エモーション・スタディーズ 第3巻第1号 pp. 2–12（2017）
doi: 10.20797/ems.3.1_2

特集論文

予測的符号化・内受容感覚・感情[1]

大平英樹（名古屋大学）

Predictive coding, interoception, and affect

Hideki Ohira (*Nagoya University*)

（2017年5月17日受稿，2017年6月12日受理）

Lisa Feldman-Barrett, who has promoted a psychological constructivism theory of affect, recently proposed the Embodied Predictive Interoception Coding (EPIC) model of affect, on the basis of the perspective of predictive coding. The theoretical framework of predictive coding argues that the brain creates inner models which can provide predictions for perception and motor movement, and that perception and behaviors are emerged from Bayesian computations rooted on the predictions. The EPIC model expands this perspective into interoception, which is perception of inner body states, and tries to explain phenomena of affect as integrative experiences based on interoception. This article introduces concepts of the EPIC model and examines the model by referencing to empirical findings.
Key words: interoception, predictive coding, affect

1. 序

Lisa Feldman-Barrettは，心理学的構成主義（psychological constructivism：本特集の対談記事を参照）に基づく感情理論で知られる論客であるが，彼女は最近，身体化された予測的内受容符号化モデル（Embodied Predictive Interoception Coding model：以下，EPICモデルと呼ぶ）[2]という理論的枠組みを提唱している（Feldman-Barrett, 2017; Feldman-Barrett, Quigley, & Hamilton, 2016; Feldman-Barrett & Simmons, 2015）。これは，近年の神経科学において優勢となりつつある予測的符号化（predictive coding）の概念，つまり脳はさまざまな階層において内的モデル（inner model）を構築し，それにより形成される予測と入力される刺激の相互作用からあらゆる機能を創発しているという考え方に基づいて，感情のみならず，知覚，運動，認知，意思決定などのすべての精神機能を統一的に説明しようとする企てである。EPICモデルはいまだ完成された理論体系には至っていないし，それ自体の妥当性を検証することは難しい一種のメタ理論である。また，EPICモデルから導出された仮説を検証した実証的研究も現在のところほとんどない。しかし，EPICモデルは極めて射程の長い一般的理論を志向している点で魅力的であり，最近の神経科学の知見との相性もよい。EPICモデルに対してどのようなスタンスを取るにせよ，このモデルを理解し考察することは，今後の感情研究において重要であると思われる。そこで本稿では，EPICモデルを紹介し，これまでの筆者らの研究知見をこのモデルに照らして検討することにより，その妥当性と示唆を考察する。

Correspondence concerning this article should be sent to: Hideki Ohira, Nagoya University, Department of Psychology, Furo-cho, Chikusa-ku, Nagoya 464–8601, Japan (e-mail: ohira@lit.nagoya-u.ac.jp)

[1] 本稿は，筆者による先の論考（大平，2017）に基づき，これをさらに発展させたものである。

[2] この名称自体はおそらく確定したものではなく，今後変更される可能性も考えられる。

2. 予測的符号化

脳は，感覚器官から入力される刺激に受動的に反応しているのではなく，これから入力される刺激を予測する内的モデルを構成し，それによる予測と入力された感覚信号を比較し，両者のずれ（予測誤差：prediction error）の計算に基づいて，知覚を能動的に創発していると考えられる。こうした脳の働きを予測的符号化と呼ぶ（Friston, 2010; Friston, Kilner, & Harrison, 2006）。この発想の起源は，19世紀の物

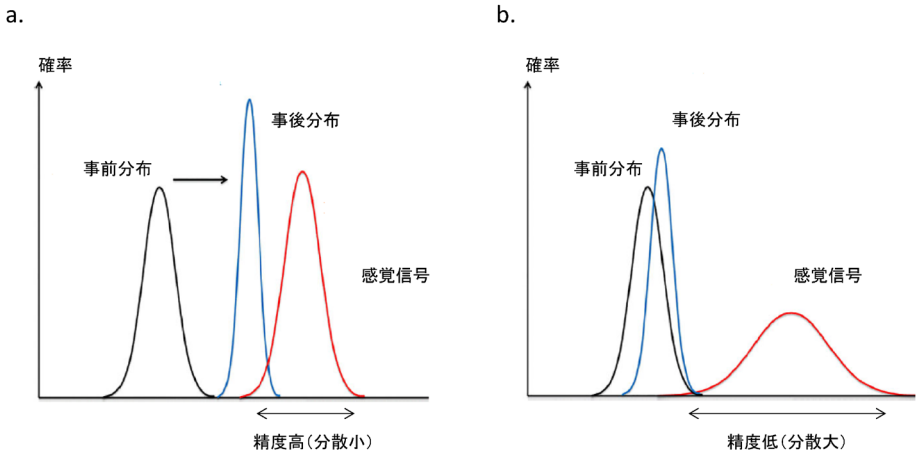

Figure 1.　予測的符号化の原理（Ainley et al., 2016に基づき筆者が作成）
事前分布は，内的モデルにより生成される，ある知覚の予測を表す。ここに感覚信号が入力されると，予測との差分である予測誤差が計算される。これらから，ベイズの定理に基づいて事前分布が事後分布に更新される。主観的な知覚経験は，こうした一連の過程，特に事前分布から事後分布への更新が意識されたものであると考えられる。
a. とb. では予測は同じであり，予測と感覚信号の平均値の距離（平均予測誤差）も等しい。しかしb. はa. に比べて感覚信号の精度が低い（分散が大きい）。感覚信号の精度が高ければ，予測が大きく更新されるが（a.），感覚信号の精度が低い場合には，主観的に経験される知覚は，入力された感覚信号とはかけ離れてほとんど予測と同じになる（b.）。

理学者ヘルムホルツが提唱した視覚の「無意識の推論（unconscious inference）」に遡る（Helmholtz, 1962 (1866)）。彼は，狭い中心視の範囲外では視覚像はぼやけており，網膜には盲点や表面を走行する血管があるなど，我々の視覚装置には大きな制約があるのに安定した視覚経験が得られるのは，脳が過去の経験に基づいて限られた視覚信号から世界像を推論しているためだと考えた。こうした処理は，気付かないうちに瞬時に行われるので，ヘルムホルツはこれを「無意識の推論」と呼んだ。

予測的符号化の理論では，脳における推論をベイズ統計学の原理になぞらえて説明する（Figure 1: Ainley, Apps, Fotopoulou, & Tsakiris, 2016）。知覚に関する内的モデルによる予測は，確率分布として表現される。これはベイズ統計学でいう事前分布（prior distribution）にあたる。感覚信号も確率分布として入力され，予測としての事前分布とのずれとして予測誤差が計算される[3]。この感覚信号がベイズ統計学でいう観測あるいは尤度（likelihood）にあたり，ベイズの定理による更新のように，事後分布（posterior distribution）が計算される。我々が主観的に経験する知覚は，この事前分布から感覚信号の入力を受けて事後分布が計算される一連の過程が意識されたものだと考えられる。ここでの予測とは，脳内の神経ネットワークが自発的に示す活動パターンを意味する。これは純粋に物理的・生物学的な現象であり，意志により生じる特定の意味内容を予見する精神活動のことではないので注意が必要である[4]。

この際，知覚経験を規定する重要な要因のひとつが，予測や感覚信号の精度（precision）である。精度は確率分布の分散を意味し，Figure 1では分布の幅として表現されている。何度も経験した事象については内的モデルの精度が高く，未知の事象では内的モデルの精度は低い。また，対象に注意を向けることで，感覚信号の精度を上げることができる。予測と感覚信号の平均間の距離が等しくても，予測と感覚信号の精度が同程度であれば主観的な知覚経験は両者の中間的なものになるのに対して（Figure 1a），予測の精度が高く（分散が小さく）感覚信号の精度が低ければ（分散が大きければ），経験される知覚は予測に大きく依存し，実際の感覚信号とはかけ離れたものになる（Figure 1b）。また，事前分布と事後分布の精度が高いほど，主観的には明瞭な知覚が経験される。

こうしたベイズ的な計算が，すべてのモダリティにおける知覚や運動の処理において行われており，また

[3]　この確率分布のずれ（距離）をカルバック・ライブラー情報量（Kullback-Leibler（KL）divergence）と呼び，次のように表される。

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

ここでP，Qをそれぞれ連続確率分布とする。p，qはそれぞれP，Qの確率密度関数を表す。

[4]　ただし，予測的符号化による処理は階層的なので，その最も上位階層においては，認知心理学でいうスキーマ（schema）の如く意味内容を伴った予測もありうるであろう。

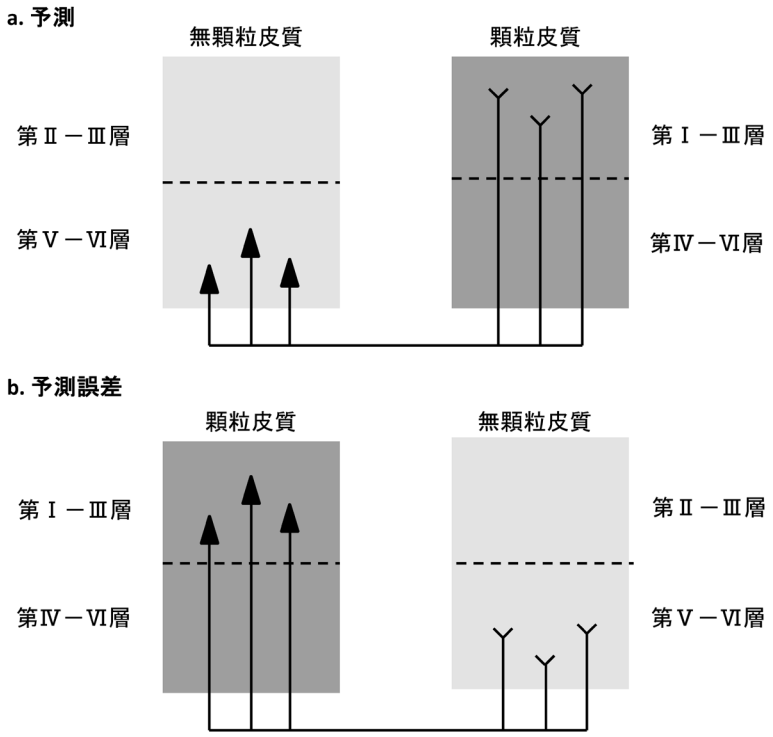

Figure 2.　予測と予測誤差の神経解剖学的基盤（Feldman-Barrett & Simmons, 2015に基づき筆者が作成）
予測は無顆粒皮質の深層に存在する錐体細胞群（図では三角形で表現されている）の活動により形成され，顆粒皮質の浅層に送られる。顆粒細胞には感覚信号が入力され，予測との比較が行われ，予測誤差が計算される。予測誤差は，無顆粒皮質に送り返され，その活動を調整することにより，内的モデルが更新される。

感覚器官の低次なレベル（例えば網膜における信号処理）から脳の連合野における高次なレベル（例えば前頭前皮質における特定の物体を認識しようとする目標）まで階層的に行われていると考えられている。ヒトを含む生体は，そうした階層的に検出される予測誤差の和を最小化することで統一的で整合的な自己像と世界像を構築し，それらを維持しようと努める。この予測誤差の和は，これもヘルムホルツの熱力学理論になぞらえて，「自由エネルギー（free energy）」と呼ばれている（Friston, 2010; Friston, Kilner, & Harrison, 2006）。予測誤差を最小化するためには，内的モデルの更新[5]と，外界への働きかけによる感覚信号の積極的な調整が行われる。前者はヘルムホルツが言った「無意識の推論」と似た過程である。一方，後者の過程は「能動的推論（active inference）」と呼ばれる。例えば，視覚を調整するために体を移動させて対象との距離や見る角度を変えたり，対象に注意して凝視したりする行為がそれにあたる。生体は，この両方の手段をダイナミックに用いることにより，全体として予測誤差を縮小するように振る舞うと想定され，これが脳の唯一の原理であるとして「自由エネルギー原理」と呼ばれる[6]。この考え方は今のところ仮説的な理論であるが，現代の認知神経科学において影響力を増しつつある。

3.　予測的符号化の解剖学的基盤

BarbasとRempel-Clower（1997）は，大脳皮質の組織学的特性によって，予測を担う脳部位と，感覚信号を受け予測誤差の計算を担う脳部位が区分されると主張している。視覚，聴覚，体性感覚など，感覚入力を受ける脳部位は顆粒皮質（granular cortex）と呼ばれる。顆粒皮質には明瞭な6層の構造があり，特

[5]　一回の事象における知覚経験による事後分布が，ただちに次の時点の事前分布になるわけではない。ベイズ統計学でいう状態方程式に従い，事後分布を一定の規則で変換することにより，次の事前分布が形成される。

[6]　認知神経科学における自由エネルギーの理論への批判として，もし脳が自由エネルギーを最小化することを目的としているならば，生体は何も刺激を受けない真っ暗な部屋に動かずにいるような状態を志向するのではないか，という考え方があり，一般に「暗室問題」と呼ばれる。しかし現実にはヒトを含めた生体が生きている環境は常に変化しているので，常に感覚の予測誤差は生じており，それを最小化するように振る舞うことが，結果として環境への適応に有益であると考えられている。また，うつ病，統合失調症，解離性人格障害などの精神疾患，あるいは高血圧，アレルギー，免疫異常などの身体疾患は，こうした予測に基づく処理の過程の不全から生じるという主張もある。

に第4層において，丸い形をした顆粒細胞（granule cell）を豊富に含む。ピラミッド形の錐体細胞（pyramidal cell）が長い軸索を持ち，異なる脳部位への出力を担うのに対して，顆粒細胞の神経連絡は局所的であり，脳部位局所で入力された信号を増幅する機能を担っている。このため顆粒皮質の構造は感覚情報の処理に適している。これに対して顆粒細胞を含まず，層構造が未分化な無顆粒皮質（agranular cortex）は，運動野や帯状皮質など出力に関連する脳部位である。脳のさまざまなレベルで，顆粒皮質の浅層から無顆粒皮質の深層へ，無顆粒皮質の深層から顆粒皮質の浅層へという双方向の神経連絡があり，感覚入力の処理とその調整という機能単位を構成している（Figure 2）。

こうした解剖学的な知見からBarbasら（1997）は，無顆粒皮質は過去の経験に基づく感覚の予測を担うと考えている。この予測は，無顆粒皮質の深層に存在する錐体細胞群の発火パターンにより符号化されている。この情報は顆粒皮質の浅層に送られ，ここに入力される感覚信号と照合されて予測誤差が計算される。この予測誤差の情報は無顆粒皮質の深層に送り返され，モデルの更新に利用される。生体が生きている限り，この過程が休まず繰り返され，それにより主観的経験が連続的に形成される。この仮説を支持する傍証として，光遺伝学（optogenetics）[7]の手法を用いてマウスの運動野から第一次体性感覚野への信号入力を阻止すると，マウスは滑らかな床とざらざらした床を区別することができなくなったという知見がある（Manita, Suzuki, Homma, Matsumoto, Odagawa, Yamada, Ota, Matsubara, Inutsuka, Sato, Ohkura, Yamanaka, Yanagawa, Nakai, Hayashi, Larkum, & Murayama, 2015）。この知見は，身体から体性感覚野への経路が正常であっても，運動野からの予測信号が無ければ，正常な触覚経験は得られないことを示唆している。運動野は運動のためだけにあるのではなく，その重要な機能のひとつは，感覚を成立させるために予測を提供することなのである[8]。

4. 内受容感覚の予測的符号化

Feldman-Barrettら（2015, 2016; Ainley et al., 2016; Seth & Friston, 2016）は，視覚や聴覚などの外受容感覚（exteroception）や運動の知覚などの固有感覚（proprioception）だけでなく，内臓や血管など身体内部の感覚である内受容感覚（interoception）も，こうした予測的符号化により成立していると主張する[9]。生体は恒常性（homeostasis）を保って生命を維持し，必要に応じて運動するために，身体状態を適切に制御する必要がある。そのために脳は，身体の現在の状態と望ましい目標状態を表象し，目標を実現するための内的モデルを構築している。そのモデルにより，状況に応じて血圧，血糖値，ホルモンの濃度，免疫機能に関わるサイトカインの濃度などを保つべき適正範囲が定められ，それらのセットポイントが目標として維持される。そこに身体からの信号が入力されると，それが内的モデルによる予測と照合され，両者のずれが予測誤差として検出される。生体は，この予測誤差を最小化することで身体状態を制御しようと努める。予測誤差の最小化のためには，内的モデルの更新と，行動による身体の変容の両方の手段が用いられる。

主観的に経験される身体内部の感覚は，事前分布と身体からの信号入力から事後分布が計算される過程が意識されたものだと考えられる。例えば，腸の蠕動運動は，通常は意識されることはほとんどない。これは普段は，内的モデルによる予測と実際の運動の予測誤差がわずかであるからである。しかし腸に感染が生じて炎症が起これば，予測誤差は増大し，我々は違和感や痛みとしてそれを知覚することになる。そうなると腸への注意により感覚信号の分布の精度が上がり，知覚は鋭敏になる。そのような場合には，腸のわずかな動きでさえ感じられる。このような時に我々は，腹部を手でさすったり押したりして，違和感を確認したり，痛みを鎮めようと試みる。これは能動的推論による予測誤差縮小のための行為であると解釈することができよう。

こうした内受容感覚の予測的符号化において重要なのが，島（insula）と呼ばれる脳部位である（詳しくは，大平，2017を参照）。島は両側の側頭葉の内側に位置する大脳皮質であり，すべての身体からの信号を受ける。島の前部は無顆粒皮質であり，身体信号が入力される後部は顆粒皮質である。また後部島には，内側前頭前皮質（medial prefrontal cortex: MPFC）や前頭眼窩皮質（orbitofrontal cortex: OFC）の無顆粒皮質からも密な神経投射がある。こうした解剖学的事実からFeldman-Barrettら（2015, 2016, 2017）は，前部島，MPFC, OFCなどが内受容感覚の内的モデルを形成し，後部島において身体信号との予測誤差が計算されると主張している。

---

[7] 特定の波長の光によって活性化されるタンパク質分子を，遺伝学的手法を用いて特定の脳部位の神経細胞に発現させ，その機能を光照射により操作する技術。任意の脳部位の機能を，促進あるいは抑制することができる。

[8] こうした知見から，視覚野，聴覚野，体性感覚野，運動野，などのように大脳皮質を機能と対応させて区分する従来の考え方を再考し，予測と予測後差の観点から再概念化すべきであるとも論じられている（Feldman-Barrett, 2017）。

[9] *Philosophical Transaction Royal Society B* 371巻（2016）において，‘Interoception beyond homeostasis: affect, cognition, and mental health（恒常性を超えた内受容感覚：感情，認知，精神的健康）’と題する特集が組まれ，この問題に関する総説論文が複数収められている。

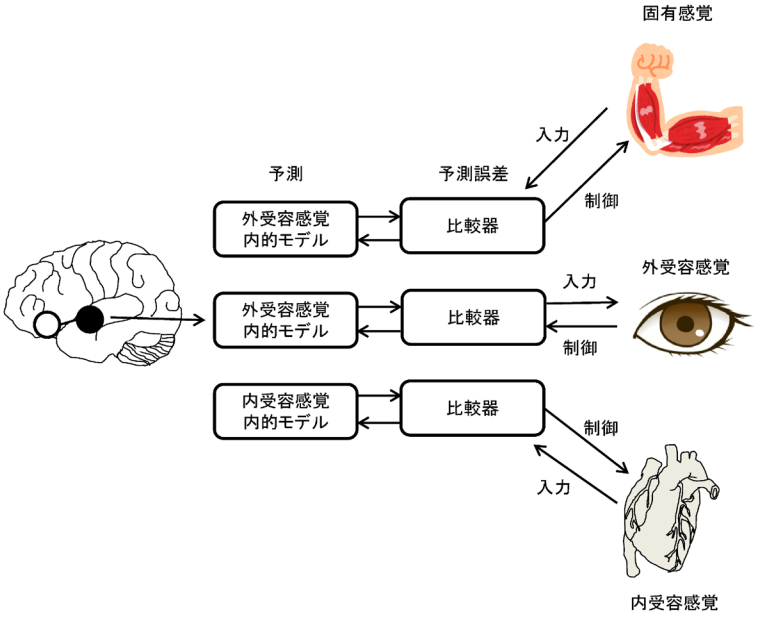

Figure 3.　外受容感覚・固有感覚・内受容感覚を統合するメカニズム
脳の各層に存在する内的モデルを担う部位（前者は白丸で表現されている）の働きにより，身体の位置や運動などの固有感覚，視覚や聴覚などの外受容感覚，内臓感覚などの内受容感覚，の予測が形成される。これらの予測と，それぞれの感覚信号との比較より予測誤差が計算される（予測誤差が計算される部位が黒丸で表現されている）。この予測誤差を縮小するように，内的モデルの更新と，諸感覚の調整が行われる。これらの過程により，諸感覚が統合される。

5.　EPICモデルにおける感情

EPICモデルは，内受容感覚が外受容感覚や固有感覚と統合され，知覚や行動を制御して整合的な状態を維持していると主張する（Figure 3）。このモデルでは，視覚野や体性感覚野において知覚が形成されるとともに，それに付随する身体状態の変化の知覚，すなわち内受容感覚が島皮質において形成される。このため，外受容感覚や固有感覚は，常に何らかの内受容感覚を伴うことになる。言い換えれば，我々が経験する意識は，内受容感覚を軸にして諸感覚を予測的符号化の原理により束ねることで成立すると考えられる。そして我々は，これらの内受容感覚，外受容感覚，固有感覚における予測誤差の和を最小化するように，反応や行動を変化させると予測される。

例えば，通い慣れた道で突然蛇に出くわしたとしよう。その瞬間，視覚野において大きな予測誤差が生じる。するとその対象を視覚的に知覚するとともに，運動野に予測誤差信号が送られ，そこにある内的モデルが更新される。その結果，運動プログラムが発動され，骨格筋や循環器の作動が変化する。そうした反応の予測的符号化によって，体性感覚野では筋のこわばりが知覚され，島皮質では心拍や血圧の上昇などが知覚される。それら一連の処理と平行して，扁桃体では事象の重要性に関する評価が進行し，線条体では価値の計算が進行する。OFCでは事象の置かれた文脈（同じ蛇でも，いつもの道で出会う場合と，初めて行った山中で出会う場合では文脈が異なる）が評価される。さらに，こうした処理における予測誤差を縮小するための能動的推論として，飛び上がる，逃げる，などの行動が前部帯状皮質（anterior cingulate cortex: ACC）などの働きにより選択される。島はこうした一連の処理において，神経ネットワークのハブとして脳と身体の働きを調整し，予測誤差の和を最小化するように働く。Figure 4は，こうした過程の神経メカニズムを表現している。

EPICモデルでは，こうした内受容感覚を基盤として脳に表象された知覚が感情（affect）の本質だと考える。感情は，我々が経験するすべての事象に常に随伴しており，我々が生きている限り，最小から最大のレベルまで連続的に変化しながら常に生じている。また感情は，知覚，認知，運動などと呼ばれる他の精神機能と区別することはできず，それらの精神機能とかなりの程度オーバーラップしている。これらの過程は物理的・生物学的な実体であり，自然物（natural kind）である。感情の多くの部分は意識されず自動的に処理が進行しており，その一部だけが意識される。感情は，快–不快，覚醒–鎮静という2次元平面に射影され経験される。この快–不快，覚醒–鎮静という軸は人間が持つ認識の仕組みであり，自然物では

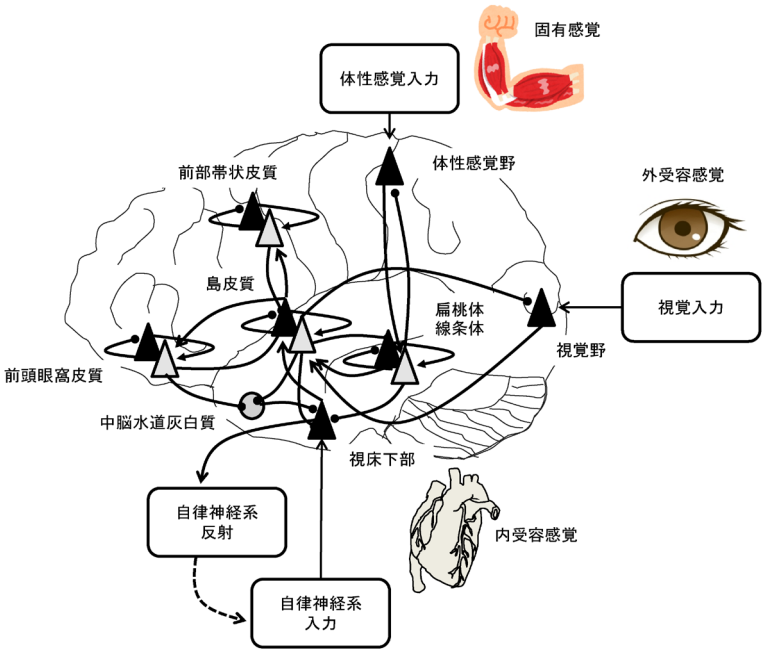

Figure 4.　予測的符号化の神経メカニズム（Seth & Friston, 2016に基づき筆者が作成）
図中の三角形は錐体細胞群を表す。グレーの三角形は予測を計算する細胞群，黒の三角形は予測誤差を計算する細胞群である。島皮質，前頭眼窩皮質，前部帯状皮質は，予測的符号化のハブ領域であり，それらの領域中に予測を計算する細胞群と予測誤差を計算する細胞群を含む。扁桃体では各神経核がそれらの役割を果たす。視覚野や体性感覚野などは顆粒皮質であり，感覚入力の処理と予測誤差の計算を担う。中脳水道灰白質を中心とする領域は内受容感覚信号の精度を調整する。

ない。さらに，言語の機能によりこの連続的な2次元平面はいくつかのカテゴリーに分節され，それらは恐怖，怒り，幸福などと呼ばれる情動（emotion）となる[10]。このように，これらの経験された感情や情動は，自然物ではなく，人間が創った概念として存在している。Feldman-Barrett（2017）は，それはちょうど「お金」が存在するのと同じ仕方であると述べている。特定の脳や身体の活動パターンが，自然物としての特定の情動（例えば恐怖）を創るのではない。恐怖と呼ばれる情動を形成するカテゴリーの輪郭線は任意に引くことができ，恐怖に伴う脳や身体の活動は，個人内でも個人間でも大きな変動がある。脳や身体の活動と，経験される情動は無関係ではないが，両者は別々の原理で作動しているのであり，研究においては，両者は概念的に峻別され検討されるべきであると主張される。

EPICモデルは，情動の末梢起源説（James, 1884）やソマティック・マーカー説（Damasio, 1994）などと共通する，身体の働きと脳におけるその表象を重視する感情理論であると言える。ただし，それらの先行する理論が脳と身体の具体的な働きについては大雑把な質的説明しか与えなかったのに対して，EPICモデルの利点は具体的な処理のメカニズムを表現し，その動きを検討できる可能性を持つことである。例えば，予測的符号化と能動的推論の原理に基づいた意思決定と，そこでのドーパミンの機能を説明する計算論モデルが提案されている（Friston, Schwartenbeck, FitzGerald, Moutoussis, Behrens, & Dolan, 2014）。そうしたモデルを，内受容感覚を包含するように拡張することは可能であろう。モデルが構築できれば，シミュレーションにより脳，身体，行動のダイナミックな変化を可視化できるし，観測されたデータによりモデルの妥当性を評価することもできる。さらには，モデルから現実のデータを予測することも可能になる。システム神経科学においては，少なくとも線虫のようなシンプルな生物に関しては，個々のニューロンのネットワークの挙動を計算論モデルにより表現し，そこから動物の具体的な運動のパターンを説明・予測することが可能になっている（Tsukada, Yamao, Naoki, Shimowada, Ohnishi, Kuhara, Ishii, & Mori, 2016）。

[10] この過程をFeldman-Barrett（2017）は，色の認識を例に挙げて説明している。我々は「虹は七色」と言うが，本来色彩の源である光の波長は連続的であり，虹に色彩の切れ目はない。我々が持つ色のカテゴリー（「赤」や「紫」など）により，その連続体を7つのカテゴリーに切り分けて認識しているのだと主張されている。

6. 予測的符号化による脳，身体，行動の制御

ここでは筆者らの研究を例として，脳と身体の生理的反応のダイナミクスに関する実証的知見を，EPICモデルに照らして検討する。

(1) 刺激－結果の随伴性と脳－身体の活動

筆者らは，刺激と結果との随伴性の評価により，行動，脳活動，自律神経系・内分泌系・免疫系などの生理的反応が，どのように形成されるかを検討してきた（Kimura, Ohira, Isowa, Matsunaga, & Murashima, 2007; Ohira, Fukuyama, Kimura, Nomura, Isowa, Ichikawa, Matsunaga, Shinoda, & Yamada, 2009; Ohira, Ichikawa, Nomura, Isowa, Kimura, Kanayama, Fukuyama, Shinoda, & Yamada, 2010）。これらの研究では，確率学習課題（stochastic learning task）と呼ばれる一種のギャンブル課題が用いられた。実験参加者は各試行において，金銭的な報酬を獲得し損失を避けるために，2つの選択肢から1つを選ぶ。一方の選択肢は有利であり，より高い確率で報酬をもたらし，他方は不利であり，報酬獲得確率は低く操作されている[11]。

この課題の遂行は，強化学習（reinforcement learning）と呼ばれる計算論モデルでよく表現できる（Lee, Seo, & Jung, 2012）。典型的な強化学習アルゴリズムのひとつであるQ学習（Q learning）は，次のように表現される。

$$Q_{(a(t+1))}(t+1)=Q_{a(t)}(t)+\alpha(R(t+1)-Q_{a(t)}(t)), \quad (1)$$

$$P(a(t))=\frac{1}{1+\exp[-\beta(Q_a(t)-Q_b(t))]}, \quad (2)$$

式(1) 中の$Q_{i(t)}(t)$は，2つの選択肢i＝a, bのうちaの，時点tにおける価値を表す。これが内的モデルによる選択肢の報酬に関する予測に該当する。もし次の時点でaを選択したことで報酬$R(t+1)$が得られたとしたら，予測$Q_{a(t)}(t)$との差が予測誤差であり（式(1) では，$R(t+1)-Q_{a(t)}(t)$と表現されている），これを縮小するように価値の更新がなされる。予測誤差が正値の場合は結果が予測より良かったことを意味し，選択肢の価値は上げられる。予測誤差が負値の場合は結果が思ったほど良くなかったことを意味し，選択肢の価値は減じられる。このように，2つの選択肢の価値は刻々と更新されていき，ある時点での意思決定，すなわちaが選択される確率は式(2) のように表される[12]。式(1) のような価値の更新は主に腹側線条体（側坐核（nucleus accumbens））により，式(2) で表現される価値に基づく選択は，背側線条体やACCを含むネットワークにより担われている（Lee et al., 2012）。すなわち，この課題を遂行する間，選択肢の価値に関する予測的符号化が線条体を中心に実行されている。

筆者ら（Kimura et al, 2007）は，この課題において，一方の選択肢が70%，他方が30%で報酬をもたらす条件（強化群）と，参加者の選択とは無関係に強化群とマッチングしたタイミングで報酬を与える統制群[13]を設けた。報酬を受け取る量とタイミングは，両群で全く同じである。しかし，強化群では課題の開始とともに大きな報酬予測誤差が生じ，それは次第に縮小されて学習が進行するのに対して，統制群では学習が困難である。強化群の反応バイアス（有利選択をした確率）は典型的な学習曲線を描き，最終的に80%程度に収束する[14]。統制群の反応バイアス（任意の一方の選択肢を選んだ確率）は課題終了時まで50%程度であり，参加者は最後まで試行錯誤を続けていたことが伺える（Figure 5a）。強化群では，免疫系反応（血液中リンパ球におけるナチュラル・キラー（natural killer: NK）細胞[15]の比率：Figure 5b）にも自律神経系反応（収縮期血圧（Figure 5c）と拡張期血圧（Figure 5d））にも，一貫して強い反応が見られた。これに対して統制群では，いずれの指標においても反応は顕著に抑制されていた。こうした刺激と結果の随伴性による生理的反応のパターン分化は，同様な課題を用いた別の研究でも再現されており（Ohira et al., 2009），頑健な知見である。この課題の遂行時には，Figure 6aに示すように，ACC, OFC，線条体，など，Figure 4に表現されている予測的符号化の

[11] この課題はいわゆるオペラント条件づけ（operant conditioning）と同様な事態であり，ヒトと動物で共通に使用することができるので，両者の比較を容易にするという利点がある。

[12] 式(1) 中のパラメータαは学習率（learning rate）と呼ばれ，報酬の予測誤差により1回につきどれくらい価値を更新するかを制御する。学習率が高いと学習が早いが，高すぎると1回ごとの結果に大きく影響され，課題遂行が不安定になる。式(2) 中のパラメータβは逆温度（inverse temperature）と呼ばれ，選択肢の選択確率に価値の差によりどの程度重みを欠けるかを制御する。βが大きいほど少しでも価値の高い選択肢を選ぶ確率が高くなり（この方略を搾取（exploitation）と呼ぶ）βが小さいと価値の低い選択肢をも選ぶ確率が高くなる（この方略を探索（exploration）と呼ぶ）。

[13] この操作をヨークト（yoked）と呼び，心理学における学習の研究ではよく用いられる。

[14] 確率的な期待値を考えれば，有利な選択肢を100%選ぶ方略が最も適応的である。しかし多くの場合，選択は強化随伴性に近い確率に収束する。これを対応法則（matching law）と呼ぶ。対応法則がなぜ生じるかについては諸説ある（Saito, Katahira, Okanoya, & Okada, 2014）。

[15] 生体内に侵入するどんな抗原にも反応し攻撃を行うことができる，免疫系において最前線防御を担う免疫細胞。感染やストレスに短時間で鋭敏に反応する（Isowa, Ohira, & Murashima, 2004; Kimura, Isowa, Ohira, & Murashima, 2005）。

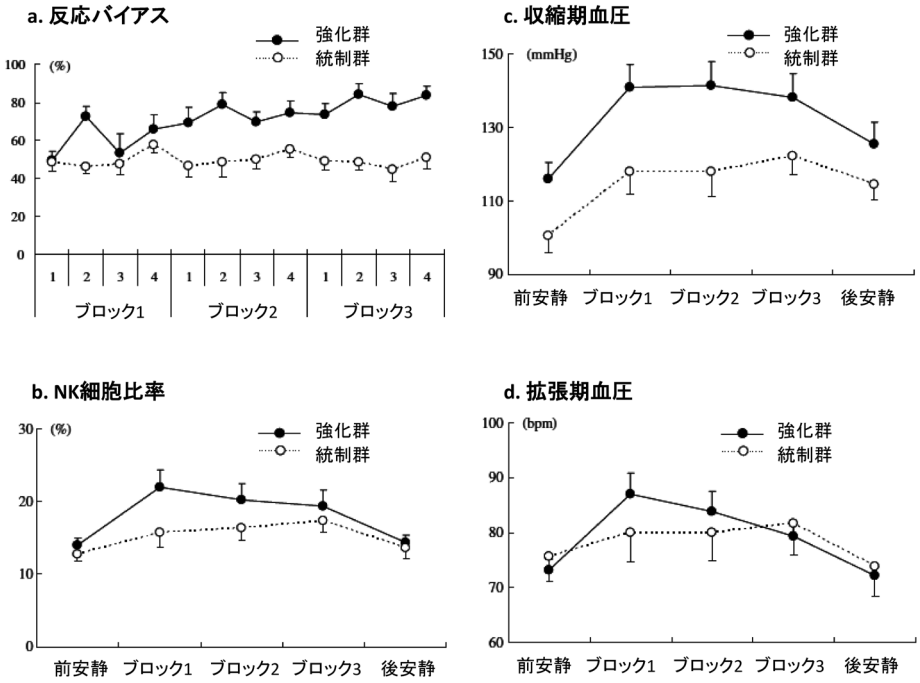

Figure 5. 確率学習課題における反応バイアス，ナチュラル・キラー細胞，血圧の変動（Kimura et al., 2007）
強化群では選択肢により確率的に報酬が与えられ，適応的な意思決定方略を学習することができる。統制群は，参加者の選択とは無関係に報酬が与えられ，学習が不可能である。a. 強化群では有利な選択肢の選択確率（反応バイアス）が漸増し学習が成立するが，統制群では2つの選択肢が等確率で選択され続ける。b，末梢血中のナチュラル・キラー（natural killer: NK）細胞は，強化群では課題開始と共に顕著に増加し，課題中高い水準に保たれる。統制群ではほとんど反応が見られない。自律神経系の指標である，c. 収縮期血圧，d. 拡張期血圧も，同様な反応パターンを示す。

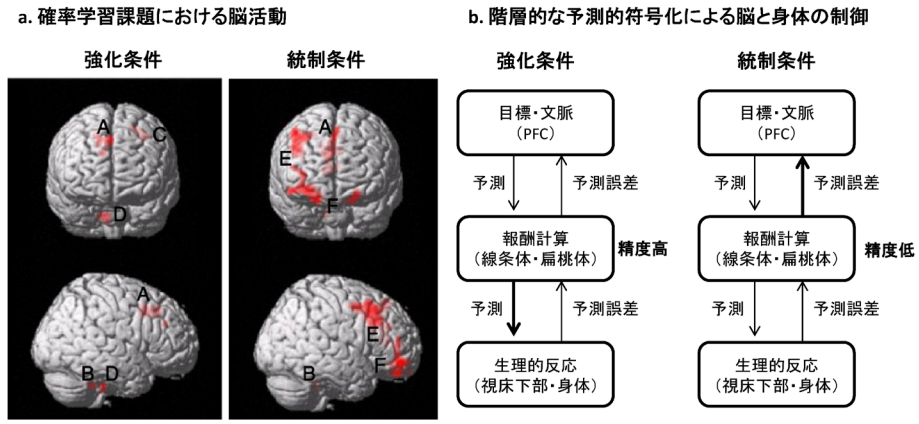

Figure 6. 確率学習課題における脳と身体の活動
a. 陽電子断層撮影法（PET）により測定された，確率学習課題を遂行している際の脳活動（Ohira et al., 2010）。A：前部帯状皮質，B：小脳，C：左背外側前頭前皮質，D：橋，E：右背外側前頭前皮質，F：前頭眼窩皮質。図では見えないが，特に統制群において線条体にも強い活動が観測されている。b. 階層的な予測的符号化の図式的表現。強化条件では報酬計算のモデル精度が高くなり，身体に強い影響を与える。統制条件では報酬計算のモデル精度が低く保たれるために，脳の高次領域では持続的なモデルの更新が行われる。身体には交感神経と副交感神経を介した異なる様相の影響が及ぶために，それらの影響が相殺されて身体への効果は小さくなる。PFC: prefrontal cortex，前頭前皮質。

ハブ領域と整合した脳部位に賦活が観測されている（Ohira et al., 2009, 2010）。また，それらの領域の賦活は，統制群でより強いことから，この群の参加者が単に学習を放棄した結果生理的反応のレベルが低下したわけではなく，報酬予測誤差を縮小する努力が継続されていることが示唆される。

これらの知見を，EPICモデルの観点から解釈してみよう。予測的符号化には階層性があり，報酬計算のシステムは，前頭前皮質（prefrontal cortex: PFC）で担われているより上位の目標や文脈を保持するシステムに対して，予測誤差の信号を提供することによって影響を与える。一方で，身体の状態を制御するより下方のシステムに対しては，予測の信号を与えることによって影響する（Figure 6b: Smith, Thayer, Khalsa, & Lane, 2017）[16]。強化群では，各試行における結果のサンプリングから，線条体における報酬計算が速やかに収束し，そこにおけるモデルの精度が向上していく。この信号は上位システムにおける目標や文脈のモデルにおける予測誤差をも縮小し，安定した対処方略を形成する。一方，身体の活動を制御する下位システムに対しては，Figure 1に示した原理により大きな予測誤差をもたらし，報酬計算モデルの予測に合致するような方向に，身体反応を強く誘導する。この場合には，報酬を獲得するために身体の活動性を増すように，交感神経系活動のセットポイントが上方修正される。Figure 5に示した，免疫系や自律神経系活動の促進は，このようにしてもたらされたのであろう。これに対して統制群では，報酬計算はいつまでも収束しない。これは，上位のシステムに対し予測誤差の信号を送り続けることになり，PFCの各領域に不断にモデルの更新を迫ることになる。Figure 6aに示した，統制条件におけるPFC領域の強い活動は，こうした事態を反映していると解釈できる。一方で，下位システムに対しては，報酬の獲得に向かわせる接近反応としての交感神経系活動を促進する信号と，損失を防ぐための回避反応としての副交感神経系活動を促進する信号が混交した形で発信されることになり，結果としてこれらの信号の効果が相殺するために，免疫系や自律神経系の生理的反応は低い水準に抑制されることになる。実際に統制群においては，心拍変動性（heart rate variability: HRV）[17]を指標として推定された副交感神経系活動が顕著に高まっていた（Ohira et al., 2010）。

このような事態では当然ながら，強化群においては快感情が，統制群においては不快感情が生じるであろう。EPICモデルの観点から考えれば，強化群における快の感情とは，階層的な予測的符号化システムにおける予測誤差の和，すなわち自由エネルギーが縮小されていく変化過程の主観的な経験にほかならない。一方，統制群における不快の感情とは，予測誤差が正負の値を揺れ動き，縮小できないことに伴う主観的経験である。つまり快と不快の感情とは，自由エネルギーの縮小という脳の原理の目標への接近と停滞を意味する。また特に強化群において強く経験される主観的な覚醒の感覚は，身体活動を制御するシステムが報酬計算システムからの予測信号の変化に伴い，各種の生理的反応のレベルを上げることにより予測誤差を縮小しようとする能動的推論の過程が認識されたものである。これらが，我々が主観的に感じる感情（affect）である。さらに，そうした経験を，我々は，意欲，満足，焦り，不安，などの名称を与えることにより認識する。これらが情動（emotion）である。この意味では，感情は予測的符号化の過程から創発された随伴物である。しかし一方で感情は，いったん形成されたならば，おそらく言語や記号を扱う最上位のシステムにおいて予測として機能し，より下位の諸システムの活動を規定すると考えられる（Pezzulo, Rigoli, & Friston, 2015）[18]。

(2)　随伴性の変化への追従：モデルの精度を反映する心拍変動性

自然環境でも社会環境でも，刺激–行動–結果の随伴性は固定されているわけではなく，むしろ時間とともに変動する。動物が，これまで餌が獲れた場所で，次回も餌にありつける保証はない。これまで異性の魅力を惹きつけられたやり方で，次回も同じようにうまくいくとは限らない。生体は環境への適応のために，このような随伴性の変化に鋭敏に追従して自らの行動を変えていく必要がある。こうした能力を目標志向的行動（goal-directed action）と呼び，学習心理学における重要なテーマのひとつとなっている（Pezzulo, van der Meer, Lansink, & Pennartz, 2014）。

目標志向行動は，PFCと背内側線条体において担われていると考えられており，予測的符号化の最上位システムに属する機能である。EPICモデルの観点からは，この最上位システムにおけるモデルの精度が高いほど，予測信号を介する下位のシステムへの影響が強くなり，随伴性の変化に応じて柔軟に行動や，それを支える生理的反応を調整していく能力が高くな

[16] この3層構造は，極めて単純化したモデルである。Smithら（2017）は，8層の構造を提唱し，詳細なメカニズムを論じている。

[17] 心臓の拍動感覚は常に一定のゆらぎを含んでいる。その原因は，圧受容体による血圧の制御と，呼吸による影響などである。このゆらぎの度合いを心拍変動性（heart rate variability: HRV）と呼び，副交感（迷走）神経系の活動水準を反映する指標と考えられている。

[18] 認知心理学や認知行動療法で対象とされる「認知（cognition）」の諸過程の多くは，この機能を扱っているのだと考えることができる。

ることが予測できる（Pezzulo, et al., 2015; Smith et al., 2017）。ここでSmithら（2017）は，HRVが，そうしたPFCを中心とした目標志向行動のモデルの精度を反映するよい指標になると論じている。HRVは心拍間隔のゆらぎであるが，単に心臓活動だけを表すのではなく，脳が身体のさまざまな活動をトップ・ダウン的に制御する広範な機能を反映することが知られているためである（Thayer, Ahs, Fredrikson, Sollers, & Wager, 2012）。

筆者ら（Ohira, Matsunaga, Osumi, Fukuyama, Shinoda, Yamada, & Gidron, 2013）はこうした問題を検討するために，HRVが高い群と低い群の参加者に，刺激−行動−結果の随伴性の変化を扱う確率逆転学習（stochastic reversal learning）と呼ばれる課題を課し，その際の脳活動と各種の生理的反応を観測した。確率逆転学習は上述した確率学習と同様なギャンブリング課題であるが，有利・不利な選択肢が課題途中で突然逆転され，そのため参加者はそれまで優勢であった自らの行動を抑制し，新たな随伴性を再学習することを迫られる。HRV高群は，随伴性の逆転の導入に伴い，自律神経系・内分泌系・免疫系の全ての指標が一貫して高い反応性を示した。さらに，OFC, ACC, 島，線条体（Figure 4と整合した脳部位である）における活動が，それらの生理的反応と顕著な相関を示した。これらの結果はSmithら（2017）の予測と整合して，この群では目標志向行動を担う上位システムが随伴性の変化を敏感に検出することができ，精度の高い予測信号を下位システムに送ることによって各種の生理的反応をトップ・ダウン的に制御しうることを示唆している。このような事態においては，随伴性の変化に伴って大きな予測誤差信号が入力されるために，いったんは不安や焦りなどの不快感情が経験されると思われる。しかし，その変化に追従することにより速やかに予測誤差が縮小されることに伴い，安堵などの快感情へと変化していくと考えられる。これに対してHRV低群は，脳においても生理的反応においても，ほとんど反応が観測されなかった。この群では目標志向行動の機能が低下しているために，随伴性の変化に鈍感であり，刺激−行動−結果の随伴性が変化したにも関わらず，それまでに獲得した習慣的な脳と生理的反応のパターンが持続されたのだと解釈できる。このような場合にはシステムの活動の変化が鈍いために，主観的な感情も鈍磨しているに違いない[19]。

これらの実証的知見は，EPICモデルと整合しているように思われる。将来においては，予測的符号化の原理に基づいて，ベイズ的推定を実行するような計算論モデルを構築して，そこに既存のデータを当てはめてその妥当性を検討することが可能になるであろう。深い思索に基づいて，適切な水準で抽象化された理論は，研究者にインスピレーションをもたらし，研究を強く刺激する力がある。EPICモデルはそうした可能性を感じさせる理論的枠組みであるが，その真価が問われるのはこれからである。

[19] 未発表データであるが，実際にHRV低群においては，主観的ストレスの評定値が一貫して低かった。

7. 結語

> 感情とは我々の身体の活動能力を増大しあるいは減少し，促進しあるいは阻害する身体の変状，また同時にそうした変状の観念であると解する。
>
> Per Affectum intelligo Corporis affectiones, quibus ipsius Corporis agendi potentia augetur, vel minuitur, juvatur, vel coercetur, et simul harum affectionum ideas.

17世紀の哲学者スピノザ（畠中尚志訳，1970）は，感情をこのように定義した。感情とは静的な状態を指すのではなく，身体の活動水準の動的な変化なのであり，精神が感受するのもその変化であるという意である。予測的符号化の原理は，事前分布による予測が観測により事後分布へと変換される不断の変化を記述する論理であり，まさにその変化のあわいに，我々の脳や身体の活動が制御されて行動が惹起され，同時に感情の主観的経験が創発されると説く。スピノザがこの原理を看取していたとすれば，慧眼であると言わざるを得ない。

21世紀の感情科学は，ようやくこの原理を顕在的に明示化する試みに着手し，それを定式化して実証的に扱う手掛りを得た。感情の予測的符号化の理論が，今後どのような展開を辿るのかはいまだ予見できないが，感情の研究に関心を持つ者は，その行方に注目せねばならないだろう。

引用文献

Ainley, V., Apps, M. A. J., Fotopoulou, A., & Tsakiris, M. (2016). 'Bodily precision': A predictive coding account of individual differences in interoceptive accuracy. *Philosophical Transactions of the Royal Society B*, **371**, 20160003.

Barbas, H., & Rempel-Clower, N. (1997). Cortical structure predicts the pattern of corticocortical connections. *Cerebral Cortex*, **7**, 635–646.

Damasio, A. R. (1994). *Descartes' error: Emotion, reason and the human brain*. New York, NY: Grosset Putnam.

Feldman-Barrett, L. (2017). *How emotions are made: The secret life of the brain*. New York, NY: Houghton Mifflin Harcourt.

Feldman-Barrett, L., Quigley, K. S., & Hamilton, P. (2016). An active inference theory of allostasis and interoception in depression. *Philosophical Transactions of the Royal Society B*, **371**, 20160011.

Feldman-Barrett, L., & Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nature Review Neuroscience*, **16**, 419–429.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Review Neuroscience*, **11**, 127–138.

Friston, K., Kilner, J., & Harrison, L. (2006). A free energy principle for the brain. *Journal of Physiology Paris*, **100**, 70–87.

Friston, K., Schwartenbeck, P., FitzGerald, T., Moutoussis, M., Behrens, T., & Dolan, R. J. (2014). The anatomy of choice: Dopamine and decision-making. *Philosophical Transactions of the Royal Society B*, **369**, 20130481.

Helmholtz, H. (1962(1866)). *Concerning the perceptions in general. Treatise on physiological optics. III*. New York, NY: Dover.

Isowa, T., Ohira, H., & Murashima, S. (2004). Reactivity of immune, endocrine and cardiovascular parameters to active and passive acute stress. *Biological Psychology*, **65**, 101–120.

James, W. (1884). What is an emotion? *Mind*, **9**, 188–205.

Kimura, K., Isowa, T., Ohira, H., & Murashima, S. (2005). Temporal variation of acute stress responses in sympathetic nervous and immune systems. *Biological Psychology*, **70**, 131–139.

Kimura, K., Ohira, H., Isowa, T., Matsunaga, M., & Murashima, S. (2007). Regulation of lymphocytes redistribution via autonomic nervous activity during stochastic learning. *Brain, Behavior, and Immunity*, **21**, 921–934.

Lee, D., Seo, H., & Jung, M. W. (2012). Neural basis of reinforcement learning and decision making. *Annual Review of Neuroscience*, **35**, 287–308.

大平英樹（2017）．内受容感覚に基づく行動の制御 *BRAIN and NERVE*, **69**, 383–395.

Ohira, H., Fukuyama, S., Kimura, K., Nomura, M., Isowa, T., Ichikawa, N., Matsunaga, M., Shinoda, J., & Yamada, J. (2009). J. Regulation of natural killer cell redistribution by prefrontal cortex during stochastic learning. *NeuroImage*, **47**, 897–907.

Ohira, H., Ichikawa, N., Nomura, M., Isowa, T., Kimura, K., Kanayama, N., Fukuyama, S., Shinoda, J., & Yamada, J. (2010). Brain and autonomic association accompanying stochastic decision-making. *NeuroImage*, **49**, 1024–1037.

Ohira, H., Matsunaga, M., Osumi, T., Fukuyama, S., Shinoda, J., Yamada, J., & Gidron, Y. (2013). Vagal nerve activity as a moderator of brain–immune relationships. *Journal of Neuroimmunology*, **260**, 28–36.

Pezzulo, G., Rigoli, F., & Friston, K. (2015). Active Inference, homeostatic regulation and adaptive behavioural control. *Progress in Neurobiology*, **134**, 17–35.

Pezzulo, G., van der Meer, M. A., Lansink, C. S., & Pennartz, C. M. (2014). Internally generated sequences in learning and executing goal-directed behavior. *Trends in Cognitive Sciences*, **18**, 647–657.

Manita, S., Suzuki, T., Homma, C., Matsumoto, T., Odagawa, M., Yamada, K., Ota, K., Matsubara, C., Inutsuka, A., Sato, M., Ohkura, M., Yamanaka, A., Yanagawa, Y., Nakai, J., Hayashi, Y., Larkum, M. E., & Murayama, M. (2015). A Top-Down Cortical Circuit for Accurate Sensory Perception. *Neuron*, **86**, 1304–1316.

Saito, H., Katahira, K., Okanoya, K., & Okada, M. (2014). Bayesian deterministic decision making: A normative account of the operant matching law and heavy-tailed reward history dependency of choices. *Frontiers in Computational Neuroscience*, **8**, 18.

Seth, A., & Friston, K. J. (2016). Active interoceptive inference and the emotional brain. *Philosophical Transactions of the Royal Society B*, **371**, 20160007.

Smith, R., Thayer, J. F., Khalsa, S. S., & Lane, R. D. (2017). The hierarchical basis of neurovisceral integration. *Neuroscience and Biobehavioral Reviews*, **75**, 274–296.

スピノザ（畠中尚志訳）（1970）．『エチカ（上下）』岩波書店

Thayer, J., Ahs, F., Fredrikson, M., Sollers, J. J. 3rd, & Wager, T. D. (2012). A meta-analysis of heart rate variability and neuroimaging studies: Implications for heart rate variability as a marker of stress and health. *Neuroscience and Biobehavioral Review*, **36**, 747–756.

Tsukada, Y., Yamao, M., Naoki, H., Shimowada, T., Ohnishi, N., Kuhara, A., Ishii, S., & Mori, I. (2016). Reconstruction of spatial thermal gradient encoded in thermosensory neuron AFD in Caenorhabditis elegans. *Journal of Neuroscience*, **36**, 2571–2581.