

マインドフルネスのメカニズムの 予測符号化モデルに基づく理解¹⁾

高橋 徹¹・萩島 大凱²

¹早稲田大学

²高知工科大学

Understanding of the mechanisms of mindfulness in the perspective
of the predictive coding model

Toru TAKAHASHI¹ and Hiroyoshi OGISHIMA²

¹Waseda University

²Kochi University of Technology

Mindfulness affects a variety of psychological problems, and many psychological concepts have been introduced to explain the mechanism of these effects. However, a unified understanding has not been attained. In this paper, we discuss the mechanism of mindfulness from the perspective of the predictive coding model, which has been attracting attention in recent years as a potential unified theory to explain various brain functions. Based on our review, we found that mindfulness is characterized by (1) predominantly minimizing prediction errors via perceptual inference, which updates predictions based on sensations rather than by active inference (which attempts to modify sensations to match predictions); (2) increasing the precision of sensory input by attention, and (3) decreasing the precision of prior predictions by not weighting past experiences. These features may lead to rapid minimization of prediction error and an accurate understanding of the inner and outer environment. By redefining mindfulness in the predictive coding model, we discuss the possibility of understanding the mechanisms of mindfulness from a unified perspective that explains human perception, behavior, cognition, and emotion.

Key words: mindfulness, predictive coding, free energy principle, prediction error, perceptual inference, active inference

キーワード：マインドフルネス，予測符号化，自由エネルギー原理，予測誤差，知覚的推論，能動的推論

1. マインドフルネスの効果

マインドフルネスに基づいた介入が，広範な心理的な問題に有効であることを示すエビデンスが蓄積してきている。マインドフルネスの実証研究の草分けである Jon Kabat-Zinn は，マインドフルネスを「意図的に，瞬間瞬間の体験に対して，評価判断することなく，注意を向けることによって得られる気づき」と定義しており（Kabat-Zinn, 2003），そのマインドフルネスを訓練する介入として，マインドフルネスストレス低減法

（Mindfulness-based stress reduction: MBSR; Kabat-Zinn, 1990）やマインドフルネス認知療法（Mindfulness-based cognitive therapy: MBCT; Segal, Williams, & Teasdale, 2002）等の標準プログラムが開発されてきた。これらのマインドフルネスに基づいた介入は，うつや不安を始めとした種々の心理的問題に小～中程度に有効であることがメタ解析によって示されている（Goldberg et al., 2018）。近年では，非臨床的なセッティングでの介入で，一般人あるいは日常的なストレスを抱えている人を対象にした介入研究のメタ解析がなされ，何も介入を行わない群に比べて，うつ，不安，ストレスに対して中程度の効果，Well-being に対して小さい効果があることが示されている

1) 本論文の執筆にあたり，第1著者と第2著者は筆頭著者として同等の貢献をした。

(Galante et al., 2021)。また、慢性疼痛やがんなどの身体疾患患者に対するマインドフルネス介入の効果研究も多くなされているが、主に検討されているアウトカムとしては、身体疾患に起因する苦悩 (Distress)、ストレス、うつ・不安状態などであり、やはり心理的症状を改善するというエビデンスが多い (高橋・熊野, 2018)。このような特定の病態によらない広範な心理的問題への効果が得られる理由を明らかにするために、マインドフルネス介入のメカニズム研究が数多くなされてきている。

マインドフルネスのメカニズム研究には、マインドフルネスという心の状態自体が、どのような構成要素で成っているかを明らかにする方向性と、マインドフルネス介入が他の心理機能にどのような影響を与えることで、最終的なアウトカムを改善するのかを明らかにする方向性があると言える。

本稿では、マインドフルネスの構成要素とメカニズムに関するこれまでの研究を簡潔に紹介した上で、脳の作動原理として注目されている予測符号化モデル、ないしは自由エネルギー原理の観点から、マインドフルネスを捉えようとする近年の動向を概観する。最後に、予測符号化の観点からマインドフルネスを再定義し、その可能性と限界について著者らの考察を述べる。

2. マインドフルネスの構成要素

マインドフルネスのメカニズム理論を初期に提唱した Shapiro et al. (2006) は、マインドフルネスは単に瞑想を指すのではなく、意識の状態のことを指していると述べた上で、本論文 (Shapiro et al., 2006) の意図は、このマインドフルネスという特定の状態の説明を洗練させ、「マインドフルネスとは正確には何なのか？そしてそれはどのように機能するのか？」という問いを探求することだとした。

Shapiro et al. (2006) は、Kabat-Zinn (1994) の有名なマインドフルネスの定義 “paying attention in a particular way: on purpose, in the present moment, and non-judgmentally” (意図的に、現在の瞬間に、評価判断せずに注意を向けること) は、次の3つのマインドフルネスの原理・公理 (Axioms) を含

んでいるとした。

1. “On purpose” 意図 (Intention)
2. “Paying attention” 注意 (Attention)
3. “In a particular way” 態度 (Attitude)

この3つの構成要素から、マインドフルネスがどのように作用するのか、推論することができる。意図、注意、態度は、分離したプロセスあるいは段階ではなく、1つの循環的なプロセスの中で密接に関連しあっている側面であり、同時に起こり得る。マインドフルネスはそのような瞬間瞬間のプロセスであると述べられている (Shapiro et al., 2006)。

まず意図は、なぜ (Why) マインドフルネスの実践をするのかという動機や目的と関連している。Kabat-Zinn (1990) は、「意図は、何が可能となるかのステージをセッティングする (著者訳)」(p. 32) と述べた上で、「かつて瞑想は実践するだけで成長や変化をもたらすと考えていたが… (中略) …ある種の個人的なヴィジョン (Personal vision) が必要だと考えるようになった (著者訳)」(p. 46) と述べている。Shapiro (1992) の研究によると、瞑想者が実践を継続するにつれて、彼らの意図は、自己制御 (Self-regulation) から自己探求 (Self-exploration)、最終的には自己解放 (Self-liberation) までの連続体に沿って転換していきことが見出されている。そしてこれらの意図はアウトカムと関連していた、つまり自己制御をゴールとしていた者は瞑想実践によって自己制御を達成し、自己探求をゴールとしていた者は自己探求を達成し、自己解放をゴールとしていた者は自己解放と慈悲的な奉仕へと向かったという (Shapiro, 1992)。このように、同じ瞑想実践を行うにしても、どのような意図を持つかによって得られるものが違うという意味で、意図は極めて重要であるということだと考えられる。

次に注意は、マインドフルネス実践の文脈では、瞬間瞬間の内的・外的経験の働きを観察することと関わっており、対象それ自体が現れるように、あらゆる解釈を中止して体験それ自体に注意を向けることであると説明される (Shapiro et al., 2006)。注意の要素は、マインドフルネスとは何 (What) をすることなのか、という部分を説明し

ていると言える。

そして態度は、どのように (How) 注意を向けるか、という部分にあたる。注意は、冷たく、批判的な性質を有することもできるし、「愛情のこもった、思いやりのある性質… (中略) …寛大な感覚、友好的な態度、関心 (著者訳)」(Kabat-Zinn, 2003, p. 145) を有することもできる (Shapiro et al., 2006)。このように注意の向け方いくつかの様態がある中で、マインドフルネス実践では、期待や評価をせずに、体験に対してオープンで、優しさや好奇心をもって受容的な態度で注意を向けることを訓練する (Shapiro et al., 2006)。

このようなマインドフルネスの構成要素を明らかにする理論的検討や質問紙研究 (e.g., Baer et al., 2006; Tran, Glück, & Nader, 2013) を経て、現在ではマインドフルネスは (1) 現在の瞬間の体験を観察する注意と、(2) 体験を受容する態度、という2つの構成要素から成るという捉え方 (Bishop et al., 2004; Lindsay & Creswell, 2017) が主流であると言える。Lindsay and Creswell (2017) が提唱した Monitor and Acceptance Theory では、このマインドフルネスの2つの要素である観察のスキルと、受容のスキルは同時に訓練されるとしながらも、観察スキルの方が先に上達し、受容のスキルの上達はより時間がかかるなどの知見 (Baer, Carmody, & Hunsinger, 2012; Desbordes et al., 2015) を元に、観察スキルと受容スキルは区別されるものであるとした。この理論の特徴の1つは、観察スキルのみの上達は、感情的でない文脈では認知機能を向上させるが、ネガティブ・ポジティブの両感情を増強させ症状を悪化させてしまう可能性もあり、観察スキルだけでなく受容スキルも向上させることで、感情的な反応性を下げることができ、効果をもたらすとしたことである。この Monitor and Acceptance Theory は、これらのマインドフルネスの構成要素自体が、直接ストレスに関連した健康指標を改善するという形で、マインドフルネスのメカニズムを説明している (Lindsay & Creswell, 2017)。

3. マインドフルネス介入のメカニズム研究

マインドフルネス自体の構成要素を明らかにしていく研究に加えて、マインドフルネス介入がど

のような変数 (マインドフルネスを含む) の変化を介して、うつや不安、ストレスなどの症状を低減するのかを明らかにしようとする研究が多くなされてきた。主に検討されてきた媒介変数としては、マインドフルネス (Baer et al., 2006; Brown & Ryan, 2003) の向上、反すうや心配などのネガティブな反復的思考 (Ehring et al., 2011) の低減、脱中心化 (思考や感情が一過性の心的事象だと観察すること; Fresco et al., 2007) の向上、認知的反応性 (抑うつ状態がネガティブな思考パターンを再活性化させる程度; Raes et al., 2009) の低減、セルフコンパッション (自身の苦しみに開放的であり、それを回避したりすることなく、和らげようとする; Neff, 2003) の向上がある。マインドフルネス介入のメカニズムを明らかにするために設計されたランダム化比較試験に限定した系統的なレビューによると、これらの変数の中でも、マインドフルネスとネガティブな反復的思考が、多くの研究で繰り返し、マインドフルネス介入と臨床症状の改善を媒介する変数であることが示されている (Alsubaie et al., 2017; Gu et al., 2015)。

また脳機能・構造の変化と対応づけながら、マインドフルネス介入は注意制御や感情制御、自己感の変容などの変化を伴って、自己制御や心身の健康に寄与する (Tang, Hölzel, & Posner, 2015; Hölzel et al., 2011) という説明もなされている。例えば、注意制御や感情制御の向上は、背外側前頭前野 (Allen et al., 2012) や前部帯状皮質 (Hölzel et al., 2007) の活性化と関連し (特に身体感覚への気づきの場合は前部島; Farb, Segal, & Anderson, 2013)、自己感の変容は後部帯状皮質の活動低下 (Brewer et al., 2011) 等が関連している可能性があるという説明されている (Tang et al., 2015; Hölzel et al., 2011)。しかし、マインドフルネス介入がこういった脳機能や構造の変化をもたらす、その変化が症状の改善に繋がっているという媒介関係を検討した研究は未だ多く見られないことが指摘されている (Alsubaie et al., 2017)。

このように概観すると、マインドフルネスのメカニズムを説明するために、多くの心理的概念が導入され複雑化しているとともに、必ずしもそれらの概念と対応する脳機能・構造の変化が、症状改善に繋がるプロセスが説明されているとは言えないと考えられる。このようなメカニズム仮説の

乱立は、マインドフルネス介入自体が、実際に複雑なプロセスを経て、症状の改善をもたらしている実態を反映しているとも考えられる。しかし本稿では、より統一的な枠組みからマインドフルネスのメカニズムを理解するために、近年人間の知覚、行動、認知、感情等の多様な機能を統一的に説明できる可能性が示唆されている予測符号化モデルの観点から考察を行う。予測符号化モデル、またそれを包括する自由エネルギー原理は、脳は外界(身体を含む)を予測、推測する器官であり、予測と実際の感覚入力との差(正確には自由エネルギー)を最小化するという原理に従って、知覚も行動も生成していると説明する理論である(Friston, 2010)。このような脳の一般的な作動原理の中で、マインドフルネスという状態がどのように位置付けられるかを検討することで、マインドフルネスをより基礎的な研究や知見と接続して検討することが可能になるというメリットがあると考えられる。

そこで以下では、予測符号化モデルと自由エネルギー原理の基礎を説明した上で、マインドフルネスを予測符号化モデルから理解しようと試みている論考を紹介する(Farb et al., 2015; Manjaly & Iglesias, 2020; Lutz, Mattout, & Pagnoni, 2019)。

4. 知覚と推論

脳は頭蓋骨の中に隠されており、直接に外界と接触できない。脳それ自身が何かを見たり、触ったりすることはできない。自分の身体でさえも例外ではない。脳は自分の身体すらも外界としてのみしか接することができない。したがって脳が外界を知るには、自身に入力される電気インパルスという単なる電気信号に頼らざるを得ない(Bear, Connors, & Paradiso, 2020)。それにもかかわらず、電気インパルスは必要な情報を的確に伝えるわけではない。もし知覚対象から常に一定の刺激量を与えられていたとしても、電気インパルスには必ずノイズが含まれ(Anderson et al., 2000)、等価な情報は与え続けられない。

このように、外界に直接アクセスできないだけでなく、与えられる情報にはノイズが多いという制約下で、脳には外界を知覚するという課題が課されている。Hermann von Helmholtz (1821-1894)



図1 白黒の模様

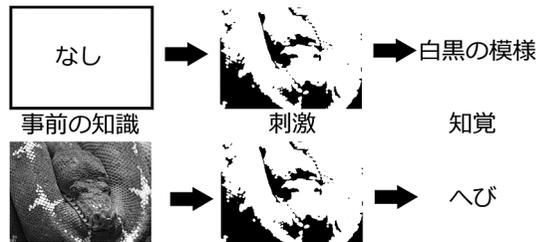


図2 事前知識によって知覚が変わる例

は、知覚対象を直接には感知できないにもかかわらず、安定した世界像を成立させる我々の知覚過程を「無意識的推論」と呼んだ(Helmholtz, 1925)。すなわちHelmholtzは、知覚も思索などと同様に、事前の知識や過去経験に依存した無意識的な推測であると理解した。

図1は「あるもの」を示している。おそらく多くの人が、ただの白黒の模様として知覚する。しかし、図2の写真を見ると認識が変わる。以後は図1を見るたびに「あるもの」が浮かびあがる。ここで特筆すべきは、知覚が変更される前後で、刺激は同一であったことである。このように我々の知覚は、無意識のうちに、事前経験に基づいた推論によって得られているという見方が、近年になって再び注目されはじめている。

5. 予測符号化モデルとは

前節のような脳が推論するプロセスを定式化したものは、予測符号化モデルと呼ばれる。予測符号化モデルとは、簡単に説明するならば、脳は事前経験に基づき外界についてのモデルを生成しており、このモデルの生成過程を「知覚」とする理論である。脳はこのモデルを使って外界から与えられるべき感覚を予測し、実際の感覚とずれがある場合は、モデルを更新し最新に保つ。そのこと

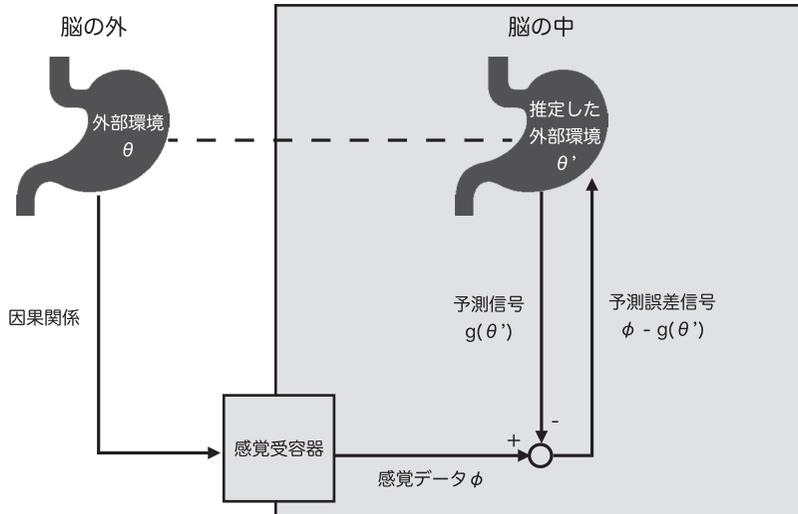


図3 予測符号化による外部環境推論のモデル(乾・阪口(2020)に基づき作成)。外部環境 θ (ここでは胃)の感覚受容器から感覚データ ϕ が入力される。一方脳内では、外部環境を推定したモデル θ' による予測信号 $g(\theta')$ によって感覚データ ϕ が予測される。予測出来なかった部分は、予測誤差として推論モデル θ' に輸入され、モデル θ' の更新が行われる。

によって、刻々と変化する外界を知覚し続けることが可能になる(Jamieson, 2016)。

例えば図3では、予測符号化の原理によって胃を知覚する様子を記述している。ここでは、胃を脳の外部にある環境として見立てている。まず、外部環境である胃 θ から、感覚受容器を通して、感覚 ϕ が入力される。一方脳内では、事前の経験を通して推論された胃のモデル θ' が存在しており、この推論モデル θ' がトップダウンに予測信号 $g(\theta')$ を送ることによって、感覚 ϕ を予測している。そして、予測された感覚 $g(\theta')$ と実際の感覚 ϕ を比較し、予測出来なかった部分はボトムアップに推論モデル θ' に送られ、モデル更新のための情報として用いられる。このように、脳は入力された感覚信号を単にボトムアップに処理するのではなく、トップダウンな予測という処理も行うことによって、ノイズの影響を抑えた効率的な情報伝達を実現している(Huang & Rao, 2011)。

前述したように、予測符号化モデル自体はHelmholtzの無意識的推論に原型がみられる。しかし、21世紀に入り再度脚光を浴び始めたことに関しては、Rao and Ballard(1999)が視知覚の原理を予測符号化モデルによって定式化したことが大きい(Friston, 2018)。視覚は詳細にその神経回路が特定されている知覚領域ではあるが

(Felleman & Essen, 1991), Rao and Ballard(1999)はその階層性に着目した。例えば、網膜に入力された刺激は、脳の視床領域にある外側膝状体を通り、大脳皮質へと投射される。次いで大脳皮質においては、低次元階層である一次視覚野(例えば、ある特定の傾きの線分のみを検出を担うといった働きをもつ)を通り、より高次元階層である二次視覚野(例えば、より複雑な形状や輪郭の特徴に反応する)へと投射される。最終的には、より高次元階層である上側頭野(例えば、視覚的な動きに反応する)などへと到達する。そして特徴的なことに、それぞれの階層は相互に情報のやりとりを行う(Huang & Rao, 2011)。

しかしながら、脳が外界から与えられた電気信号を処理した結果、知覚が生成されると素朴に理解するならば、低次から高次階層へのボトムアップな情報伝達のみで十分なのではないだろうか。高次から低次階層へのトップダウンな情報伝達は、これまでの理解では運動指令が送られていると考えられていた(乾・阪口, 2020)。もしそうならば、視覚という知覚に特化したシステムにおいては、高次階層からのトップダウンな情報伝達はどのような役割を持つのか不明である。

Rao and Ballard(1999)は、このトップダウンな信号を、低次階層の活動を「予測」する信号だと

した。そしてこの想定に立ち、低次から高次階層へのボトムアップ信号は、予測した活動のうち予測できなかった部分である「予測誤差」を伝えているとした。実際に、Rao and Ballard (1999) は、このような予測符号化モデルに基づいたニューロンモデルを数式化することで実装した結果、このニューロンモデルが画像にさらされると、第一次視覚野の受容野内において光刺激が処理される様相と近い反応を示すことを明らかにした。

このように、予測符号化モデルによって、視知覚における多くの疑問が解決される可能性が示された。予測符号化モデルを採択すると、「どのようにノイズが多い感覚入力の信号から安定した世界像を構築するか」という情報処理的な疑問だけでなく、「なぜ感覚処理を行う皮質領野において上位階層から下位階層への投射が存在するのか」という脳構造的な疑問が解決される可能性がある。しかもそれだけでなく、ボトムアップな信号は予測誤差を伝えているのみであるため、脳は従来想定していたよりも情報伝達をスリム化しており、より効率よく情報を伝えている可能性も示唆される。

しかしながら、今日では予測符号化モデルは視知覚の原理を説明するにとどまらず、知覚、行動、学習、意思決定、感情形成などの様々な脳機能を理解する説得力をもった仮説として積極的に検討されている。このようなかつてない学際的広がりを見せるに至った理由については、Karl Fristonの自由エネルギー原理を理解する必要がある。

6. 自由エネルギー原理とベイズ推定

Karl Fristonは、Rao and Ballard (1999) に従うならば、個々のニューロンの活動や階層間の情報伝達を説明するにあたって、脳はただある1つの働きのみを行っている想定すればよいと指摘した (Friston, 2018)。その働きこそが、「予測誤差を最小化する」ことであった (Friston, 2018)。やがて、彼は予測誤差の最小化をエネルギー関数の最適化として理解し、より一般化することで、自由エネルギーという概念を提唱するに至った (Friston, 2018)。

自由エネルギー原理とは、Fristonによって提唱された、行動、知覚、学習を統一的に説明するこ

とを試みる理論的枠組みである (Friston, 2010)。自由エネルギー原理と Rao and Ballard (1999) の予測符号化モデルとの最も大きな違いは、理論の適用範囲が知覚という受動的プロセスにとどまらず、行動という能動的プロセスにまで至ることである。

Fristonが予測符号化モデルを一般化するにあたって、脳が予測誤差を最小化する方法として、ベイズ脳という考え方が用いられた。これは脳が外界を推論するにあたって、ベイズ推定という確率計算を行なっているとする説である (Buckley et al., 2017)。なお、ベイズ推定の式は以下である。

$$P(\text{外界 } \theta | \text{ 感覚データ } \phi) = \frac{P(\text{感覚データ } \phi | \text{ 外界 } \theta) P(\text{外界 } \theta)}{P(\text{感覚データ } \phi)} \quad (1)$$

この式は端的には、外界についてのモデルを得られた感覚をエビデンスとして最適化するということを示している。この式を概念的にあらわしたものは、図4である。

この式をより詳細に説明する。左辺 $P(\theta | \phi)$ は事後確率分布であり、ある感覚データ ϕ が得られた際に、脳が事前経験に基づいて事後的に推論した外界 θ がどの程度観察されるかについての信念 (Belief)²⁾ を示す。したがって、事後モデルと呼ばれることもある。一方で、右辺 $P(\theta)$ は事前確率分布であるが、ある外界状態 θ が観察される確率についての事前の信念を表している。したがって、事前モデルとも呼ばれる。右辺 $P(\phi | \theta)$ は、外界状態が θ であった場合に、感覚データ ϕ が得られるであろう確率を示している。したがって、ある意味で外界状態が θ であるエビデンスを示しているとも言え、外界状態が θ である尤もらしさを示すとも考えられるため尤度と呼ばれている。

右辺の分母は周辺尤度と呼ばれるものである

2) このように、ベイズ統計学においては、事後 (もしくは事前) 確率分布を信念と呼ぶ。これは、事後 (もしくは事前) 確率分布は、ある外界についての主観的な確率分布 (Subjective probability)、すなわち「経験に基づいた推測」であるからである。認知行動療法等における「信念」とは異なる意味であるが、大平 (2017) も述べているように、予測符号化処理の上位階層の事前信念は、いわゆるスキーマのような意味内容を伴った「信念」に相当している可能性もある。

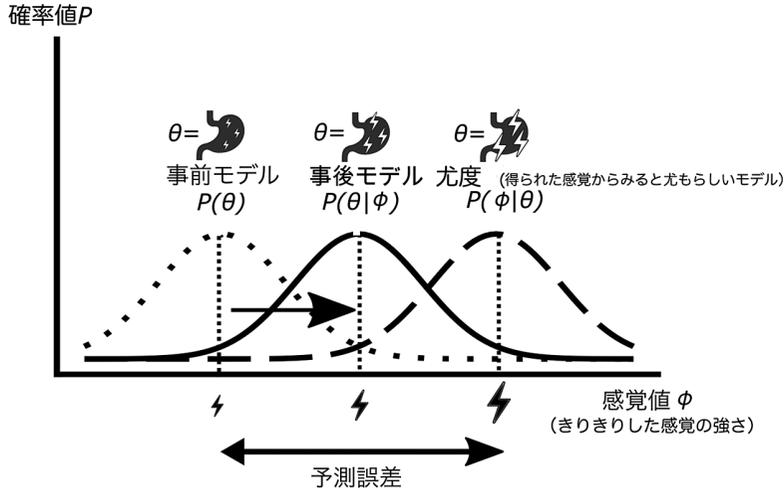


図4 ベイズ推定概念図。事後モデル P (外界 θ | 感覚データ ϕ) は、事前モデル P (外界 θ) と尤度 P (感覚データ ϕ | 外界 θ) から推論された結果として成立する。ここでは、胃の状態を θ とした時に、きりきりとした感覚 ϕ を観察したことから、胃の状態が推論される様子を示している。

が、一般には導出が難しいので、簡単化のために以下の式で表されることがある。

$$P(\text{外界}\theta | \text{感覚データ}\phi) \propto P(\text{感覚データ}\phi | \text{外界}\theta)P(\text{外界}\theta) \quad (2)$$

\propto は左辺と右辺は比例関係にあることを表している。したがって、直感的に述べるならば、ベイズ推定とは、

$$\text{事後モデル} \propto \text{得られた感覚のエビデンス} \times \text{事前モデル} \quad (3)$$

という規則によって、我々の事後モデル、すなわち知覚が得られるというものである³⁾。つまりベ

3) (3)式はベイズ推定式を予測符号化モデルにしたがって直感的に理解した場合であるため、一般のベイズ統計学における理解と異なる可能性がある。例えば、前述したように予測符号化モデルにおける研究では、尤度を外界状態 θ であるエビデンスとして理解するものがある(e.g., Barrett, Quigley, & Hamilton, 2016)が、ここでいうエビデンスは、ベイズ統計学におけるエビデンス(Evidence)とは異なった意味である(ベイズ統計では周辺尤度をエビデンスと表現する)。自由エネルギー原理において尤度は、「ある外界状態 θ と対応する脳状態から予測された感覚」を平均、「実際の得られた感覚」と「ある外界状態 θ と対応する脳状態から予測された感覚」の差の分散(すなわち、感覚のノイズ)を分散とした正規分布によって定義される(Buckley et al., 2017)。したがって、ある意味で外界状態を θ としてよいかを反映しているのもであると思なせるため、尤度がエビデンスとして表現されることがあると考えられる。

イズ推定式は、我々の知覚(事後モデル)は、現在の経験(尤度)をエビデンスとして、過去経験(事前モデル)の正確さを評価し最適化することによって推論されていることを示すものである。

ベイズ脳的働きの直感的な理解のために、先ほどと同様に胃を例に挙げる(図4)。例えば、胃がきりきりするという状態を考える。これをベイズ推定式にあてはめると以下のような式となる。

$$P(\text{胃の状態} | \text{きりきり}) \propto P(\text{きりきり} | \text{胃の状態})P(\text{胃の状態}) \quad (4)$$

$P(\text{胃の状態})$ は、胃の状態の事前モデルなので、普段通りの比較のおだやかな胃の状態がモデル化されているはずである。一方で、普段とは違う信号(きりきりとした感覚)が入力されたため、 $P(\text{感覚データ} | \text{外界})$ によって、その感覚を与える尤もらしい胃の状態を推測してみる。すると、どうやら事前モデルとは大きく異なるようだ。このようにして、ある感覚信号が入力された時に、それをエビデンスとして胃の状態 $P(\text{胃の状態} | \text{きりきり})$ を推論し、胃のモデル $P(\text{胃の状態})$ を更新する(図4)。このように、我々の知覚(すなわち事後モデル)は得られた感覚をエビデンスとして事前モデルを更新することによって生成されているとするのが、ベイズ脳という考え方の主張であり、Fristonが想定した脳の働きであった。

7. 自由エネルギー原理に基づく ベイズ推定の脳内表現

しかしながら、同時に注意が必要なのは、実際に脳がベイズ推定式によって事後モデルを計算するのは困難だと考えられることである。なぜならば、事前モデル P (外界 θ) と尤度 P (感覚データ ϕ | 外界 θ) は経験的な事実に基づくため、脳内に近似的に表象されているのは想定できる (Buckley et al., 2017) 一方で、(1) のベイズ推定式の分母 $P(\phi)$ は外界から感覚データ ϕ が得られる真の確率であり、神のみぞ知るような値だからである。実際に、正確な計算を行うには計算量が膨大で、脳がそのような妙技を行えるとは考えられないことが指摘されている (Buckley et al., 2017)。

したがって、自由エネルギー原理では、脳は外界を推定するために、正確な事後モデルではなく、近似的な事後モデルである認識事後モデル $q(\theta)$ を算出しているとする。そして、脳は事後確率分布 $P(\theta | \phi)$ と認識事後確率分布 $q(\theta)$ を一致させるようにして働くとする。つまり、事後モデル-認識事後モデルをゼロにしようとする。事後モデルを直接求めるのは困難であるので、まずは認識事後モデルによってあたりをつけて、それから漸次修正して事後モデルに近似していくことによって、事後モデルを予想していくのである。

しかしながら、結局のところ事後モデルを計算できないのならば、「そもそも認識事後モデルを事後モデルに近づけることはできないのではないか」という疑問がある。ここで、「自由エネルギー」という概念が登場する。認識事後モデルを算出した際に外界について予測できなかった部分が「自由エネルギー」として脳には符号化されていると考えるのである (Buckley et al., 2017)。そして、「自由エネルギーを最小化させる」ことによって、この認識事後モデルを事後モデルに近似させることを試みているとする。

したがって、脳にとっては、自由エネルギーは予測の不確実性を表す情報量として経験されるのみであるとも言えるが、それを厳密に理解するならば、自由エネルギーは次のような情報量として定義できる。まず、事後モデルと認識事後モデルは確率分布であるため、その差はKLダイバージェンスという分布の距離のようなものから求め

られる。事後モデルと認識事後モデルの距離 $D_{KL}(q(\theta) \| p(\theta | \phi))$ は、定義によって次のように表される。

$$D_{KL}(q(\theta) \| p(\theta | \phi)) = \int d\theta q(\theta) \ln \frac{q(\theta)}{p(\theta | \phi)} \quad (5)$$

これを書き換えると、次のように表される。

$$D_{KL}(q(\theta) \| p(\theta | \phi)) = F + \ln p(\phi) \quad (6)$$

ここで、

$$F \equiv \int d\theta q(\theta) \ln \frac{q(\theta)}{p(\theta, \phi)} \quad (7)$$

である。この F が自由エネルギーと呼ばれるものである。正確には、この自由エネルギーは変分ベイズ法によって計算されるため、変分自由エネルギーと呼ばれている。

(6) 式における $\ln p(\phi)$ を一旦脇において、 F を小さくしていくと、事後モデル-認識事後モデルがゼロに近づくことがわかる。このようにして、脳は予測の不確実性 (自由エネルギー) を符号化しており、それを最小化するために機能しているとする、脳がベイズ脳的働きを持っていることは、現実的にあり得ることがわかる。

8. 自由エネルギー原理に基づく 脳機能の理解： 知覚的推論と能動的推論

脳が最小化を試みている自由エネルギーとは何であるか。実は、自由エネルギーは式変形を行うことで複数の意味を見いだすことができる (Friston, 2010)。ここでは (6) の式変形を行うことで、自由エネルギーの意味を明らかにする。

(6) において F を左辺に持ってくると、

$$F = D_{KL}(q(\theta) \| p(\theta | \phi)) - \ln p(\phi) \quad (8)$$

となる。 $-\ln p(\phi)$ はシャノンサプライズとよばれ、 $p(\phi)$ の値が大きくなるほど (つまり、確率が1に近づくほど)、値は小さくなる。つまりシャノンサプライズは、経験した感覚 ϕ の物珍しさを示す

情報量である⁴⁾。誤解を恐れずに簡略化するならば、自由エネルギーは次のように表現できる。

$$\begin{aligned} & \text{推論が外れている程度 (ダイバージェンス)} \\ & + \text{感覚の物珍しさ (シャノンサプライズ)} \end{aligned} \quad (9)$$

さて、この式から自由エネルギーを最小化することには2つの方向性があることがわかる。まず、推論の誤差 (ダイバージェンス) を最小化させるという方向性である。これは、認識事後モデルを事後モデルへと正確に近似させていく、すなわち外界について最善の推論を行うというプロセスである。自由エネルギー原理では、このようにして自由エネルギーを最小化させるためのプロセスが、「知覚」であると考えられている。

このプロセスをベイズ推定の考えと共に理解する (図4)。感覚入力信号を一定としたときに、図4のようなベイズ推定を繰り返すと、やがて事後モデルは、ほとんど正確に感覚信号を推定できるようになる。知覚はこのような事後モデルを、認識事後モデルによって近似的に推論するプロセスであると考えられている。したがって自由エネルギー原理では、「知覚」は知覚的推論 (Perceptual inference) と呼ばれている (Friston, 2010)。

一方で、予測誤差を最小化するには、(9) 式の感覚の物珍しさ (シャノンサプライズ) を小さくする方向性もある。これは、長時間座ったために痛みという「物珍しい感覚」が生じた時に、立ち上がるなどの運動を実行することで、痛みが無いという「なじみ深い感覚」を得ることなどが該当する。予測符号化モデルでは、運動はこのように外界を変化させることから、感覚信号の性質を変えるなどして、自由エネルギーを最小化させるためのものであると考えられている (Friston, 2010)。

その他にも、例えば血圧や心拍を一定にコントロールするなどのホメオスタシス制御も、感覚の物珍しさ (シャノンサプライズ) を小さくし、なじみ深い感覚信号を得るためのものであると言える。事前モデルは、これまで想定してきた、外界の推測という意味だけでなく、外界への期待あ

るいは望ましい状態を意味する場合がある (Seth, 2013)。後者の意味では、事前モデルはホメオスタシス制御の基準となる目標 (セットポイント) を意味する (Seth & Friston, 2016)。例えば、前述の血圧などはある程度の範囲に収まるように、その平均と分散が定められたモデルが事前モデル (予測・信念) として保持されている (e.g., 大平, 2019)。このように、事前モデルは「望ましい状態」を表象しており、事前モデルの分散はいわゆる生体の許容範囲を意味していると考えられる (大平, 2019)。したがって、実際の感覚入力はその範囲から逸脱する (例えば、心拍の増幅による血圧の上昇) と、なじみ深い感覚を再び得ようと、予測誤差を最小化し外界を制御 (例えば、心拍の鎮静化) しようとすることで、常に一定の身体状態が得られていると考えられる。

以上のように、運動や制御信号は、なじみ深い感覚刺激を得るために、環境 (身体という内的状態を含む) に働きかけることによって、自由エネルギーを最小化する方法である (Friston, 2010)。これは、自ら能動的に予測誤差を最小化させる方法であるため、能動的推論 (Active inference) と呼ばれる (Friston, 2010)。

まとめると、脳は予測通りではない現実に出会った時、2通りの戦略をとる。1つは、知覚的推論と呼ばれ、外界に沿って事後モデルを形成していく方法であり、もう1つは、能動的推論と呼ばれ、外界 (身体という内的環境を含む) に働きかけることによって、事前モデルにとってなじみ深い外界環境を実現する方法である。このように自由エネルギー原理は、知覚だけでなく行動も、自由エネルギーの最小化という目的の下で、予測符号化モデルによって統一的に説明できる可能性を示した。このことによって、予測符号化モデルは様々な脳機能を説明しうる統一理論として検討が進められるようになった。

ところで、本節では自由エネルギー最小化と予測誤差最小化にほとんど区別を設けず説明した。予測誤差は自由エネルギーを構成するものではあるが、実際には同義であるとは言えない。しかしながら、ここではその違いについては深くは立ち入らない (詳しくは、Buckley et al. (2017) を参照のこと)。本稿では、以後予測符号化モデルの適用例について概観するが、自由エネルギー最小化

4) $p(\phi)$ は確率値であり1以下であることから、 $\ln p(\phi)$ は負の値をとる。したがって、シャノンサプライズ $-\ln p(\phi)$ は正の値を取ることに注意。

と予測誤差最小化に触れる場合、これらはほとんど同じ意味として説明されている。

9. 予測符号化モデルと精神疾患

前述のように脳機能の創発を理解することによって、最近では知覚や行動の異常として定義される精神疾患の理解へ予測符号化モデルが適用され始めている。

例えば統合失調症は、予測符号化モデルから理解が最も進められている疾患の1つである。統合失調症の幻覚・幻聴症状は、事前モデルの精度が過度に高いために、事前モデルによる予測に対応するような知覚が得られていると考えられている(乾・阪口, 2020)。実際、Powers, Mathys, and Corlett (2017)において視覚刺激と聴覚刺激を対呈示するパブプロフ型条件づけ研究が行われたが、元々幻覚症状のある人は、聴覚刺激を取り除いた後でも、視覚刺激の呈示に対して聞こえるはずのない音が聞こえる条件づけ幻聴をより経験しやすいことが示されている。また、Cassidy et al. (2018)においても統合失調症患者の事前モデルの精度の高さが幻聴を予測することが同様に確かめられたが、この研究では結果を予測しにくい場合において、より幻聴症状がみられることが示されている。ただし幻聴については、事前モデルの精度が高いのではなく、内言の感覚入力としての精度が高すぎるために生じるとする知見も存在することには注意が必要である(e.g., Shergill et al., 2014; Stephan, Friston, & Frith, 2009)。しかしながら、妄想(Corlett et al., 2009; Corlett et al., 2010; Schmack et al., 2013)など幻聴以外にも様々な症状が予測符号化モデルから検討されており(Sterzer et al., 2018)、今後の予測符号化モデルに基づく検討によって、理論の精緻化が図られるとともに、様々な統合失調症の症状理解と治療方略の提案が得られるものと考えられる。

また、実験的なエビデンスは少ないものの、うつ病も予測符号化モデルによる理解が試みられている。例えばBarrett, Quigley, and Hamilton (2016)は、うつ病を内受容感覚の予測符号化が不正確な疾患として位置づけている。ここで内受容感覚とは、内臓、自律神経、内分泌系などの身体内部の信号に関する脳内処理と、その結果生じる感覚を

意味している(大平, 2018)。例えば、心臓がドキドキする感じなど、内臓由来の情動的要素を持った感覚はそれに該当する。内受容感覚は、古くから感情形成の基盤であることが指摘されており(e.g., James, 1890)、気分・感情の問題を持つうつ病患者に内受容感覚の予測の問題がある可能性は容易に推測される(Harshaw, 2015)。Barrett et al. (2016)は、うつ病患者においては、(1)事前モデルにおける異常(例えば、持続するネガティブ気分によって外界をネガティブに予測するモデルが形成されているなど)、(2)不正確な予測誤差信号(例えば、不正確なモデルによる不正確な予測信号によって、上行する予測誤差信号がバイアスされ疲労感を正確に知覚できないなど)、(3)予測の不正確性(例えば、報酬量を予測できないなど)という複数の異常を持つ可能性を考察している。そして、これらは能動的推論によってホメオスタシスを一定に保つことである「アロスタシス」(例えば、疲労を感じたら休むなど)の不全につながり、気分、運動、自律神経、免疫、代謝、概日リズムの異常といったうつ病の諸症状に影響を及ぼしていると考えている。実際にStephan et al. (2016)では、うつ病における慢性的な疲労様症状は、アロスタシスによってホメオスタシスを維持できないことから生み出される可能性を、予測符号化モデルの原理にしたがった数理シミュレーションによって示している。このシミュレーションによれば、ホメオスタシスの調整の失敗は、身体的な症状を引き起こすだけでなく、身体症状をアロスタシスによって調整できないという信念の形成、すなわちアロスタシスに対する自己効力感の低下を生み出すとされている。

なお、うつ病を能動的推論から整理する知見は、Barrett et al. (2016)以外にもみられる。例えば、Badcock et al. (2017)では、一般に能動的推論によって予測誤差を解消しない場合、知覚的推論によってモデルを更新することになると指摘する。その上で、例えばある社会集団が受容的ではないことを感知した際に、うつ病者は行動化(能動的推論)によって社会集団の自身への態度・行動を元の受容的なものへと戻すのではなく、社会集団への認識(事前モデル)をネガティブなもの(事後モデル)へと更新する(知覚的推論)とする。これは、その集団の非受容的な態度・行動に対す

る感度（すなわち、注意）を高める一方で、事前モデルの正確性を低くし、集団からの社会的報酬に対する予測を低下させる。そして、社会集団の受容的でない態度・行動に敏感になったうつ病者は、悲観主義、自己価値を低下させるなどによって、集団からの社会的報酬を得られる可能性が低いという信念を自己維持的に強める。そのことによって、うつ病者においては社会的報酬の探索という能動的推論がさらに行われなくなるとする。また、Kube et al. (2020) も同様に、うつ病者はポジティブな経験を得るなどして予測誤差が生じた時にも、経験をネガティブな方に意味づけることで事前モデルを維持するという「認知的免疫化 (Cognitive immunization)」戦略をとることから、うつ病の能動的推論において異常がみられる可能性を議論している。

その他にも、様々な精神疾患が予測符号化モデルから検討されている。例えば自閉スペクトラム症は、統合失調症と同様に、予測符号化モデルによる理解が比較的進められており、感覚刺激の精度が過度に高いという特徴によって整理がされている (Friston et al., 2014)。加えて、不安症状 (Cornwell et al., 2017)、慢性疼痛 (Hechler, Endres, & Thorwart, 2016; Song et al., 2021)、依存症 (Miller, Kiverstein, & Rietveld, 2020) など多くの疾患が、不正確な事前モデルによる不正確な事後モデルの生成という観点から予測符号化モデルによって検討されている。

このような知覚と行動の異常性の理解を背景に、最近ではそれらを改善するマインドフルネスを予測符号化モデルによって理解する試みがみられる。以下では、予測符号化モデルの観点から、マインドフルネスと内受容感覚、心身の健康の関連を論じた Farb et al. (2015)、MBCT の作用機序を論じた Manjaly and Iglesias (2020)、集中瞑想で行われていることを再解釈した Lutz et al. (2019) を主な論考として、著者らの解説も加えながら説明する。

10. マインドフルネスは知覚的推論が優位になる：Farb et al. (2015)

マインドフルネス介入では、呼吸に注意を向けることから始まり、身体のあらゆる感覚（内臓感

覚や筋肉の緊張などを含む）を観察する訓練をすることから、内受容感覚の精度を高める訓練をしているとも言える。Farb et al. (2015) は、内受容感覚の自己表象の正確性は、一瞬一瞬の適応を促進し、正確な表象からの乖離は機能不全をもたらすと述べた。実際に、先述したようにうつ病などの精神病理は、内受容感覚の異常や制御不全がその背景にあることが主張されている (Barrett et al., 2016; Stephan et al., 2016)。したがって Farb et al. (2015) は、Seth (2013) の提示した内受容感覚の予測符号化の枠組みに基づいて、マインドフルネスの作用機序を考察した。

Farb et al. (2015) は、マインドフルネスの作用機序の1つとして、現在の瞬間との繋がりの感覚であるプレゼンス (Presence) を想定している。プレゼンスは、個人の Well-being の感覚を決定するのに重要であるとされており、内受容あるいは外受容感覚の予測誤差信号がうまく最小化された場合に生じると考えられている (Seth, Suzuki, & Critchley, 2012)。これまでは、それは能動的推論によって成し遂げられると説明されてきたが (Seth et al., 2012)、Farb et al. (2015) は、瞑想的な観点からすれば、プレゼンスは知覚的推論によってもアクセスできると提案した。

ベイズ更新においては、事前モデルの精度が高く（分散が小さい）、感覚入力の精度が低い（分散が大きい）場合、あまり大きくモデルは更新されないのに対して、事前モデルの精度が低く（分散が大きい）、感覚入力の精度が高い（分散が小さい）場合、モデルは大きく更新される（図5）。マインドフルネス瞑想は、感覚入力に注意を向けて感覚入力の精度を高める。さらに、多様な内受容感覚を通じたプレゼンスは、内受容感覚の予測の分布を広げる、つまり事前モデルの精度を下げる可能性がある。これは、マインドフルネス瞑想によって高い精度の感覚入力を繰り返し経験して、移り変わる刺激に応じてモデルを更新していくと、モデルの分散が大きくなり、より迅速に知覚的推論によって予測誤差を最小化できるようになることを意味していると考えられる。したがって、素早く予測誤差を解消できるため、マインドフルネスにおいては能動的な調整が必要なような大きい予測誤差が生じにくくなるだろう、と考えられている (Farb et al., 2015)。能動的推論には、

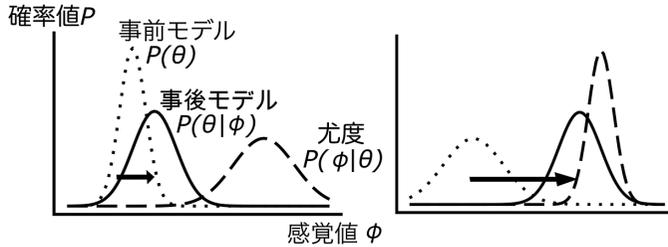


図5 事前モデルと尤度の精度の違いによるモデル更新の違い。事前モデルの精度が高い（事前モデルの分散が小さい）とモデルの更新の程度は小さいが、得られた感覚の精度が高い（尤度の分散が小さい）と得られた感覚にしたがって、モデルは大きく更新される。

うつや不安の維持要因とされている、ネガティブな内的体験を消そうとする体験の回避 (Hayes et al., 1999) 等の不適応的な対処も含まれていると考えられ、そのような能動的推論が減少することがうつや不安の低減に繋がるということは十分に考えられる。

このように、Farb et al. (2015) の論考は、マインドフルネスは知覚的推論での予測誤差の最小化を優位にすることで、それ自体が健康に繋がるプレゼンス感を高めるとともに、不適応に繋がりをよめるような能動的推論を少なくするという形で、マインドフルネスの作用メカニズムを説明していると言える。

11. 予測符号化モデルによるマインドフルネス認知療法 (MBCT) の説明：Manjaly and Iglesias (2020)

Farb et al. (2015) は、マインドフルネスによって、知覚的推論が優位になることの重要性を述べたが、Manjaly and Iglesias (2020) ではこれをさらに拡張して、うつ病の再発防止を目的としている MBCT (Segal et al., 2002) の Being モード、脱中心化、反応性 (Reactivity) の理解にまで適用した。

まず、MBCT の通常メカニズムの説明を行う。抑うつ処理活性仮説によると、うつ病経験者は、抑うつ感情が生じると、抑うつ的な情報処理が活性化しやすい (Teasdale, 1999)。このような認知的反応性が高い人ほど、うつ病の再発率が高いことが示されており (Segal, Gemar, & Williams, 1999)、MBCT はこの認知的反応性を低減することを目的としている。

高い認知的反応性は、Doing モードによって維

持される。Doing モードとは、望ましい状態と実際の体験の乖離 (Discrepancy) によって活性化される心的状態であり、現実を目標に近づけようとする認知や行動を駆動させる。外的な対象に対しては Doing モードは有効だが、内的な対象、すなわち自身の思考や感情などに対しては、うまく機能しない (e.g., 思考抑制の逆説的効果: Wegner et al., 1987)。ネガティブな感情に対して Doing モードが起動し、それを消そうとする考え (なぜこの感情がまた生じたのか? という原因追求等) や、望ましい状態 (ネガティブ感情なし) と実際の体験 (ネガティブ感情あり) の乖離が減少したかどうかをモニタリングすることは、反すうとして現れる。

これに対して、MBCT の中では、「どのような感覚が生じて、それを判断したり変化させたりしようとすることなく、許容し受け入れられる」心のモードである Being モードを訓練する。これはマインドフルネスの受容的な態度 (Shapiro et al., 2006) で観察することそのものであると考えられるが、この心のモードで、身体感覚だけでなく感情や思考といった体験が、変えようとせずとも過ぎ去っていくことを繰り返し経験することで、「これらの体験は、自分と同一ではなく、過ぎ去っていく心的事象である」という洞察が生じる。この洞察あるいはこの洞察のもとで体験を観ることが、メタ認知的気づきあるいは脱中心化と呼ばれ、認知的反応性の低減に繋がると考えられている。実際に、MBCT はこの脱中心化を高める (Bieling et al., 2012) とともに、うつ病の再発率を下げることを示されている (Teasdale et al., 2000)。

11.1 Being モード

Manjaly and Iglesias (2020) では、ベイズ推定の更新プロセスから、まず Being モードがどのような状態かを説明しようとしている。具体的には、ベイズ推定の更新プロセスは、以下のような計算式で記述できることがわかっている (乾・阪口, 2020)。

$$\begin{aligned} \text{予測の修正量} = & \\ & \frac{\text{感覚信号の精度}}{\text{予測の精度} + \text{感覚信号の精度}} \\ & \times (\text{予測誤差}) \quad (10) \end{aligned}$$

予測誤差は精度で重みづけられており、感覚信号の精度が高ければ高いほど (値が大きければ大きいほど)、予測の修正量 (モデル更新の期待値) は、予測誤差と近い値になっていく。すなわち、感覚信号の精度が高いほど、得られた感覚信号と同期するようにして、モデルは修正されていく (図5)。

それでは、感覚信号の精度はどのようにして高められるかという点、予測符号化モデルでは注意がその役割を果たすと考えられている (Feldman & Friston, 2010)。神経生理学的には、注意によって引き起こされる感覚信号の精度の変化は、前脳基底部や脳幹からの神経性調節 (特にコリン系) が変化し、ニューロンの興奮性が変化することに依存していると考えられている (Friston, 2009)。そしてこの興奮性の変化は、ニューロンの活動のばらつきを統制し、よりノイズの少ない (つまり、分散の小さい) 感覚信号の入力を可能にしていると考えられている。

実際、MBCTにおいては、まず最初に感覚に注意を維持しつづけることが強調される。これは感覚入力 of 精度を高める操作であると考えられ、このために変化 (予測誤差) に対して柔軟にモデルを更新できるようになり、どのような感覚も受け入れることができるような Being モードが達成されるのだと考えられる。

11.2 脱中心化

Manjaly and Iglesias (2020) は、Being モードによって、感覚信号の変化に応じて事前モデル (信念) が刻々と変化していくと、最終的に信念が緩

む (事前モデルの分散が大きくなる) と考えている。さらに、様々な階層においてモデルが急速に変化することを考えると、非常に高いレベルの階層において形成されていた「自己」という信念でさえも、変化するものであるという自覚に至ることを想定している。自己という信念でさえも変化することを自覚することによって、以前は重要視され、自分のアイデンティティに不可欠なものとして理解されていた、自分自身の主体性や制御に関する高次の信念への執着が緩む可能性がある。Manjaly and Iglesias (2020) は、この状態が「思考や知覚をただ単純に心に生じては消えていく現象として経験する能力」と定義される脱中心化なのではないかと考えている。

このように Manjaly and Iglesias (2020) において脱中心化は、知覚などを行う低次の階層を制御するより高次の階層、すなわちメタ認知レベルで起こると想定されているが、これは脱中心化を「メタ的気づき」と理解する MBCT とも共通するものである (Bernstein et al., 2015; Teasdale, 1999)。また、予測符号化モデルから検討することで、脱中心化は Being モードの直接的な帰結である可能性が示されたと言える。

11.3 反応性

最後に、予測符号化モデルは、認知的・生理的な反応性を新たに理解する方法としても役立つとされる。

一般的に、事前モデルの精度が高いほど、予測誤差が生じた際に行動が活性化されやすくなる (Stephan et al., 2016)。例えばアロスタシスによる身体調整では、事前モデルの精度が高い (ホメオスタシス信念が強固な) ほど、身体からの感覚入力が期待から逸脱する可能性が高く、制御が反射的に展開されることが示されている (Stephan et al., 2016)。このような前提に立つと、MBCT を実施した結果、認知的反応性が低くなることは、事前モデルの精度が低下した (信念が緩んだ) ためであると考えられる。事前モデルの分散が大きければ、外界状態の想定範囲も広くなるため、事前モデルの平均 (予想) から外れた感覚も想定範囲内になり、入力された感覚に対して反射的な反応が低下することが期待される (Stephan et al., 2016) (図6)。これは、生理学的には、例え

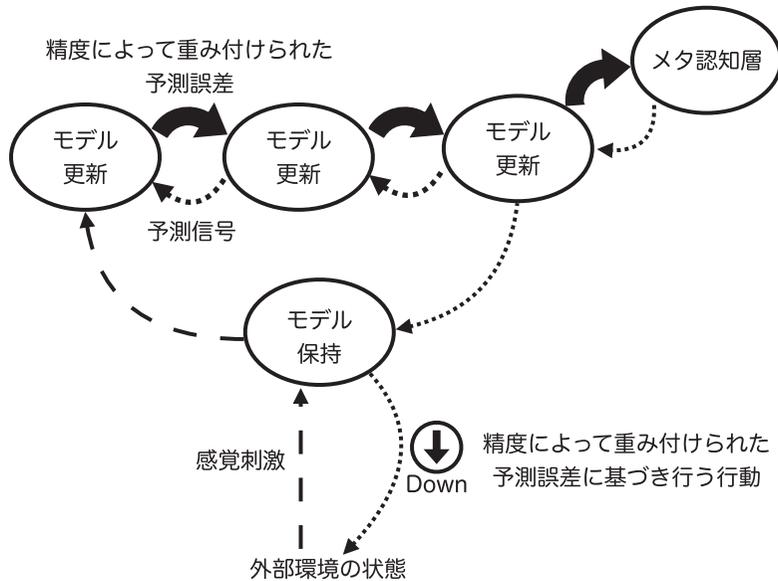


図6 Manjaly and Iglesias (2020) に基づくマインドフルネスの理解 (Manjaly and Iglesias (2020) の図6, 7, 8に基づき作成)。感覚刺激へ注意を向け、感覚刺激の入力精度が高まることによって、より重みづけられた予測誤差信号が高次の階層に送られる。したがって、得られた感覚刺激に応じて、事後モデルは迅速に更新されるようになる (Being モード)。予測誤差がより高次の階層に伝搬されると、やがて低次の階層を統制しているメタ認知層における自己の主体性や制御に関する信念の精度が低下する (脱中心化)。反応を出力する効果器においては通常、期待される感覚信号のモデル (信念) を保持できるように、感覚信号の方を変える反応が出力されるが、期待される感覚信号の信念精度が低下することで、反応が出力されにくくなると考えられる (認知的・生理的反応性の低減)。

ば前帯状回からの下降投射によって視床下部や脳幹における局所的な興奮-抑制のバランスを変化させたり、ホメオスタシス信念を符号化するニューロン活動のばらつきを変化させたりすることが対応する (Seth & Friston, 2016)。このようにして事前モデルの精度を低いものにすると、反射的な反応が減少し、感覚変化に対する過剰な反応性が弱まると考えられる。

このMBCTによる事前モデルの精度の低下は、Being モードの醸成と脱中心化の達成の結果であると考えられる。具体的には、予測誤差に応じて事前モデルが刻々と変化する状態である Being モードによって、より高次のレベルの信念 (自己など) でさえも変化するものであるというメタ認知的気づき (図6) を経た結果、事前モデルの精度が低下して反応性が低下すると考えられる。言い換えると、知覚階層 (最も低次の階層) において常に信念が更新されることを経験し、多少の予測誤差を「許容できる」ようにメタ認知的な期待を調整することによって、事前モデルを保持しようとする傾向が弱くなることが考えられる。こ

れは、高次の階層から効果器へのトップダウン信号が弱くなる (精度が低くなる) ことに繋がる可能性があり、感覚刺激にフォーカスするMBCTによって、刺激、感情、思考に対してまで反射的な反応が減少することのひとつの説明となると考えられる (図6)。

このManjaly and Iglesias (2020) の説明は、マインドフルネスにより感覚入力精度が高まり、事前モデルの精度が弱まるというFarb et al. (2015) の説明と大きくは一致しているが、階層的な処理のモデルを明示し、MBCTで変化する心理的概念がどの処理の変化に相当しているかを詳述した点で有意義であったと言える。

12. 予測符号化モデルによる集中瞑想の説明: Lutz et al. (2019)

ここまでの2つの論文では、マインドフルネスの訓練によって、全体的にどのような変化がみられるかを説明していると言えるが、Lutz et al. (2019) は、集中瞑想を行なっているときにどの

ようなことが起きているかを、予測符号化モデルで理解しようと試みている。マインドフルネス瞑想は、特定の1つの対象（呼吸の感覚等）に集中する集中瞑想と、特定の1つの対象ではなく意識にのぼるあらゆる体験を評価判断せずに気づいていく観察瞑想の2つで構成されている（Lutz et al., 2008）。この中でも Lutz et al. (2019) では、特にプロセスがシンプルにモデル化されている、集中瞑想に着目している。

集中瞑想では、練習者は背筋を伸ばして安定した座位をとる（Phase 1a）。続いて、注意を、局所的な呼吸の感覚など、選択した対象に意図的に向け（Phase 1b）、注意の質と気ざらしの発生を継続的に監視することで維持する（Phase 1c）。それにもかかわらず、遅かれ早かれ、注意は意図したターゲットから外れてしまう（Phase 2）。そのような出来事に対処するための典型的な指示は、それを認識し（Phase 3a）、注意を気の散る内容から「外す」（Phase 3b）、そして注意を選択した瞑想の対象に戻す（Phase 4）ことである。Lutz et al. (2019) では、このような集中瞑想のプロセスの中で実践者は「行動しない（Non-action）」という行為を主体的に選択することによって、実利的価値（Pragmatic value）と認識的価値（Epistemic value）を追求していると説明されている。

なお、Lutz et al. (2019) の論文内では明示的に記載されていないが、実利的価値と認識的価値は、将来の成果の不確実性を表す「期待自由エネルギー」を構成する2つの価値に相当すると考えられる。ここで実利的価値は、ある行動規則をとることによって、「どれだけ成果（事後分布）を期待できるか」の期待値を示し、一方で認識的価値は、「どれだけモデルの更新（事後モデルと事前モデルのダイバージェンス）が期待できるか」の期待値を示すとされる（e.g., Parr & Friston, 2017）。乾・阪口（2020）の説明によると、期待自由エネルギーは以下のように解釈することができ、人間はこれら2つの価値の和が最大となるような行為系列を選択することで、期待自由エネルギーを最小化している。

$$\begin{aligned} (\text{期待自由エネルギー}) = \\ - (\text{認識的価値}) - (\text{実利的価値}) \quad (11) \end{aligned}$$

Lutz et al. (2019) で言及されている価値は、この定義を汲みながらもより広く定義されたものであるが、実利的価値は、注意を規律正しく調整することで心を落ちつけるというものであり、認識的価値は、精神機能の構築的・動的な性質について親しむこと（例えば、思考の無常を認識する）であると説明されている。

12.1 集中瞑想の実利的価値

集中瞑想は、姿勢を正し呼吸に意識を向けるといった注意集中のための行動規則と、対照的に思考にしたがって意識をあちらこちらへと移ろわせるといった、単純な行動規則のみから成り立っている。集中瞑想は、このように単純化された状況において行動規則の選択を調整することで、心の統制を行う訓練をしている。Lutz et al. (2019) では、この注意を規律正しく調整することで心を落ちつけるという要素が実用的な価値であるとしている。そして、この行動調整に大きな役割を担うのが、「行動しない」ことであるとしている。

具体的には、注意とさまよいという相容れない行動規則が存在することから、以下のような弁証法的プロセスが行われていると説明されている。まず Phase 1（テーゼ）では、安定した姿勢で座り所定のターゲットに注意が向けられる。このフェーズでは、すぐに感覚に関する予測誤差が減少する。そして Phase 2（アンチテーゼ）では、気を散らす思考や感覚に注意が奪われ、それに伴い身体面にも変化が現れ始め（例えば、不安によって自律神経が覚醒し、筋肉が緊張するなど）、感覚に関する予測誤差が生じ始める。そして、Phase 3（ジンテーゼ）において、注意が拡散し、知らず知らずのうちに心がさまよっていたことに気づく。このフェーズは、予測誤差がピークに達した段階である。しかし、集中瞑想においては、「行動しない」ために、身体的な活動によってこの予測誤差を調整することはできない。よって、事前モデルをわずかに修正することで、予測誤差を調整することになる。このようにして、予測誤差を減少させることによって、注意を再び戻すことができるが、「行動しない」は、気を反らせるものを手放し取り除くということに重要な役割を果たしており、ある意味で Phase 4（注意を瞑想対象に戻す）の一側面であるとも考えられる。

このような弁証法的プロセスをたどることによって、注意とさまよいという矛盾した行動規則は両立するものとして存在させることができる(アウフヘーベン)。実際に、熟練した瞑想者においては、自発的な精神活動の発生は、必ずしも注意集中を妨げるものではない(Lutz et al., 2019)。注意ネットワークの活動とデフォルト・モード・ネットワーク(Default mode network: DMN)の活動とで負の相関がみられない人は、注意を維持しやすいという知見(Pagnoni, 2012)は、予期できないものの関連性のある刺激を検出する能力(Corbetta, Patel, & Shulman, 2008)が、DMNを神経基盤と言われるマインドワンダリング(目の前の課題から注意が逸れて無関係なことを考えること)からの自律性と関連するという見解を支持するものである。この矛盾した行動規則の両立は、以下に説明するような集中瞑想における認知的価値の獲得につながる。

12.2 集中瞑想の認知的価値

Lutz et al. (2019)では、集中瞑想における認知的価値とは、例えば自分の思考の無常を認識することであると説明されている⁵⁾。

一般に、集中した状態(Phase 1)からマインドワンダリング(Phase 2)に至る状態は、予測される感覚状態および予測誤差精度の両方の変化を伴うものとみなすことができる。まずPhase 1では、姿勢と呼吸への注意を維持しながら、「行動しない」ことをあえて選択しているようにみえる。これは、予測符号化の観点から検討すると、身体状態や注意対象に関する予測誤差を最小化させようとしていると考えられる。ボトムアップの予測誤差とトップダウンの予測の影響は常に競合関係にあるとされており、注意を向けてボトムアップの予測誤差信号に重みづけることは、相対的に事前モデルの精度を低めることに寄与すると考えられる。これはFarb et al. (2015)やManjaly and Iglesias (2020)の主張と共通している。このようにして予測誤差を最小化させることによって、行動せず

とも予測誤差を最小化できることを学んでいき、予測誤差に対して「行動しない」という行動規則はさらに促進される。

しかしながら、このような予測誤差の迅速な解消が、Phase 2の契機ともなる。瞑想中は外部環境の変化が少ないこともあり、珍しい感覚に出会うことはそうない。したがって、比較的低次の階層で処理される感覚刺激を容易に処理できるようになるため、DMNの活動に関連する自伝的な情報に関連するような、いま現在予測誤差を解消できていない対象へと情報処理の階層が移行すると考えられる。この移行によって、瞑想者は自分の身体と周囲の環境の両方を忘れるのだと考えられる。

しかしながら、再び感覚刺激に関する予測誤差を十分に自覚するほどの感覚入力の変化が生じる(例えば、座位の姿勢が崩れる等)。この時、瞑想者は無意識のうちに注意がそれていたことに気づき、マインドワンダリングを中断させる(Phase 3)。このようにマインドワンダリングの生起自体は、自発的な制御の対象外であるが、その持続時間については、集中瞑想の行動規則によって、短縮できると考えられる。実際に、いくつかの研究では瞑想者のマインドワンダリングとDMN活動の両方の制御能力が向上していることが示されている(Brewer et al., 2010; Mrazek et al., 2013; Pagnoni, 2012)。

このようにして迅速にフェーズ移行が繰り返されると、注意集中とマインドワンダリングとの間で絶妙な動的バランスを保つことができるようになる。精神的な出来事は、その内容には意識的にアクセスできるが、その構成のプロセスにはアクセスできない場合、「透明(Transparent)」であると言われる(Metzinger, 2003)。透明性は、精神的な出来事に直接「触れている」という感覚を作り出し、したがってその「現実」に対する我々の主観的な確信と結びついているとされる(Lebois et al., 2015)。瞑想中に様々な精神的な出来事に触れ、手放すことで、それらが仮設的で構築されたものであり、自律性の範囲外にあることに気づき、究極的にはそれらが「無常」であることを理解できるようになる(Lutz et al., 2019)。Lutz et al. (2019)では、このようにして、集中瞑想は認知的価値を形作ると考えられている。

5) 仏教においては、止瞑想(集中瞑想)ではなく、観瞑想(観察瞑想)が無常・苦・無我の理解を目的としているとされているが、Lutzらは主に注意のあり方や体験への態度の違いによって、2つの瞑想を分類していることに注意が必要である(e.g., Lutz et al., 2008)。

以上のように、Lutz et al. (2019) においては、集中瞑想は弁証法的プロセスによって成り立っているとされ、予測符号化モデルの観点から実利的価値と認知的価値を追求するプロセスとして理解することが試みられている。まとめると、まず Phase 1 (テーゼ) においては、注意を対象に向けることによって、感覚処理という低次な階層における予測誤差が減少する。一方で、低次な階層における予測誤差が最小化されることによって、予測誤差がまだ解消されていない高次な階層の処理へと注意が移る。これが Phase 2 (アンチテーゼ) における、注意対象と外部環境に関して注意を喪失した状態であると考えられる。これらの両立しない状況は、注意を集中し予測誤差に気づきやすくなることで、フェーズの移行を素早いものとし、集中とマインドワンダリング間において絶妙なバランスを保つことによって解消される (Phase 3: ジンテーゼ)。この素早いフェーズ移行によって、心の統制を保つことが容易になる (実利的価値) ことに加えて、集中とマインドワンダリング間のバランスによって、思考のありのままの性質、すなわち予期できない精神的事象の変遷に触れることが可能となることから、思考の無常性という理解が培われる (認知的価値) のだと考えられる。

13. 予測符号化モデルから見た マインドフルネス：まとめ

これまで説明してきた論考は、以下の点で共通点が見られる。(1) マインドフルネスでは、能動的推論よりも、知覚的推論による予測誤差の最小化が主に行われている。その知覚的推論の際には、(2) 感覚入力に注意を向けることで、ボトムアップな入力に重みづけをしている (精度が高い)。(3) 事前モデルには重みづけをしない (精度が低い)。この (2) と (3) の結果として、モデルは迅速に更新され、予測誤差は迅速に最小化される。

Farb et al. (2015) や Manjaly and Iglesias (2020) では、刻一刻と変わる感覚信号に合わせて、上記の仕方モデルを更新していくことで、最終的にはより事前モデルの精度が低くなっていく (事前分布の分散が大きくなる) ことが想定されてい

る。しかし、Manjaly and Iglesias (2020) 論文に対する反論論文が発表されている (Verdonk & Trousselard, 2021)。この論文では、(1) もしモデルを感覚入力に重みづけて更新するならば、事前モデルは尤度と近似するように正確なものになる、(2) マインドフルネスは、瞬間瞬間に注意を向けることから、環境変化に応じて適切なモデルを抽出することができ、むしろモデルは正確になって事前モデルの精度も高まり (分散が小さい)、最終的に適応的な能動的推論ができるようになることがマインドフルネスのメカニズムとしてあるのではないかと考察されている (Verdonk & Trousselard, 2021)。

これらの意見の相違は、モデルを構成する対象の違いによって説明できると考えられる。内受容感覚のような、常にある程度の変動が微細には存在するような入力信号に対しては、それを精度高く観察していれば、実際の内受容感覚の揺らぎに合わせて、分散の広がった事前分布が形成されると考えられる。一方で、あまり変動がなく、明確に定まっている外的対象や事象に対しては、その実際に合わせて、分散の小さい事前モデルが形成されると考えられる。これらの考察を踏まえると、マインドフルネスは高い学習率によって、より正確な世界の理解に繋がる、と考えることができる。

14. 予測符号化モデルによる マインドフルネスの再定義

ここまでで、予測符号化モデルの観点からマインドフルネスのメカニズムを考察した論文を説明してきた。これらの理論的な検討は近年始まったばかりで、その実証的な研究はほとんど見られない。著者らは、今後の実証的な研究のために、Kabat-Zinn (2003) のマインドフルネスの定義「意図的に、瞬間瞬間の体験に対して、評価判断することなく、注意を向けることによって得られる気づき」を、予測符号化モデルの観点から以下のように再定義する (図7)。感覚入力の精度が高い (分散が小さい) ことを、マインドフルネスにおける「注意」とし、事前分布の精度が低い (分散が大きい) ことを、マインドフルネスにおける「評価判断しない」とする。これらによって得ら

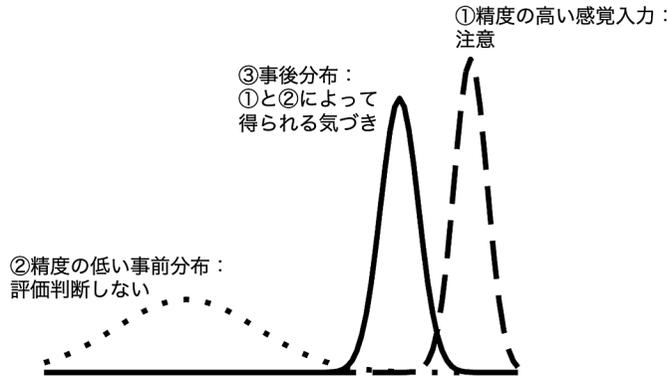


図7 予測符号化モデルの観点からのマインドフルネスの再定義。Kabat-Zinn (2003) の定義「意図的に、瞬間瞬間の体験に対して、評価判断することなく、注意を向けることによって得られる気づき」に基づく。過去の経験から形成された予測（事前分布）に重みづけず（精度を低め）、現在の感覚入力に重みづける（精度を高める）ことによって、柔軟なモデル更新を可能にしている。

れる気づきが、事後分布が生成された時の知覚あるいは認識に相当すると考えられる。マインドフルネス訓練における、体験を（事前の予測に従って）解釈したり決めつけたりしない、初めて体験するかのように経験するなどの教示は、事前分布への重みづけを下げることを誘導していると考えられる。このように、意図的に事前モデルの精度を下げて「評価判断しない」こともできる一方で、Farb et al. (2015) や Manjaly and Iglesias (2020) が述べたように、感覚入力の変化に伴って、高い精度の感覚入力で分布を大きく更新していった結果、事前分布の精度が低くなり、「評価判断しない」は達成されるものでもある。マインドフルネスの注意と受容的な態度は相互に影響し合うというモデルはいくつか発表されている (Shapiro et al., 2006 ; Teper, Segal, & Inzlicht, 2013) が、著者らの再定義ではこのように注意が受容に繋がる過程を理解することができる。またこのように予測符号化の観点から再定義することで、マインドフルネスという心の状態は、特別なものではなく、通常の認識の仕方の一形態であると考えることができ、基礎研究との接続や今後の実証研究の足がかりになることが期待される。

15. 仏教の真理との関連

マインドフルネスの起源である仏教の実践は、無常・苦・無我という真理を体得することを目的としている。

あらゆる現象は常に変化しているという無常の真理に関して、予測符号化の観点から捉えると、外界の現象に対する無常の理解に加えて、観察する主体自体も無常であることを表現することができる。もし事前分布の精度が高い、つまりこれまでの経験に固執するならば、分布はあまり更新されず、常に同じような感覚入力があると体験されるであろう。一方で、事前分布の精度が低く、感覚入力の精度が高いならば、実際の感覚の変化を検出することができる。低次の階層において、刻一刻と変わる感覚の変化に気づく体験を繰り返すことは、より高次の階層における環境の変動性についての信念に影響を与え、「世界は無常である」という信念を形成すると考えられる。また、ある感覚刺激を知覚した後は、事前分布が更新されているため、完全に同じ感覚刺激が再び入力されたとしても、事後分布が異なり、また違った刺激として知覚される。このように、仏教における無常というものは、観察される対象自体も無常であることに加え、観察する主体自体も無常であることを意味していると考えられる。

苦の真理は、あらゆる現象は苦しみである、というものである。苦は仏典が書かれたパーリ語では *dukkha* と呼ばれ、これは不満足とも訳される (魚川, 2015)。我々は世界に関する完全なモデルを構築するという終わりのないプロセスの中で生きることを余儀なくされており、無常が真であるならば、予測誤差はゼロになることはないという意味で、常に不満足している存在であると言える

のではないだろうか。

無我は、完全に制御できる固定的な自己という実体はないことを意味している(魚川, 2015)。仏教の3つの真理は1つの真理を違う側面から表現しているものとされており、苦に関して述べた、外界(自身の身体を含む)とモデルの誤差がゼロになることはないという意味で、完全に制御できるものは存在せず、無常に関して述べた、観察する主体も無常であるということから、固定的な自己は存在していないと言える。また、Manjaly and Iglesias (2020) が述べた、高次の階層には自己の主体性や制御に関する信念が存在し、低次の階層からの予測誤差信号が重みづけられて入力され続けることで、自己の主体性や制御に関する信念も緩むという脱中心化の説明も、予測符号化の観点から無我の理解を説明していると言えるだろう。

16. 今後の展望・課題

未だ実証研究がほとんどなされていない中で、今後の展望・課題を述べる。まず感覚信号の精度を高める注意に関する研究は、マインドフルネスの文脈とは別に多くの研究がなされている一方で、内受容感覚の精度を測定する課題や指標に関しては、統一的な見解が得られていない。多く使われてきた課題として、心拍知覚課題があるが、マインドフルネス介入によって心拍知覚課題成績が向上することは、わずかな例(Bornemann et al., 2015)を除き頑健に示されていない(Melloni et al., 2013; Parkin et al., 2014)。一方で、マインドフルネス瞑想による島の活性化は一貫して示されており(Fox et al., 2016)、前部島と後部島で信号をやりとりすることで、モデルを更新するということを盛んに行っていると考えられ、Farb et al. (2015) や Manjaly and Iglesias (2020) の論考と一貫している。今後は、心拍以外を含めた内受容感覚の精度を測定する課題の精緻化によって、このような脳の指標と一貫するような結果が得られるかどうかを検討する必要がある。

注意に加えて、評価判断しないことによって、モデルが大きく更新される点に関連する研究としては、マインドフルネス実践者は報酬予測に反する結果が与えられたときの線条体の活動の変化

が、非実践者に比べて小さいことが示されている(Kirk & Montague, 2015)。この結果は、マインドフルネス実践者は報酬予測誤差に対して内的モデルを更新する学習率が低くなっているという解釈(大平, 2016)がある一方で、マインドフルネス実践者は感覚入力に重みづけていて、すぐに予測誤差が最小化される、つまり学習率が高いために、複数回の試行全体で見ると報酬予測誤差信号が小さくなっているという解釈も考えられる。今後は、時系列的な反応データに対して、カルマンフィルターや階層ガウシアンフィルターなどの計算論モデルを用いた検討を行うことによって、事前分布と感覚入力のそれぞれの分布のパラメータを推定し、著者らが提案したマインドフルネスの再定義が妥当かどうかを検討する必要がある。

さらに今後は、脳内の処理の具体的にどの階層で、事前分布の精度が下がる、ということが生じているのかを検討できることが望ましい。高次の階層で事前分布が広がることは、例えば、「これが正しい考えだ」という内的モデル・信念が緩み、許容範囲が広がる(どのような思考があっても良い)状態を表現している可能性がある。これは、どのような体験も許容するセルフ・コンパッションや平静さ(Equanimity: 藤野論文参照)の状態に近いと言えるだろう。こういった点に関しては、fMRIのEffective connectivity解析などによって高次の階層(腹内側前頭前野など)から低次の階層への影響性の強さが、マインドフルネス介入によってどう変化するか、などによって検討できる可能性がある(Manjaly & Iglesias, 2020)。

また、マインドフルネスによる自己感の変容(Hölzel et al., 2011)を検討するためには、自己というものが高次の階層でどのように表現されているかも考える必要がある。例えばうつ病者において、身体のだるさの入力が継続的に与えられていて、精度の高い強固な事前分布が形成されていたならば、そのだるさの変化に気づくこともできず、高次の階層でも「私はうつ病である」という信念が確立されていると考えられる。この事前分布に近づけるように知覚的推論や能動的推論を行うことが、ネガティブな記憶、認知、注意バイアス(Disner et al., 2011)などの、あたかもうつ病であるという現状を維持しようとしているように見える認知や行動に対応している可能性がある。だ

るさがない状態をセットポイントとした、ホメオスタシス制御のためのだるさの事前分布も存在するはずだが、うつ病の期間が長引くほど、うつ病である自分が通常という予測の分布が強固に確立されてしまい、その予測を支持するような感覚を集めることで、うつ病が維持されている可能性が考えられる（この理解は、うつ病が再発するほど再発率が高くなるという現象も説明するかもしれない）。今後は、それぞれの階層での事前分布の精度を推定することができるような課題や測定法を開発することができれば、マインドフルネスがどのようなプロセスを経て、うつ病などの自己に関する信念の異常も見られる病態を治癒するのかを詳細に明らかにできると考えられる。

最後に、本稿で述べた予測符号化モデルから見たマインドフルネスの理解では、Shapiro et al. (2006) のマインドフルネスの構成要素の「意図」、すなわち、なぜ（何を指して、あるいは何が原因となって）マインドフルネスをするのかは説明できていない。また Shapiro et al. (2006) は、マインドフルネスが実現すると、脱中心化に加えて、(1) 自己制御、(2) 価値の明確化、(3) 認知・情動・行動の柔軟性、(4) エクスプロージャーの4つの付加的なメカニズムによって効果に繋がると説明している。この4つのメカニズムは、大まかにまとめると「目の前の刺激や内的体験の影響を受けすぎずに、自分のやりたいこと、やるべきことができるようになる」と表現できるだろう。しかしここで、「自分のやりたいこと、やるべきこと」という事前の期待や信念に重みづけて、環境（身体を含む）に左右されにくくなる、というのはこれまで述べてきたマインドフルネス理解と反するようにも見える。MBCTにおいても、あくまで Doing モードと Being モードのギアチェンジの練習をしているという説明がなされている (Segal et al., 2002) が、このような能動的推論と知覚的推論の2つの戦略のバランスを取ることを、どのようにマインドフルネスが実現しているのかを明らかにすることも、今後の検討課題であると言えるだろう。

謝辞

本稿を執筆するにあたり、有益なご示唆をくださったトロント大学准教授 Norman Farb 先生に記して感謝いたします。

文献

- Allen, M., Dietz, M., Blair, K. S., van Beek, M., Rees, G., Vestergaard-Poulsen, P., ... Roepstorff, A. (2012). Cognitive-affective neural plasticity following active-controlled mindfulness intervention. *Journal of Neuroscience*, 32, 15601–15610.
- Alsubaie, M., Abbott, R., Dunn, B., Dickens, C., Keil, T. F., Henley, W., & Kuyken, W. (2017). Mechanisms of action in mindfulness-based cognitive therapy (MBCT) and mindfulness-based stress reduction (MBSR) in people with physical and/or psychological conditions: A systematic review. *Clinical Psychology Review*, 55, 74–91.
- Anderson, J. S., Lampl, I., Gillespie, D. C., & Ferster, D. (2000). The contribution of noise to contrast invariance of orientation tuning in cat visual cortex. *Science*, 290, 1968–1972.
- Badcock, P. B., Davey, C. G., Whittle, S., Allen, N. B., & Friston, K. J. (2017). The depressed brain: An evolutionary systems theory. *Trends in Cognitive Sciences*, 21, 182–194.
- Baer, R. A., Carmody, J., & Hunsinger, M. (2012). Weekly change in mindfulness and perceived stress in a mindfulness-based stress reduction program. *Journal of Clinical Psychology*, 68, 755–765.
- Bear, M., Connors, B., & Paradiso, M. A. (2020). *Neuroscience: Exploring the brain*. Burlington, MA: Jones & Bartlett Learning.
- Baer, R. A., Smith, G. T., Hopkins, J., Krietemeyer, J., & Toney, L. (2006). Using self-report assessment methods to explore facets of mindfulness. *Assessment*, 13, 27–45.
- Barrett, L. F., Quigley, K. S., & Hamilton, P. (2016). An active inference theory of allostasis and interoception in depression. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371, 20160011.
- Bernstein, A., Hadash, Y., Lichtash, Y., Tanay, G., Shepherd, K., & Fresco, D. M. (2015). Decentering and related constructs: A critical review and metacognitive processes model. *Perspectives on Psychological Science*, 10, 599–617.
- Bieling, P. J., Hawley, L. L., Bloch, R. T., Corcoran, K. M., Levitan, R. D., Young, L. T., ... Segal, Z. V. (2012). Treatment-specific changes in decentering following mindfulness-based cognitive therapy versus antidepressant medication or placebo for prevention of depressive relapse. *Journal of Consulting and Clinical Psychology*, 80, 365–372.
- Bishop, S. R., Lau, M., Shapiro, S., Carlson, L., Anderson, N. D., Carmody, J., ... Devins, G. (2004). Mindfulness: A proposed operational definition. *Clinical Psychology: Science and Practice*, 11, 230–241.

- Bornemann, B., Herbert, B. M., Mehling, W. E., & Singer, T. (2015). Differential changes in self-reported aspects of interoceptive awareness through 3 months of contemplative training. *Frontiers in Psychology, 5*, 1504.
- Brewer, J. A., Bowen, S., Smith, J. T., Marlatt, G. A., & Potenza, M. N. (2010). Mindfulness-based treatments for co-occurring depression and substance use disorders: What can we learn from the brain?. *Addiction, 105*, 1698–1706.
- Brewer, J. A., Worhunsky, P. D., Gray, J. R., Tang, Y. Y., Weber, J., & Kober, H. (2011). Meditation experience is associated with differences in default mode network activity and connectivity. *Proceedings of the National Academy of Sciences, 108*, 20254–20259.
- Brown, K. W., & Ryan, R. M. (2003). The benefits of being present: Mindfulness and its role in psychological well-being. *Journal of Personality and Social Psychology, 84*, 822–848.
- 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, 81*, 55–79.
- Cassidy, C. M., Balsam, P. D., Weinstein, J. J., Rosengard, R. J., Slifstein, M., Daw, N. D., ... Horga, G. (2018). A perceptual inference mechanism for hallucinations linked to striatal dopamine. *Current Biology, 28*, 503–514.
- Corbetta, M., Patel, G., & Shulman, G. L. (2008). The reorienting system of the human brain: From environment to theory of mind. *Neuron, 58*, 306–324.
- Corlett, P. R., Krystal, J. H., Taylor, J. R., & Fletcher, P. C. (2009). Why do delusions persist? *Frontiers in Human Neuroscience, 3*, 12.
- Corlett, P. R., Taylor, J. R., Wang, X. J., Fletcher, P. C., & Krystal, J. H. (2010). Toward a neurobiology of delusions. *Progress in Neurobiology, 92*, 345–369.
- Cornwell, B. R., Garrido, M. I., Overstreet, C., Pine, D. S., & Grillon, C. (2017). The unpredictable brain under threat: A neurocomputational account of anxious hypervigilance. *Biological Psychiatry, 82*, 447–454.
- Desbordes, G., Gard, T., Hoge, E. A., Hölzel, B. K., Kerr, C., Lazar, S. W., ... Vago, D. R. (2015). Moving beyond mindfulness: Defining equanimity as an outcome measure in meditation and contemplative research. *Mindfulness, 6*, 356–372.
- Disner, S. G., Beevers, C. G., Haigh, E. A., & Beck, A. T. (2011). Neural mechanisms of the cognitive model of depression. *Nature Reviews Neuroscience, 12*, 467–477.
- Ehring, T., Zetsche, U., Weidacker, K., Wahl, K., Schönfeld, S., & Ehlers, A. (2011). The Perseverative Thinking Questionnaire (PTQ): Validation of a content-independent measure of repetitive negative thinking. *Journal of Behavior Therapy and Experimental Psychiatry, 42*, 225–232.
- Farb, N. A., Daubenmier, J., Price, C. J., Gard, T., Kerr, C., Dunn, B. D., ... Mehling, W. E. (2015). Interoception, contemplative practice, and health. *Frontiers in Psychology, 6*, 763.
- Farb, N. A., Segal, Z. V., & Anderson, A. K. (2013). Mindfulness meditation training alters cortical representations of interoceptive attention. *Social Cognitive and Affective Neuroscience, 8*, 15–26.
- Feldman H., & Friston K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience, 4*, 215.
- Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex, 1*, 1–47.
- Fox, K. C., Dixon, M. L., Nijeboer, S., Girn, M., Floman, J. L., Lifshitz, M., ... Christoff, K. (2016). Functional neuroanatomy of meditation: A review and meta-analysis of 78 functional neuroimaging investigations. *Neuroscience & Biobehavioral Reviews, 65*, 208–228.
- Fresco, D. M., Moore, M. T., van Dulmen, M. H., Segal, Z. V., Ma, S. H., Teasdale, J. D., & Williams, J. M. G. (2007). Initial psychometric properties of the experiences questionnaire: Validation of a self-report measure of decentering. *Behavior Therapy, 38*, 234–246.
- Friston, K. (2009). The free-energy principle: A rough guide to the brain? *Trends in Cognitive Sciences, 13*, 293–301.
- Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience, 11*, 127–138.
- Friston, K. (2018). Does predictive coding have a future? *Nature Neuroscience, 21*, 1019–1021.
- Friston, K., Stephan, K. E., Montague, R., & Dolan, R. J. (2014). Computational psychiatry: The brain as a phantastic organ. *The Lancet Psychiatry, 1*, 148–158.
- Galante, J., Friedrich, C., Dawson, A. F., Modrego-Alarcón, M., Gebbing, P., Delgado-Suárez, I., ... Jones, P. B. (2021). Mindfulness-based programmes for mental health promotion in adults in nonclinical settings: A systematic review and meta-analysis of randomised controlled trials. *PLoS Medicine, 18*, e1003481.
- Goldberg, S. B., Tucker, R. P., Greene, P. A., Davidson, R. J., Wampold, B. E., Kearney, D. J., & Simpson, T. L. (2018). Mindfulness-based interventions for psychiatric disorders: A systematic review and meta-analysis. *Clinical Psychology Review, 59*, 52–60.
- Gu, J., Strauss, C., Bond, R., & Cavanagh, K. (2015). How do mindfulness-based cognitive therapy and mindfulness-based stress reduction improve mental health and well-being? A systematic review and meta-analysis of mediation studies. *Clinical Psychology Review, 37*, 1–12.
- Harshaw, C. (2015). Interoceptive dysfunction: Toward an integrated framework for understanding somatic and affective disturbance in depression. *Psychological Bulletin*,

- 141, 311–363.
- Hayes, S. C., Bissett, R. T., Korn, Z., Zettle, R. D., Rosenfarb, I. S., Cooper, L. D., & Grundt, A. M. (1999). The impact of acceptance versus control rationales on pain tolerance. *The Psychological Record*, 49, 33–47.
- Hechler, T., Endres, D., & Thorwart, A. (2016). Why harmless sensations might hurt in individuals with chronic pain: About heightened prediction and perception of pain in the mind. *Frontiers in Psychology*, 7, 1638.
- Helmholtz, H. (1925). *Treatise on physiological optics* (Vol. 3, pp. 1–37). Retrieved from <https://echo.mpiwg-berlin.mpg.de/ECHOdocuView?url=/permanent/library/HS7FH69N/pageimg&pn=9&mode=imagepath> (July 4, 2021.)
- Hölzel, B. K., Lazar, S. W., Gard, T., Schuman-Olivier, Z., Vago, D. R., & Ott, U. (2011). How does mindfulness meditation work? Proposing mechanisms of action from a conceptual and neural perspective. *Perspectives on Psychological Science*, 6, 537–559.
- Hölzel, B. K., Ott, U., Hempel, H., Hackl, A., Wolf, K., Stark, R., & Vaitl, D. (2007). Differential engagement of anterior cingulate and adjacent medial frontal cortex in adept meditators and non-meditators. *Neuroscience Letters*, 421, 16–21.
- Huang, Y., & Rao, R. P. N. (2011). Predictive coding. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2, 580–593.
- 乾 敏郎・坂口 豊 (2020) 脳の大統一理論——自由エネルギー原理とはなにか 岩波書店.
- James, W. (1890). *The principles of psychology*. Retrieved from <http://psychclassics.yorku.ca/James/Principles/> (July 4, 2021.)
- Jamieson, G. A. (2016). A unified theory of hypnosis and meditation states: The interoceptive predictive coding approach. *Hypnosis and Meditation* (pp. 313–342). Oxford, England: Oxford University Press.
- Kabat-Zinn, J. (1990). *Full catastrophe living: Using the wisdom of your body and mind to face stress, pain, and illness*. New York, NY: Delacorte Press.
- Kabat-Zinn, J. (1994). *Wherever you go, there you are: Mindfulness meditation in everyday life*. New York, NY: Hyperion.
- Kabat-Zinn, J. (2003). Mindfulness-based interventions in context: Past, present, and future. *Clinical Psychology: Science and Practice*, 10, 144–156.
- Kirk, U., & Montague, P. R. (2015). Mindfulness meditation modulates reward prediction errors in a passive conditioning task. *Frontiers in Psychology*, 6, 90.
- Kube, T., Schwarting, R., Rozenkrantz, L., Glombiewski, J. A., & Rief, W. (2020). Distorted cognitive processes in major depression: A predictive processing perspective. *Biological Psychiatry*, 87, 388–398.
- Lebois, L. A., Papiés, E. K., Gopinath, K., Cabanban, R., Quigley, K. S., Krishnamurthy, V., ... Barsalou, L. W. (2015). A shift in perspective: Decentering through mindful attention to imagined stressful events. *Neuropsychologia*, 75, 505–524.
- Lindsay, E. K., & Creswell, J. D. (2017). Mechanisms of mindfulness training: Monitor and Acceptance Theory (MAT). *Clinical Psychology Review*, 51, 48–59.
- Lutz, A., Mattout, J., & Pagnoni, G. (2019). The epistemic and pragmatic value of non-action: A predictive coding perspective on meditation. *Current Opinion in Psychology*, 28, 166–171.
- Lutz, A., Slagter, H. A., Dunne, J. D., & Davidson, R. J. (2008). Attention regulation and monitoring in meditation. *Trends in Cognitive Sciences*, 12, 163–169.
- Manjaly, Z. M., & Iglesias, S. (2020). A computational theory of mindfulness based cognitive therapy from the “Bayesian brain” perspective. *Frontiers in Psychiatry*, 11, 404.
- Melloni, M., Sedeño, L., Couto, B., Reynoso, M., Gelormini, C., Favaloro, R., ... Ibanez, A. (2013). Preliminary evidence about the effects of meditation on interoceptive sensitivity and social cognition. *Behavioral and Brain Functions*, 9, 1–6.
- Metzinger, T. (2003). Phenomenal transparency and cognitive self-reference. *Phenomenology and the Cognitive Sciences*, 2, 353–393.
- Miller, M., Kiverstein, J., & Rietveld, E. (2020). Embodying addiction: A predictive processing account. *Brain and Cognition*, 138, 105495.
- Mrazek, M. D., Franklin, M. S., Phillips, D. T., Baird, B., & Schooler, J. W. (2013). Mindfulness training improves working memory capacity and GRE performance while reducing mind wandering. *Psychological Science*, 24, 776–781.
- Neff, K. D. (2003). The development and validation of a scale to measure self-compassion. *Self and Identity*, 2, 223–250.
- 大平英樹 (2016) 内受容感覚とマインドフルネス 貝谷久宣・熊野宏昭・越川房子 (編) マインドフルネス——基礎と実践—— (pp. 33–50) 日本評論社.
- 大平英樹 (2017) 予測符号化・内受容感覚・感情 エモーション・スタディーズ, 3, 2–12.
- 大平英樹 (2018) 内受容感覚の予測的符号化——福島論文へのコメント—— 心理学評論, 61, 322–329.
- 大平英樹 (2019) 脳と身体の予測的符号化とその不全——守谷・国里・杉浦論文へのコメント—— 心理学評論, 62, 132–141.
- Pagnoni, G. (2012). Dynamical properties of BOLD activity from the ventral posteromedial cortex associated with meditation and attentional skills. *Journal of Neuroscience*, 32, 5242–5249.
- Parkin, L., Morgan, R., Rosselli, A., Howard, M., Sheppard, A., Evans, D., ... Dunn, B. (2014). Exploring the rela-

- tionship between mindfulness and cardiac perception. *Mindfulness*, 5, 298–313.
- Parr, T., & Friston, K. J. (2017). Uncertainty, epistemics and active inference. *Journal of The Royal Society Interface*, 14, 20170376.
- Powers, A. R., Mathys, C., & Corlett, P. R. (2017). Pavlovian conditioning–induced hallucinations result from over-weighting of perceptual priors. *Science*, 357, 596–600.
- Raes, F., Dewulf, D., Van Heeringen, C., & Williams, J. M. G. (2009). Mindfulness and reduced cognitive reactivity to sad mood: Evidence from a correlational study and a non-randomized waiting list controlled study. *Behaviour Research and Therapy*, 47, 623–627.
- Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2, 79–87.
- Schmack, K., de Castro, A. G. C., Rothkirch, M., Sekutowicz, M., Rössler, H., Haynes, J. D., ... Sterzer, P. (2013). Delusions and the role of beliefs in perceptual inference. *Journal of Neuroscience*, 33, 13701–13712.
- Segal, Z. V., Gemar, M., & Williams, S. (1999). Differential cognitive response to a mood challenge following successful cognitive therapy or pharmacotherapy for unipolar depression. *Journal of Abnormal Psychology*, 108, 3–10.
- Segal, Z. V., Williams J. M. G., & Teasdale J. D. (2002). *Mindfulness-based cognitive therapy for depression: A new approach to preventing relapse*. New York, NY: Guilford Press.
- 越川房子 (監訳) (2007) マインドフルネス認知療法 うつを予防する新しいアプローチ 北大路書房。
- Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, 17, 565–573.
- Seth, A. K., & Friston, K. J. (2016). Active interoceptive inference and the emotional brain. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 371, 20160007.
- Seth, A. K., Suzuki, K., & Critchley, H. D. (2012). An interoceptive predictive coding model of conscious presence. *Frontiers in Psychology*, 2, 395.
- Shapiro, D. H. (1992). A preliminary study of long term meditators: Goals, effects, religious orientation, cognitions. *Journal of Transpersonal Psychology*, 24, 23–39.
- Shapiro, S. L., Carlson, L. E., Astin, J. A., & Freedman, B. (2006). Mechanisms of mindfulness. *Journal of Clinical Psychology*, 62, 373–386.
- Shergill, S. S., White, T. P., Joyce, D. W., Bays, P. M., Wolpert, D. M., & Frith, C. D. (2014). Functional magnetic resonance imaging of impaired sensory prediction in schizophrenia. *JAMA Psychiatry*, 71, 28–35.
- Song, Y., Yao, M., Kempreos, H., Byrne, Á., Xiao, Z., Zhang, Q., ... Chen, Z. S. (2021). Predictive coding models for pain perception. *Journal of Computational Neuroscience*, 49, 107–127.
- Stephan, K. E., Friston, K. J., & Frith, C. D. (2009). Dysfunction in schizophrenia: From abnormal synaptic plasticity to failures of self-monitoring. *Schizophrenia Bulletin*, 35, 509–527.
- Stephan, K. E., Manjaly, Z. M., Mathys, C. D., Weber, L. A. E., Paliwal, S., Gard, T., ... Petzschner, F. H. (2016). Allostatic self-efficacy: A metacognitive theory of dyshomeostasis-induced fatigue and depression. *Frontiers in Human Neuroscience*, 10, 550.
- Sterzer, P., Adams, R. A., Fletcher, P., Frith, C., Lawrie, S. M., Muckli, L., ... Corlett, P. R. (2018). The predictive coding account of psychosis. *Biological Psychiatry*, 84, 634–643.
- 高橋 徹・熊野宏昭 (2018) マインドフルネス 吉内一浩 (編) 医学のあゆみBOOKS 今日から実践！ 日常診療に役立つ行動医学・心身医学アプローチ (pp. 73–78) 医歯薬出版株式会社。
- Tang, Y. Y., Hölzel, B. K., & Posner, M. I. (2015). The neuroscience of mindfulness meditation. *Nature Reviews Neuroscience*, 16, 213–225.
- Teasdale, J. D. (1999). Metacognition, mindfulness and the modification of mood disorders. *Clinical Psychology & Psychotherapy: An International Journal of Theory & Practice*, 6, 146–155.
- Teasdale, J. D., Segal, Z. V., Williams, J. M. G., Ridgeway, V. A., Soulsby, J. M., & Lau, M. A. (2000). Prevention of relapse/recurrence in major depression by mindfulness-based cognitive therapy. *Journal of Consulting and Clinical Psychology*, 68, 615–623.
- Teper, R., Segal, Z. V., & Inzlicht, M. (2013). Inside the mindful mind: How mindfulness enhances emotion regulation through improvements in executive control. *Current Directions in Psychological Science*, 22, 449–454.
- Tran, U. S., Glück, T. M., & Nader, I. W. (2013). Investigating the Five Facet Mindfulness Questionnaire (FFMQ): Construction of a short form and evidence of a two-factor higher order structure of mindfulness. *Journal of Clinical Psychology*, 69, 951–965.
- 魚川祐司 (2015) 仏教思想のゼロポイント 「悟り」とは何か 新潮社。
- Verdonk, C., & Trousselard, M. (2021). Commentary: A computational theory of mindfulness based cognitive therapy from the “Bayesian brain” perspective. *Frontiers in Psychiatry*, 12, 575150.
- Wegner, D. M., Schneider, D. J., Carter, S. R., & White, T. L. (1987). Paradoxical effects of thought suppression. *Journal of Personality and Social Psychology*, 53, 5–13.