションでは FEP の骨格を説明するために、非

図 6 (a) 視線を右上に向けてみないと蝶か蛾かはわからない. (b) 行動の結果を考慮した自由エネルギーの期待値 $E(F(q,s))$. (c) 反実仮定の過程を考慮した自由エネルギーの時間推移.

常に限定された環境について、大幅に簡略化した説明を行った。FEP の詳細については本特集号の他の記事を参照していただきたい。この例で示したような視線移動はこれまで視覚サリエンスに基づいた行動選択としてモデル化されてきた。FEP における視覚サリエ

ンスの扱いに関しては付録 A.1 において補足を行っている。

本セクションでは agent による認識や行動が結果として自由エネルギー F を下げているということを図示することを目的とした。しかしこれは agent が F の値

をモニターしているという意味ではない。Agent が実際に行っているのは、脳の階層構造を用いた予測誤差および予測分散の最適化であり、その結果として $q(x)$ や s が変化する。この過程で F が最小化していくところを図示した。脳の階層構造を用いた予測誤差および予測分散の最適化については、本特集号の他の記事を参照していただきたい。

3. FEP に基づいた意識の理論

ここまで説明してきたことからわかるように、FEP 自体は「意識についての理論」そのものではない。しかし FEP の枠組みの上で意識が可能となる条件を議論できるならば、FEP は意識の理論にとって必須であると言えるだろう。何人かの研究者はその条件が反実仮想であると考えている。

3.1 反実仮想と意識

たとえばサセックス大学の Anil Seth は反実仮想が知覚経験の持つ「存在感」(perceptual presence) という特性に重要であると主張している¹⁷⁾。存在感とは、知覚の content が主観的にリアルであると感じること、存在していると感じることを指す。たとえば、トマトを見たとして、我々は見えない面も確かに存在していると感じる。これは現象学における射映構造と地平の考えにきわめて近い。現象学的に言えば、見えない面のリアルさは、「こうしたらこう経験されるはずだ」という蓋然性の高い行為的知のシステムに対応している。

Seth は存在感の大きさは FEP がどの程度まで深く反実仮想をできるか（「反実仮想的な豊かさ」 counterfactual richness）によって決まると主張している¹⁷⁾。FEP の枠組みでは、トマトの見えていない部分も含めた上で、これに手を伸ばして掴んだり、引き寄せて裏側を見る、といった潜在的な行動が可能であるということが、トマトの裏面を aware しているということになる。SMC の習熟には「(まだ行っていない) ある行動がどのような感覚入力を引き起こすか」という反実仮想的な知識が必要である。

Friston は反実仮想に基づく「時間的な厚み」が意識に必要であると提唱している¹⁸⁾。FEP においては、将来の感覚入力 s とその原因 x の推定のためには、実際に行わなかった行動 a によってどのように感覚入力 s とその原因 x が起きるかを予測する反実仮想的な推論が必要になると考える。このような将来への予測、そして過去の結果への後付的な予測 (postdiction) の時間幅を「時間的厚み」(temporal thickness) と表現している。そのうえで、「意識が生まれるのは、外界についてのサプライズを最小化する inference が「実行さ

れていない行動が引き起こす未来の感覚入力 s とその原因 x についての反実仮想的な推測」を持つことができるような時間的厚みを持っているときだ。」と主張している¹⁸⁾。

Friston¹⁸⁾や Seth¹⁷⁾の理論では、反実仮想は「行動によって将来世界の状態がどのように変わるかを推測する能力」として捉えられている。しかしここで本当に必要となる反実仮想が可能になるためには、介入による因果推論¹⁹⁾を実装する必要があるが、これはまだ実現していない¹⁹⁾。

Friston¹⁸⁾や Seth¹⁷⁾は反実仮想を意識の有無を決めるものとして議論している。一方、著者たちは FEP における推測 $q(x)$ と生成モデル $p(x, s)$ の距離を近づけることそのものが意識に関わっていると考えている。本記事の以降の部分では我々の提案する意識説について説明する。そのため、セクション 1.2 で言及した意識の感覚運動随伴性 (SMC) 説および表象説に戻って、FEP の意義を考えるところから議論を始めたい。

3.2 FEP は SMC 説と整合的である

SMC 説についてあらためてパラフレーズすると、「環境との ongoing の相互作用からなる流れのパターンから出来ているスキル」である。「赤い」という経験は「赤いものとの相互作用が明らかにしてゆく法則性に習熟し、それを実践すること」となる。FEP との関連で考えるならば、FEP では agent と外部世界の ongoing な相互作用の重要性を強調している¹⁷⁾。

FEP における生成モデル $p(x, s)$ とは、たとえば眼を動かしたら、どのように視覚入力に変化するか、といった世界と交互作用する agent との法則的關係についての知識であった。これはまさに感覚運動随伴性の習熟そのものと言えるだろう。SMC 説における感覚運動的知識への習熟について、FEP ではそれが生成モデル $p(x, s)$ の確立、維持であるという計算論的な裏付けを提供することができるようになった。これはセクション 1.2 で挙げた SMC 説への批判 (3) 「SMC は定式化された議論になっていない」という問題点につい

¹⁹⁾ Friston によれば¹¹⁾、FEP で駆動する agent は感覚入力 s から外的世界にあるその原因 x を推定しているというのがこれは正確な意味での因果推論ではない。FEP では生成モデル $p(x, s)$ 、つまり s と x の同時確率という「観察」に基づいて x の推定を行っているのに過ぎない。因果推論¹⁹⁾とは本質的には「世界にある介入をした場合の結果」と「介入しなかった場合の結果」の比較である。しかし両者は同時には存在し得ないため、代わりに「実際に介入したときの結果」と「介入しなかったら起きたであろう仮想的な結果」の比較を行う。よって因果推論には介入と反実仮想が必要となる。

ての答えとなっている。

3.3 FEP は表象説とも整合的である

「意識経験を持つとは、脳内に内的表象を持つことである」という意識の表象説は、脳は外部世界について写真のようなコピーを持っているわけではないという点で批判されてきた²⁰⁾。しかし現在の表象主義的な考えではそのような古典的な考えは取っていない。たとえばミリカンの生物学的意味論²¹⁾では、表象の内容を決定するのは表象をどのように生成するかではなく、表象がどのように消費されるかであるとしている。鈴木貴之 (2015) によるミニマルな表象主義²²⁾では、このようにして消費される内的表象の志向的内容が意識経験の内容となる (p.158)²²⁾。そしてこの内的表象は「世界をありのままに写し取る」再現表象ではなく「みずからの生存にとって有用な仕方世界を分節化する」分節表象である (p.142)²²⁾。

FEP ではイマココの志向の対象を推測 $q(x)$ という形で表現している。この意味では推測 $q(x)$ とは世界についての表象であるということが出来る。しかし推測 $q(x)$ は外界をすべて写し取っているわけではない。つまり、表象主義のうち、外的な刺激をそっくり写しとるような「再現表象」を要請しないミニマルな表象主義は FEP と整合的である。

なお、Friston¹¹⁾や Hohwy²³⁾は FEP における視覚の枠組みとしてヘルムホルツの視覚観を採用している。ヘルムホルツの視覚観は表象説の一形態として捉えることができる。ヘルムホルツの視覚観と FEP の関係については付録 A.2 において議論する。

3.4 FEP は意識の構造からみても妥当である

以上をまとめると、推測 $q(x)$ は表象説におけるイマココの世界の表象の内容と捉えることができる。一方で生成モデル $p(x, s)$ は学習で獲得した「世界のモデル」であり、環境と agent の相互作用のダイナミクスを記述した感覚運動随伴性そのものであり、拘束条件と経験との総体である「来歴」²⁴⁾である。FEP は SMC 説と表象説を無理なく統合した計算理論となっている。

しかしこのままでは FEP は視覚に基づいた行動に関する理論に留まるのであって、FEP を元に意識の理論を構築してゆくためには、意識経験自体がどのような構造を持っているのかについて、第一人称的立場から、意識に直接与えられたものに基づいて、理論構成をすることが必須であるといえるだろう。このような営みをしてきたのがフッサール以降の現象学だ。

Glossary: 現象学関連の各種用語

- 志向性, 志向的对象: 「志向性」とは、「何かについての意識」と定義される。我々は、そのつど与えられている感覚をその向こうへ乗り越える仕方、その感覚が「何について」のものであるかを意識しているが、この「何について」が「志向的对象」と呼ばれる。いま見えている赤が「薔薇の」赤として捉えられているとき、「薔薇」は志向的对象である。そのつど経験されるのは薔薇の部分的現われであり、薔薇そのものがまるごと経験に与えられているわけではない。
- 主題的, 非主題的: 意識が何らかのものに焦点を当て、注意の主要な対象にしていることを「主題的」と呼ぶ。意識のなかでテーマとしてフォーカスを当てられてはいないが、それでも働いている経験のあり方を「非主題的」と呼ぶ。たとえば移動する対象を注視しているとき、見ている対象は主題的だが、それを注視するための眼球運動や身体運動の自己受容感覚は非主題的である。
- 先反省的自己意識, 反省的自己意識: 自己自身をあえて振り返る仕方での明示的 (explicit) に意識することが「反省的自己意識」と呼ばれる。これに対し「先反省的自己意識」とは、このようにはっきりと自己を振り返ることがなくても、自己自身 (や自分の身体) を implicit に意識していることを言う。それは反省に先立つが、反省しようと思えば反省できる (振り返って明示的に意識できる) という性格を持つ。

現象学の考え方によれば、経験は今与えられている感覚を絶えず志向的に乗り越える仕方での客観的現実の意識を生み出している。その意識は、身体的行為との相関で、次に起こる (または起こりうる) 事態を予測するという仕方では生み出されている。この予測は、「当てはずれ」に開かれている。赤いボールを裏返してみたら背面は緑色だった、という場合のように、予測とは異なる事態が起こったとしても、そこで我々の経験は機能不全に陥ったりしない。経験の予測システムの一部に修正が加えられるだけである¹⁾。予測からのズレは、経験の予測システムをアップデートしていくための貴重なデータなのである。このような意識の構造は FEP におけるペイズの枠組みと整合性が高い²⁵⁾。

フッサール現象学における知覚論においては、意識

経験が「主観的なもの」と「非主観的なもの」からなる構造を持っていることが探求されてきた²⁶⁾。視覚意識を例に取れば、1) 注意を向けられている対象、2) 注意を向けては居ないが目に写っていると思われるような周辺視、3) 視覚世界を成り立たせている非主観的な前提条件、という三つの契機が互いに緊密に結びついた構造が作り出されている。

1) は FEP での「イマココの推測 $q(x)$ 」、3) は「生成モデル $p(x, s)$ 」とすると、2) は両者の相互浸透であり、どちらか片方だけで説明できるようなものではない。1) だけが意識であるとするのが（ミニマルな）表象説で、3) が意識であるとするのが SMC 説であるとする、現象学は 1)、2)、3) を一つの構造と捉えている点に優位性があり、それは FEP とも整合的である。

3.5 「FEP と現象学に基づいた意識理論」

以上を元にして、表象説でもなければ SMC 説でもない、「FEP と現象学に基づいた意識理論」を提案したい。意識とは、イマココの推測 $q(x)$ と生成モデル $p(x, s)$ とが一体となって知覚のたびにオンラインで照合され続ける process であり、それによって意識は意識化されない非主観的な前提条件に支えられて図と地の関係のように浮かび上がった志向的对象という構造を持つようになる。意識経験が部分と全体を持ち、通常は単一のものとして経験され、世界と交互作用する主体からの一人称的な視点に拘束されている（先反省的自己意識を持つ²⁷⁾）という意識の特性は FEP の特性から自然と生まれてくる。

「FEP と現象学に基づいた意識理論」では、推測 $q(x)$ と生成モデル $p(x, s)$ との間の差異こそが意識を生き生きと成り立たせているものであると提案する。この差異がなくなったら、主体は環境と単に地続きとなり、意識も消滅してしまう。つまり意識は、異なるものの「間」にこそ生まれてくる。この意味で、意識はそれだけで自立した存在者（状態や対象）ではなく、むしろ本質的に「媒介的」と考えられる²⁸⁾。つまり意識とは、異なるもの（主体と環境、推測 $q(x)$ と生成モデル $p(x, s)$ ）の間を「媒介」するプロセスであると考えられる。

図 7 に以上の考えを図式化したものを示す。この図では、現象学的に明らかにされた意識の構造と、FEP における $q(x)$ と生成モデル $p(x, s)$ の時間的推移とをおおまかに繋げることを意図している。各時点での感覚入力 s が与えられた状態で perceptual inference は推測 $q(x)$ を変化させる。Active inference は視線移動によってサンプルされる感覚入力 s を変化させる。両者が協調して F を下げてゆく。意識の構造における非

主観的な前提条件は生成モデル $p(x, s)$ として複数の階層で表現されている。我々の意識経験はこのような推測 $q(x)$ と生成モデル $p(x, s)$ の両方を含んだものとなっている。

また、たとえば change blindness²⁹⁾ の例にあるように、我々は周辺視野が持つ情報に必ずしもアクセスできているわけではない。しかし周辺視野には穴が空いていない、つまり意識経験が欠けているわけではない。我々の理論ではこのような周辺視野も将来的に視線移動のターゲットとなる場所として意識経験の一部分に組み込まれている。ゆえに我々の理論は「アクセス意識」の理論というよりも「現象的意識」の理論たり得ている。

「FEP と現象学に基づいた意識理論」は意識の単一性や一人称性といった意識の特性と、現象学的に明らかにされた意識の構造と整合的であるという点で、知覚 perception の理論というよりは意識 consciousness の理論たり得ていると言える。しかしこの理論は意識の「内容」についてなにと言えるだろうか？ つまり意識的な情報処理と無意識的な情報処理の区別をつけることができるだろうか？

Hohwy³⁰⁾ は両眼視野闘争の説明として、競合する 2 つの視覚刺激のうちどちらが視覚経験として選択されるかの説明として、perceptual inference に基づいたモデルを用いて、2 つの感覚入力のうち脳内の推測が予測誤差を減らしている方が意識に上る、と説明している。このようにして FEP 自体は意識的な情報処理と無意識的な情報処理の区別をつけることが可能となっている。

しかし「FEP と現象学に基づいた意識理論」はむしろ、視線を向けている場所と周辺視野とそれらの知覚の前提条件とを繋いでひとつかたまりとして扱うという意味において、意識的な情報処理と無意識的な情報処理がどうつながるかについての理論であるといったほうがよい。そのような連続性は人間の意識、動物の意識、機械の意識といったことを考えるときにも同様に役立つだろう。

ここでは知覚を例にとりて説明しているが、ここでの議論は知覚意識を一例とする意識一般に拡張可能であると考えている。意識と無意識を截然と区別するのではなく、主観的な処理と非主観的（無意識的）な処理とが相互浸透するところに意識が成立すると本稿では考えている。意識は孤立的な対象ではなく、むしろ媒介現象なのである。

図7 「FEPと現象学に基づいた意識理論」

3.6 「FEPと現象学に基づいた意識理論」における「介入」の意義

「FEPと現象学に基づいた意識理論」をより意識の構造と特性に整合的にするためには何が必要か。ここで前セクションでの自由エネルギー原理についての三段階の説明に戻ってみよう。自由エネルギー原理はperceptual inference, active inference, 反実仮想の三段階に分けて捉えることができる。よって論理的には「FEPと現象学に基づいた意識理論」における意識の十分条件として、3種類の可能性、つまりperceptual inferenceができること、active inferenceができること、反実仮想ができること、の3通りが考えられる(図8)。ここで我々は、意識の構造について現象学的に妥当であるためには、世界への「介入」という概念が重要であると考えられる。

我々人間を含む生物はactive inferenceによって世界に介入することが可能である。カエルのようなステロタイプな感覚運動変換を行う動物ですら、餌となる虫を見たら舌を伸ばし、捕食者を見たら逃避行動を行うことで世界に介入する。そうして世界の状態または

	意識の充分条件	ヒト	カエル	VAE
Perceptual inference説	生成モデルを持つこと	✓	✓	✓
Active inference説	介入によって作り上げた生成モデルを持つこと	✓	✓	
反実仮想説	未来の行動の結果についての推測を持つこと	✓		

図8 FEPに基づいた意識論としてありうる3種類のリスト

世界と主体の関係を変えることで生存の可能性を高める。そしてこれは結果的に自由エネルギー F を下げる方向に働く。このようにしてagentが自らの介入によって世界を操作することは、世界を自分に関係のあるものとする(groundedなものとする)ことに必須となるだろう。つまり「FEPと現象学に基づいた意識理論」では「active inferenceによって推論と生成モデルの照合が可能な生物はすべて程度の差はある種の意識を持つ」とするactive inference説(図8)の立

場に賛成する^{†10}。

一方で perceptual inference だけが可能なシステムでは、外部の世界を自分に関係あるものとすることができない。ゆえにコンピュータ上の生成モデルに基づくニューラルネットワーク（たとえば variational autoencoder, VAE）はそのままでは意識を持たないだろう（図 8）。そしてそれは「active inference が不可能なシステムでは、世界に介入することで生成モデルを作り上げることができないから」と説明することができる。

図 8 は、「FEP と現象学に基づいた意識理論」が先述した Friston¹⁸⁾や Seth¹⁷⁾による反実仮想説とも異なっていることを示している。しかし active inference においては、行動の結果としてまだ実現していない感覚入力がかどのように変化するかを予測するという反実仮想の成分が含まれている。よって反実仮想説と active inference 説の違いは表面上は無いと言えるかもしれない。しかし、反実仮想説が反実仮想の能力を意識の条件と置くのと比べると、「FEP と現象学に基づいた意識理論」における active inference 説では、agent として世界に介入できることを意識の条件として置き、反実仮想の能力はあくまでその帰結として捉える点に違いがある。また、前述したように、反実仮想説が正しく反実仮想であるためには Pearl の言うような意味での因果推論をする必要がある。このような意味での反実仮想説はより認知的な過程が含まれた、active inference 説とは異なるものとして位置づけられることになるだろう。

3.7 「FEP と現象学に基づいた意識理論」の優位な点

「FEP と現象学に基づいた意識理論」はセクション 1.2 で挙げた SMC 説の弱点を克服している。

まず (1) の「夢」議論についてだが、「SMC 説は意識を環境との相互作用という探索的行動と捉えているが、夢を見ているときは行動していないのに経験があるのではないか」という批判があった。たしかに、我々が夢を見ているときに行動はしていない。しかし生成モデル $p(x, s)$ はすでにそれまでの経験から構築されているので、推測 $q(x)$ と生成モデル $p(x, s)$ の照合は可能である。これが行動していないときでも知覚経験が

^{†10} 「身体を通して世界に ground する」ことの重要性は現象学において深く追求されてきた。たとえばメルロー＝ポンティはこれを以下のように表現している。「我々の身体が空間のなかにあるとか、時間のなかにあるとかと、表現してはならない。我々の身体は、空間や時間に住み込むのである」(p.162)。「意識とは、身体を媒介として事物のもとにある」ということである」(p.233)³¹⁾。

可能である説明となる^{†11}。同じ説明が閉じ込め症候群 (locked-in syndrome)³²⁾や最小意識状態 (minimally conscious state)³³⁾の患者が意識経験を持っていることの説明にも適用できる^{†12}。

つぎに (2) 「必要条件と充分条件」についてだが、感覚運動随伴性は系統発生や個体発生に分類されるものであり、神経科学が明らかにしようとしている至近メカニズムではないのではないかという批判があった。たしかに SMC 説を「生成モデル $p(x, s)$ を技能知として獲得すること」と理解するならば、この批判からは逃れられない。しかし「FEP と現象学に基づいた意識理論」においては、生成モデル $p(x, s)$ は認識の度に推測 $q(x)$ とオンラインで照合される（ティンバーゲンの意味での）至近メカニズムの一部であり、単なる認識の前提条件とはなっていない。

そして (3) 「SMC は定式化された議論になっていない」についてはすでに言及しているが、SMC における感覚運動的な技能知への習熟について、FEP ではそれが生成モデル $p(x, s)$ の確立、維持であるという計算論的な裏付けを提供することができるようになった。

哲学的な観点からの利点としては、「FEP と現象学に基づいた意識理論」では「意識とは process である」と主張していることが挙げられる。これによって、表象説が「意識とは state である」とすることで二元論の問題に直面するのに対して、解決策を提案することができる。これについては付録 A.3 にて議論する。

3.8 「FEP と現象学に基づいた意識理論」に足りないもの 1: Agent にとっての情報と介入

以上のようにして、「FEP と現象学に基づいた意識理論」には意識の理論として優位な点が数多くあることを示すことができた。しかしこの理論にはまだ不十分なところが複数ある。

まず考慮すべきは、もともと FEP は脳を外側から見た第三者的な工学的な理論として構築されたという点だ。FEP ではすでに稼働している脳神経系をどうモデル化するかという観点からスタートしているため、物

^{†11} ただし、知覚入力がないのに推測 $q(x)$ をする特殊なモードの perceptual inference であることが、veridical な意識経験と夢の意識経験の質的違いの起源となる可能性を示唆している。

^{†12} 前述のとおり、開眼手術の直後に患者はすぐに正確な視覚経験を持つことはできない⁴⁾。他の知覚モダリティーの生成モデル $p(x, s)$ があっても、視覚の生成モデルがないので、イマココの推測 $q(x)$ と照合させることができない。短期間の視覚経験（1 日程度）によって生成モデルができると、視覚経験を持つことができるようになる³⁴⁾。

理世界の中にどうやって表象の世界を創発させるかという問題は後回しになっている^{†13}。

FEP では情報理論的な枠組みを全面的に援用しているが、その情報とは、agent にとって意味をなす違いによって分けられた事象についてのもの (“a difference which makes a difference”)³⁷⁾に改める必要があるだろう。それによって、その生物にとっての解像度で分節化された世界での情報量に基づいての推測 $q(x)$ と生成モデル $p(x, s)$ を作ることが可能になるはずだ。

同様な問題として、介入をどう FEP に組み込むかという問題がある。前述したように「FEP と現象学に基づいた意識理論」では、agent が自らの介入によって世界を操作することは、世界を自分に関係のあるものとする (grounded なものとする) ことに必須であり、原始的な知覚的意識を持つことの条件となると主張した。しかしそのような自らの介入によって確立、維持されている生成モデルと、コンピュータ上の生成モデルとは FEP の作動の上で区別が可能となっていない。

3.9 「FEP と現象学に基づいた意識理論」に足りないもの 2: 自己の境界問題

FEP での推測 $q(x)$ と生成モデル $p(x, s)$ を計算するためには agent と環境の境界が前提として必要だ。Friston は、情報のやり取りが比較的隔絶されたマルコフ・ブランケットによって agent と環境の区別が規定されるとしている³⁵⁾。一方で Friston は単細胞における細胞外との相互作用³⁵⁾や進化における自然選択¹⁸⁾さえある種の推論であるとしている。これは FEP における主体の境界が相対的または恣意的であることを示している。

これに対して「FEP と現象学に基づいた意識理論」では agent が世界への「介入」を行うことを意識の条件とした。これによって agent と環境の境界も介入可能な個体として規定される。しかしこの考えは agent が行動主体として働く主体性 agency を仮定しており、論点先取となっていないだろうか？ これに対する我々の答えはこうだ：進化、発達の段階でどう自己ができたかを想像してみれば、世界への介入と主体性は同じ根っこから同時に生まれたと考えたほうが妥当なはずだ。世界への介入に基づく推論、つまり active inference が可能になったときに、その agent は原始的な意識を持つようになり、それと同時に agent からみた世界は agent と環境に切り分けられて主体性の単位が決まる。

^{†13} この点については人工知能、人工生命的なアプローチでミニマルな FEP エージェントを創発させるというのが一つの方策だろう^{7, 35, 36)}。

つまり意識の理論は同時に主体性の理論となる^{†14}。

4. さ い ご に

この記事の前半では、眼球運動を行いながら視覚シーンを構成してゆくという例をとって、自由エネルギー原理を非常に簡単な言葉で説明することを目指した。この総説を読んでさらに FEP について知りたいという方にはまずは本特集号の各総説にあたることをおすすめしたい。また、Buckley ら (2017)³⁸⁾がごまかしなく丁寧に FEP の説明を行っているのでこれに当たることをおすすめしたい。

後半ではこのようにして理解した FEP をもとに「FEP と現象学に基づいた意識理論」を提案した。しかしこれはまだ意識の「計算理論」とはなっていない。つまり、統合情報理論において意識の統合の度合いを Φ で表現する³⁹⁾のと同じような意味では F が意識の強度を表すことをこの理論が主張しているわけではない。代わりに今回の議論では、意識の特性と意識の構造と整合的であるような意識の計算理論の候補として FEP の理論構造が有望であることを提案した。そして agent にとっての推測、agent にとっての介入などの拘束条件を抽出してこれから意識の「計算理論」を構築するための基礎固めを行った。今後の進歩を期待していただけたらと思う。

謝辞

本記事を執筆するにあたっては、論文討論会等においてコメントを下された特集号メンバーの皆様に感謝します。

本研究の一部は国立研究開発法人科学技術振興機構 (JST) の研究成果展開事業「センター・オブ・イノベーション (COI) プログラム」および国立研究開発法人日本医療研究開発機構 (AMED) の「革新的技術による脳機能ネットワークの全容解明プロジェクト」(JP18dm0207011) および JSPS 科研費 JP17K02153 の支援によって行われた。

参 考 文 献

- 1) 田口 茂 (2014): 現象学という思考: <自明なもの>の知へ, 筑摩書房, 東京。
- 2) Mach, E. (1886): Die Analyse der Empfind-

^{†14} この考え方にはたとえば以下の利点がある。Agent と環境の境界は介入が単位となって規定されるため、マルコフ・ブランケットによる境界で起こるようなサブパーソナルな境界が多数作られることを回避できる。また、介入の範囲が道具使用によって拡張しうることもうまく説明できる。

- ungen und das Verhältnis des Physischen zum Psychischen (『感覚の分析』(叢書・ユニベルシタス; 26), 須藤吾之助, 廣松渉訳, 法政大学出版, 1971. 図の URL: <https://publicdomainreview.org/collections/self-portrait-by-ernst-mach-1886/>)
- 3) Findlay, J. M., Gilchrist, I. D. (2003): Active vision: The psychology of looking and seeing, Oxford University Press (フィンドレイ, J. M.・ギルクリスト, I. D. 本田仁視(監訳) (2006), アクティブ・ビジョン-眼球運動の心理・神経科学- 北大路書房).
 - 4) Noë, A. (2004): Action in Perception. Cambridge, MA: MIT Press (アルヴァ・ノエ, 『知覚のなかの行為』, 門脇俊介, 石原孝二監訳, 春秋社, 2010).
 - 5) O'Regan, J. K. (2001): "Experience is not something we feel but something we do: a principled way of explaining sensory phenomenology, with Change Blindness and other empirical consequences," <http://nivea.psych.univ-paris5.fr/ASSChtml/Pacherie4.html>
 - 6) Friston, K., Adams, R. A., Perrinet, L., Breakspear, M. (2012): Perceptions as hypotheses: saccades as experiments, *Front Psychol.*, Vol.3, p.151.
 - 7) McGregor, S., Baltieri, M., Buckley, C. L. (2015): A Minimal Active Inference Agent, *arXiv:1503.04187* <https://arxiv.org/abs/1503.04187>
 - 8) 乾 敏郎 (2018): 知覚・認知・運動・感情・意思決定をつなぐ自由エネルギー原理, *日本神経回路学会誌*, Vol.25, No.3, pp.123-134.
 - 9) 磯村拓哉 (2018): 自由エネルギー原理の解説: 知覚・行動・他者の思考の推論, *日本神経回路学会誌*, Vol.25, No.3, pp.71-85.
 - 10) 大羽成征 (2018): 自由エネルギー原理は脳型人工知能の基盤となるか〜アクティブインファレンスの意義について〜, *日本神経回路学会誌*, Vol.25, No.3, pp.113-122.
 - 11) Friston, K. (2010): The free-energy principle: a unified brain theory?, *Nature Reviews Neuroscience*, Vol.11, No.2, pp.127-138.
 - 12) Rao, R. P. (1999): An optimal estimation approach to visual perception and learning, *Vision Research*, Vol.39, No.11, pp.1963-1989.
 - 13) Mumford, D. (1992): On the computational architecture of the neocortex. II. The role of cortico-cortical loops, *Biol. Cybern.*, Vol.66, No.3, pp.241-251.
 - 14) 川人光男, 乾 敏郎 (1990): 視覚大脳皮質の計算理論, *電子情報通信学会論文誌 D-II*, Vol.J73-D-II, No.8, pp.1111-1121.
 - 15) Kawato, M., Hayakawa, H., Inui, T. (1993): A forward-inverse optics model of reciprocal connections between visual cortical areas, *Network, Computation in Neural Systems*, Vol.4, No.4, pp.415-422.
 - 16) Sommer, M. A., Wurtz, R. H. (2008): Brain circuits for the internal monitoring of movements, *Annu. Rev. Neurosci.*, Vol.31, pp.317-338.
 - 17) Seth, A. K. (2015): The cybernetic Bayesian brain: from interoceptive inference to sensorimotor contingencies, In *Open MIND*, Metzinger, T., Windt, J., eds., Frankfurt a.M., GER: MIND group.
 - 18) Friston, K. (2018): Am I Self-Conscious? (Or Does Self-Organization Entail Self-Consciousness?), *Frontiers in Psychology*, Vol.9, p.47, <http://doi.org/10.3389/fpsyg.2018.00579>.
 - 19) Judea P. (2000): Causality: Models, Reasoning, and Inference, Cambridge University Press (黒木学 訳, 「統計的因果推論 -モデル・推論・推測」, 共立出版, 2009 年).
 - 20) O'Regan, J. K., Noë, A. (2001): A sensorimotor account of vision and visual consciousness, *Behav. Brain Sci.*, Vol.24, No.5, pp.939-973; discussion, pp.973-1031.
 - 21) Millikan, R. G. (1993): Biosemantics. In *White Queen Psychology and Other Essays for Alice*, Cambridge, Mass., MIT Press, pp.83-101 (ルース・ギャレット・ミリカン/前田高弘訳, 「バイオセマンティックス」 in 信原幸弘編「シリーズ心の哲学 III 翻訳編」, 勁草書房, 2004 年).
 - 22) 鈴木貴之 (2015): ぼくらが原子の集まりなら, なぜ痛みや悲しを感じるのだろうか, 勁草書房, 東京.
 - 23) Hohwy, J. (2013): *The Predictive Mind*, Oxford University Press.
 - 24) 下條信輔 (1999): 「意識」とは何だろうか-脳の来歴, 知覚の錯誤, 講談社現代新書, 講談社, 東京.
 - 25) Madary, M. (2016): *Visual Phenomenology*, Cambridge, MA, MIT Press.
 - 26) Husserl, E. (1972): *Erfahrung und Urteil*, Hamburg: Felix Meiner.
 - 27) Zahavi, D. (2005): *Subjectivity and Selfhood*, Cambridge MA, Bradford Book/MIT Press.
 - 28) 田口 茂 (2015): 媒介論的現象学の構想, *ハイデガー・フォーラム*, Vol.9, pp.37-48.
 - 29) Rensink, R. A. (2000): Seeing, sensing, and scrutinizing, *Vision Res.*, Vol.40, no.10-12, pp.1469-1487.
 - 30) Hohwy, J., Roepstorff, A., Friston, K. (2008): Predictive coding explains binocular rivalry: an epistemological review, *Cognition*, Vol.108, No.3, pp.687-701.

- 31) Merleau-Ponty, M. (1945): *Phénoménologie de la perception*, Paris: Gallimard (竹内芳郎, 小木貞孝訳, 「知覚の現象学 1」, みすず書房, 1967 年).
- 32) Chatelle, C., Chennu, S., Noirhomme, Q., Cruse, D., Owen, A. M., Laureys, S. (2012): Brain-computer interfacing in disorders of consciousness, *Brain Inj.*, Vol.26, No.12, pp.1510–1522; doi: 10.3109/02699052.2012.698362.
- 33) Boly, M., Garrido, M. I., Gosseries, O., Bruno, M. A., Boveroux, P., Schnakers, C., et al. (2011): Preserved feedforward but impaired top-down processes in the vegetative state, *Science*, Vol.332, pp.858–862; doi: 10.1126/science.1202043.
- 34) Held, R., Ostrovsky, Y., de Gelder, B., Gandhi, T., Ganesh, S., Mathur, U., Sinha, P. (2011): The newly sighted fail to match seen with felt, *Nat. Neurosci.*, Vol.14, No.5, pp.551–553; doi: 10.1038/nn.2795.
- 35) Friston, K. (2013): Life as we know it, *J. R. Soc. Interface*, Vol.10, No.86, pp.20130475; doi: 10.1098/rsif.2013.0475.
- 36) Baltieri, M., Buckley, C. L. (2018): A probabilistic interpretation of PID controllers using active inference, *bioRxiv*, 284562. <http://doi.org/10.1101/284562>
- 37) Bateson, G. (1987): Form, substance, and difference. In: Bateson, G. (ed.), *Steps to an ecology of mind: Collected essays in anthropology, psychiatry, evolution, and epistemology*, Jason Aronson, Northvale NJ, pp.455–471 (佐藤良明訳 (2000), 精神の生態学 (改訂第2版), 新思案社).
- 38) Buckley, C. L., Kim, C. S., McGregor, S., Seth, A. K. (2017): The free energy principle for action and perception: A mathematical review, *Journal of Mathematical Psychology*, Vol.81, pp.55–79
- 39) Oizumi, M., Albantakis, L., Tononi, G. (2014): From the phenomenology to the mechanisms of consciousness: Integrated Information Theory 3.0, *PLoS Comput. Biol.*, Vol.10, No.5, e1003588; doi: 10.1371/journal.pcbi.1003588.
- 40) Yoshida, M., Itti, L., Berg, D.J., Ikeda, T., Kato, R., et al. (2012): Residual attention guidance in blindsight monkeys watching complex natural scenes, *Curr. Biol.*, Vol.22, pp.1429–1434.
- 41) 吉田正俊 (2014): サリエンス・マップの視覚探索解析への応用, 日本神経回路学会誌, Vol.21, No.1, pp.3–12.
- 42) Itti, L., Koch, C. (2001): Computational modelling of visual attention, *Nat. Rev. Neurosci.*, Vol.2, No.3, pp.194–203.
- 43) Veale, R., Hafed, Z.M., Yoshida, M. (2017): How is visual salience computed in the brain? Insights from behaviour, neurobiology and modelling, *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, Vol.372, No.1714. pii: 20160113. doi: 10.1098/rstb.2016.0113.
- 44) Itti, L., Baldi, P. (2012): Bayesian surprise attracts human attention, *Vision Res.*, Vol.49, No.10, pp.1295–1306.
- 45) Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., Pezzulo, G. (2015): Active inference and epistemic value, *Cogn. Neurosci.*, Vol.6, No.4, pp.187–214.
- 46) Hohwy, J. (2016): The Self - Evidencing Brain, *Nous*, Vol.50, pp.259–285; doi:10.1111/nous.12062.
- 47) Bruineberg, J., Kiverstein, J., Rietveld, E. (2018): The anticipating brain is not a scientist: the free-energy principle from an ecological-enactive perspective, *Synthese*, Vol.195, pp.2417–2444.
- 48) Allen, M., Friston, K.J. (2018): From cognitivism to autopoiesis: towards a computational framework for the embodied mind, *Synthese*, Vol.195, No.6, pp.2459–2482.
- 49) Kirchhoff, M.D., Froese, T. (2017): Where there is life there is mind: In support of a strong life-mind continuity thesis, *Entropy*, Vol.19, No.4, p.169.
- 50) Kiverstein, J. (2018): Free Energy and the Self: An Ecological-Enactive Interpretation *Topoi*, pp.1–16. <https://doi.org/10.1007/s11245-018-9561-5>.
- 51) Baltieri, M., Buckley, C. L. (2018): The modularity of action and perception revisited using control theory and active inference, *arXiv:1806.02649*.
- 52) Gallagher, S., Allen, M. (2018): Active inference, enactivism and the hermeneutics of social cognition, *Synthese*, Vol.195, No.6, pp.2627–2648.
- 53) Linson, A., Clark, A., Ramamoorthy, S., Friston, K. (2018): The active inference approach to ecological perception: General information dynamics for natural and artificial embodied cognition, *Frontiers in Robotics and AI*, 5, 21. DOI: 10.3389/frobt.2018.00021.

付 録

A.1 FEP における視覚サリエンス

セクション 2 では画像を見ているときの視線移動についての単純化した状況を題材にして FEP の説明を

行った．このような状況は視覚サリエンスを元にした視線移動としてモデル化されてきた^{40,41)}．つまり，画像自体の情報によって眼を惹く場所を視覚サリエンスが高いと呼び，画像の位置ごとの視覚サリエンスの大きさを表現したものをサリエンシー・マップと呼ぶ^{42,43)}．我々の視線は視覚サリエンスの高いところに誘引されながら情報を収集するプロセスとしてモデル化される．

サリエンシー・マップは画像内の視覚情報の空間的配置によって決まる静的な特性である．一方で Itti and Baldi による Bayesian Surprise⁴⁴⁾では動画において視覚情報の空間的，時間的特性から眼を惹く場所を計算している．これが FEP の式 (3) で出てくる Bayesian Surprise そのものである．ここでは FEP におけるサリエンシー・マップがどのような位置づけにあるか，そしてサリエンシー・マップと Bayesian Surprise との関係について説明する．

まず，行動 a は行動計画の確率分布 control state $u(x)$ の中から F を最小化するものとして選ばれる．Agent は control state $u(x)$ にはアクセスできるが行動 a 自体にはアクセスできない． $u(x)$ は prior η_u と変動項 ω_u を持っている．ここで a , u , η_u の次元は眼球運動の方向ベクトルになっている．

$$u = \eta_u + \omega_u$$

FEP の枠組みにおいて「視覚サリエンスが高い」とは「そこを見るような行動選択をすれば将来の $q(x')$ のエントロピー $H(q(x')) = -\sum_{x'} q(x') \log(q(x'))$ を下げることができる場所」のことと捉える⁶⁾．よって FEP における視覚サリエンス S は可能な行動 (=現在位置からの眼球運動の方向ベクトル) ごとに計算可能な値であり，最も視覚サリエンスの高い場所が次の視線移動の prior η_u となる．

$$S(\eta_u) = -H(q(x')|\eta_u) \\ \eta_u = \operatorname{argmax}_{\eta_u} S(\eta_u)$$

FEP の式表現の (3) 式は $F(s) = \text{Uncertainty}(s) + \text{Bayesian surprise}$ となっていた．ここで Bayesian surprise と salience には以下の関係がある．

$$\begin{aligned} \text{Bayesian Surprise} \\ &= D_{KL}[q(x)||p(x)] \\ &= -\sum_x \{q(x) \log(p(x))\} + \underbrace{-H[q(x)]}_{\text{Salience}} \end{aligned}$$

もし $p(x)$ が informative ではない場合，つまり分布が x に依存せず平らである場合，第 1 項は $q(x)$ に依

存しない定数となる．このとき Bayesian Surprise と Salience は等価になる．

Salience(=サリエンシー・マップ)⁴²⁾は画像情報の空間的特性によってきまる静的な特性である．一方で Bayesian Surprise⁴⁴⁾では動画において視覚情報の空間的，時間的特性から眼を惹く場所を計算している．よって， $p(x)$ が informative ではないとはつまり以前の情報が意味が無いときのことから，時間を考えずに空間だけ考えたときに Salience が Bayesian Surprise と等価になるということは理にかなっている．

本記事は Friston et al. (2012)⁶⁾に準拠した説明を行っているが，より新しい論文⁴⁵⁾においては，行動選択についての理論的整備が進められている．そこでは行動は行動の policy π に関する expected free energy $G(\pi)$ を最小化するように選択されるとモデル化したうえで，視覚探索における 2 つの value として extrinsic value と epistemic value に分けている．視覚サリエンスおよび Bayesian surprise は後者の epistemic value に相当する．詳細については本特集号の他の記事を参照していただきたい．

A.2 ヘルムホルツ的視覚観と SMC 的または現象学的視覚観について

本記事では FEP を SMC 的または現象学的視覚観とつなげることを提案した．一方で Hohwy²³⁾や Friston の初期の FEP 論文の説明においては，FEP における視覚の枠組みがヘルムホルツ的視覚観，つまり外界に直接アクセスできない脳が外界のモデルを作るのが視覚であるという考えに基づいていると主張してきた．

たとえば Hohwy⁴⁶⁾によれば，perceptual inference における予測誤差最小化 (PEM) の過程はヘルムホルツ的視覚観を反映しており，視覚には外的世界との交互作用を考える必要はないと主張する．

しかし「FEP と現象学に基づいた意識理論」においてはヘルムホルツ的視覚観よりも，SMC 的，現象学的な視覚観が FEP の特性をよりうまく捉えていると著者は考えている．

この理由としてまず，ヘルムホルツ的視覚観は perceptual inference のみに閉じた系の説明にはよく合致しているが，active inference を含めた FEP 全体の構造の説明にはなっていない点が挙げられる．Active inference においては行動 a の結果外界の原因 $\mathbf{X}$ が変化して感覚入力 s をどう変化するか予測する際には $s = \mathbf{g}(\mathbf{X})$ つまり外界の生成過程を脳内の生成モデル $s = g(x)$ と比較する必要がある．これが行動の不確定性の原因であり，active inference は脳活動だけでは説明することができない．

そして Bruineberg ら⁴⁷⁾が議論しているように, FEP の本体は active inference を含んだ能動的なプロセスにある。なぜならば, FEP の式 (2) からわかるように, perceptual inference では推測 $q(x)$ を変化させることで *Relative entropy* を最小化することはできるが, *Sensory surprise* $= -\log(p(s))$ を変えることはできないからだ。つまり, FEP が目的としている *Sensory surprise* の最小化は active inference による感覚入力 s の変化でのみ可能となっている。よって active inference は, FEP の式 (3) における *Uncertainty* を下げることによって, perceptual inference を促進するという補助的役割を果たしているというよりは, active inference こそが FEP における *Sensory surprise* の最小化という主役の機能を果たしていると言う方が正しい。

このように, FEP とヘルムホルツ的視覚観は必ずしも分かちがたく結びついているものではない。また, FEP を SMC 的, 現象学的考えとつなげて考える論文はここ数年で急速に増えていて, そのなかには Friston 本人が著者に入っているものも含まれている⁴⁷⁻⁵³⁾。

A.3 意識とは state であるか, process であるか

「FEP と現象学に基づいた意識理論」の優位な点として, 「意識とは process である」と主張していることが挙げられる。これによって, 表象説が「意識とは state である」とすることで二元論の問題に直面するのに対して, 解決策を提案することができる。これについて説明しておきたい。

表象説における「意識経験を持つとは, 脳内に内的表象を持つことである」という考えは, 意識とは脳の state であることを前提としている。この点は統合情報理論³⁹⁾においても変わらない。「FEP と現象学に基づいた意識理論」がこれらと大きく異なっているのは, 意識とは process であると主張する点にある。意識とは推測 $q(x)$ による志向的対象ではなくて, 推測 $q(x)$ と生成モデル $p(x, s)$ とを照合する process である。Friston も最新の論文¹⁸⁾において, 意識とは state ではなくて process であると提案しているが, これは我々の説とよく合致している。

一方で SMC 説に関してはこの点で紆余曲折があっ

図 A.9 State としての意識 (a) と process としての意識 (b)

た。SMC 説を最初に提案した O'Regan and Noë (2001)²⁰⁾では意識が感覚運動随伴性 (sensorimotor contingency) という探索的行為 (process) である点が強調されていた。しかしその後の Noë (2004)⁴⁾では, 感覚運動随伴性に関する技能的知識 (感覚運動依存性: sensorimotor dependency) を獲得することを強調するように力点が変わっている。これによって Noë (2004)⁴⁾は「夢」議論で攻撃されたような, 意識には常に行動が必要であるという考え方を回避することには成功したが, 意識が探索的行為 (process) であるとする主張が弱められており, この点で理論的な後退が見られる。しかし「FEP と現象学に基づいた意識理論」では推測 $q(x)$ と生成モデル $p(x, s)$ とを照合する過程であると主張するため, 意識とは process であるという主張を保持できている。

Noë (2004)⁴⁾による, 意識の神経相関 (Neural correlates of consciousness, NCC) による研究アプローチに対する批判のひとつに, NCC アプローチが二元論的な枠組みから逃れることができない点がある。NCC アプローチの枠組みは, 脳の state と意識の state との対応関係を探るというものだ (図 A.9a)。このため, 脳の state と意識の state とは存在論的にどのような関係にあるのか, どのように両者を架橋するのか (explanatory gap) という問題を引き起こす。しかし意識について脳の state 間を媒介する process と捉える場合 (図 A.9b) にはこのような対応付け問題に囚われなくて済むようになる。