

特集 1 精神症 (サイコーシス) の精神病理学を神経科学的に再考する

3. 精神症 (サイコーシス) の精神病理学を計算論の枠組みで解剖する

山下 祐一*

抄録: 脳の計算理論は、精神病の病態メカニズム解明に新たな視点を提供する研究アプローチである。本稿では、もっとも有力視される脳の計算理論である「予測情報処理」理論に基づき、精神病の症状形成における予測、予測誤差、予測精度の異常に焦点を当てる。幻覚や妄想などの精神病症状は、これらの要素の異常が階層的に相互作用することで生じる可能性が示唆されている。ニューロロボティクスを活用した実験的枠組みを通じ、予測、予測誤差、予測精度の各異常が精神病の症状形成に帰結する可能性を検証した研究を紹介する。これらの結果に基づいて、精神病の症状形成における非特異性 (多元的経路) と異質性 (等価的経路) の重要性を指摘し、精神病の統合的理解に向けた新たな道として、単一の経路に依存しない多面的な理解の必要性を論じる。

日本生物学的精神医学会誌 36 (3) : 105-108, 2025

Key words : computational psychiatry, neural network model, predictive processing, predictive coding, prediction error

1. 精神病の病態を解明するための計算論的枠組み

計算機性能の飛躍的な向上や数理学の洗練により、数理・データ科学を用いた研究手法を精神医学に適用することへの期待が高まっている。このような研究は「計算論的精神医学 (computational psychiatry)」と総称され、精神医学研究において重要な位置を占める分野として発展を遂げている^{11, 17)}。計算論的精神医学が注目される理由の一つは、精神障害の病態理解における“水準間の説明のギャップ”を埋める可能性を秘めているからである。精神障害に関連する生物学的知見が多数蓄積されているにもかかわらず、それがどのようにして行動や症状レベルの変調に結びつくのかは十分に解明されていない。このギャップを架橋することで、計算論的精神医学は精神医学の抱える根本問題に対処し得る強力な研究方略として期待されている¹¹⁾。

計算論的精神医学の中でも特に注目されている理論が「予測情報処理」理論^{7, 16, 19)}である。予測情報処理理論とは、脳が内部モデルに基づいて外界を予

測し、実際の感覚入力とのズレ (予測誤差) を最小化することで学習や知覚、行動を制御するという考え方である。この理論によれば、人の脳は外界との相互作用を通じて内部モデルを構築する「予測器」であり、そのモデルを用いて迅速かつ正確な認知と行動を生成する。予測が感覚フィードバックと異なる場合、そのズレである予測誤差に基づき内部モデルは更新される。この予測-誤差-更新のプロセスが脳の一般的な計算原理として捉えられる。階層的に構成される予測情報処理プロセスは、人の高度な認知、特に複雑な社会環境における適応行動に大きな利点をもたらすと考えられる一方、その障害は精神障害の症状を引き起こす可能性がある。実際、近年の計算論的精神医学では、統合失調症をはじめとする精神病の症状 (幻覚、妄想、自己意識の変容など) は、予測情報処理の異常と関連づけて理解することが試みられている^{1, 3, 15)}。本稿では、計算論的精神医学の視点から精神病の病態を予測情報処理プロセスにおける「予測」「予測誤差」「精度」という3つの要素に分けて解説し、計算理論の実験的検証としてニューロロボティクス実験を用いた研究成果

Computational frameworks for dissecting the psychopathology of psychosis

* 国立研究開発法人 国立精神・神経医療研究センター 神経研究所疾病研究部第七部 (〒187-8502 東京都小平市小川東町 4-1-1) Yuichi Yamashita : Department of Information Medicine, National Center of Neurology and Psychiatry, National Institute of Neuroscience, 4-1-1 Ogawa-Higashi, Kodaira, Tokyo 187-8502, Japan

【山下 祐一 E-mail : yamay@ncnp.go.jp】

を紹介する。

2. 精神病における予測処理の異常

a. 予測の障害

Frith らの「フォワードモデル障害」仮説では、行動の意図に基づく感覚予測の異常が、実際の感覚フィードバックとのズレを生じることで行為の意思作用感 (sense of agency: SoA) の障害や被影響体験症状を引き起こすとされる⁸⁾。同様に、統合失調症では、感覚予測と関連するとされる事象関連電位 N1 抑制の遅延が確認され、予測信号の伝達遅延が示唆されている²⁰⁾。

b. 予測誤差の異常

予測誤差信号処理の変調が、幻覚や妄想といった精神病症状に帰結する可能性が提案されている。たとえば、神経生理学的観点では、特に前頭前野の予測誤差信号の異常が陽性症状と関連していることが報告されている⁵⁾。また、ミスマッチ陰性電位 (mismatch negativity: MMN) が減衰することが観察されており、予測誤差関連神経応答の異常が示唆される²⁾。

c. 精度の異常

精度とは、感覚入力や予測の信頼性を意味し、信頼すべき情報とそうでない情報を区別するという意味で、予測情報処理プロセスにおいて、予測誤差を重み付けする重要なパラメータとして機能する。予測精度が低い場合 (不確実性が高い) は予測誤差を無視し、高い場合 (不確実性が低い) は反応する必要がある。このメカニズムは、脳の適切な情報処理と環境への行動適応を可能にすると理解できる。精度推定の異常が精神病症状に関連すると考えられているが、メカニズムについては議論が分かれる。たとえば、予測精度が過大評価される場合、過剰な信念の精度が幻覚の原因となりうる (ハイパー・プライア仮説) とされる一方で、予測の精度が過小評価される場合、相対的に過剰に評価された感覚精度が誤った異常な顕著性や信念を引き起こすという仮説 (ハイポ・プライア仮説) も存在する^{4, 6, 18)}。

3. ニューロロボティクスによる実証研究

計算論的精神医学の理論を実験的に検証する強力な枠組みとして、ニューロロボティクスがある^{10, 22)}。この枠組みでは、脳の予測情報処理がどのように症

状に寄与するかを明らかにするために、人工ニューラルネットワークモデルを用いて、環境からの感覚入力 (感覚フィードバック) をもとに次の状態の予測 (トップダウン予測) とその誤差 (予測誤差) を計算するという予測情報処理プロセスを実装する。さらに、この人工ニューラルネットワークを脳とみたとて、ロボットの物理的な身体を通じて環境との動的な相互作用を再現し、抽象的な計算理論を具体的な現象に結びつけることが可能となる。この枠組みは、計算理論、神経動態、行動や症状の間の橋渡しを行い、予測情報処理の異常がどのように症状に結びつくかを統合的に検討するための手段を提供する。

a. 予測シグナルの遅延

SoA を定量的に評価する方法として、エージェンシー判断課題 (慶應メソッド) が用いられる^{12, 13)}。この課題では、被験者のボタン押しに対する視覚フィードバックのタイミングを操作する。健常者では、短い遅延の場合には行為と感覚の間に自己行為の作用感が生じる (SoA+) 一方で、遅延が長い場合には、感覚フィードバックは外的要因に影響を受けているように感じられる (SoA-)。Maeda ら¹²⁾ は、この課題を用いて統合失調症では、陽性症状が優勢の患者では、SoA が過剰となる一方、陰性症状が優勢の患者では、SoA が減弱することを明らかにし、予測信号の遅延が、自己行為の SoA 変調と関連している可能性を仮説している。これに対して、Okimura ら¹⁴⁾ は、このエージェンシー判断課題を、ニューロロボティクス実験として実装し、予測情報処理プロセスにおける予測信号の遅延が SoA に与える影響を検証した。実験の結果、予測シグナルの遅延があると、統合失調症で観察されていた SoA の二方向性の変調が再現され、統合失調症の SoA 変化に予測シグナルの遅延が関与している可能性が示唆された。

b. 階層間の機能的断裂と異常な予測誤差シグナルの生成

予測誤差の変調が精神病症状に帰結する可能性を検証するため、予測情報処理プロセスを実装した階層的ニューラルネットワークモデルにおいて、階層間の機能的切断を導入し、その影響を評価する実験が行われた^{21, 22)}。軽微な機能的断裂では、自発的な予測誤差が生じ、その予測誤差が妄想気分や被影響体験といった病的体験を引き起こす可能性が示唆された。一方、重度の機能的断裂では、カタレプシー

様の行動や常同行為といった、統合失調症の重症例でみられる行動異常が再現された。これらの結果から、精神病の多様な症状が、予測誤差最小化プロセスの障害とそれに対する代償的な応答によって説明される可能性が示された^{21, 22)}。これらの知見は、予測誤差の変調が精神病症状の形成における中心的なメカニズムであることを示唆し、精神病の病態生理学を解明するうえで重要な視点を提供するものである。

c. 精度の階層的モデル

上述したように、精度の異常が精神病の症状形成を説明する可能性については議論が分かれている¹⁸⁾。これに対し筆者らは、精度の階層構造を詳細に考慮したモデルを導入し、階層間の精度推定の影響の相対的な不均衡が与える影響を検証した^{9, 10)}。実験の結果、精度の階層構造における不均衡が、精神病の症状形成において重要な役割を果たす可能性が示された。具体的には、中間レベルの精度の推定が減弱することで、下位の感覚運動レベルでの予測誤差への過剰な反応と同時に、中間レベルでの未解決の予測誤差が波及する結果、上位レベルの活動の不安定化を引き起こし、知覚の顕著性の増加や行為の誤帰属といった、統合失調症における行動および神経的变化と類似の現象が再現された^{9, 10)}。

4. 精神病の多様な発現メカニズム

予測情報処理の異常として精神病の症状形成メカニズムの可能性を列挙してきたが、読者はどれが正しい経路なのか疑問を抱くかもしれない。しかし、唯一の正解を求める必要は必ずしもない。むしろ、精神病の病態を理解するうえで重要なのは、生物学的な非特異性と異質性という構造である。非特異性(多元的経路: multifinality)は、同じ生物学的因子が異なる表現型(症状)を引き起こす現象を指し、異質性(等価的経路: equifinality)は、異なる生物学的因子が同じ表現型をもたらす現象を指す^{10, 17)}。これらの構造は、精神障害の症状形成を理解する際に必ず直面する困難であり、常にこの点を考慮している必要がある。予測情報処理は、予測、予測誤差、および予測精度といった要素が相互に関連し、精神病の発症に寄与する複数の経路の可能性を提案する。このような非特異性と異質性を受け入れつつ、それぞれの要素が相互にどのように作用するかを探ることで、精神病の統合的な理解に貢献することが期待される。

本論文に記載した筆者らの研究についてすべて倫理的配慮を行っている。開示すべき利益相反は存在しない。

文 献

- 1) Adams RA, Stephan KE, Brown HR, et al (2013) The computational anatomy of psychosis. *Front Psychiatry*, 4 : 47.
- 2) Baldeweg T, Klugman A, Gruzelier JH, et al (2002) Impairment in frontal but not temporal components of mismatch negativity in schizophrenia. *Int J Psychophysiol*, 43 (2) : 111-122.
- 3) Brown H, Adams RA, Parees I, et al (2013) Active inference, sensory attenuation and illusions. *Cogn Process*, 14 (4) : 411-427.
- 4) Corlett PR, Horga G, Fletcher PC, et al (2019) Hallucinations and strong priors. *Trends Cogn Sci*, 23 (2) : 114-127.
- 5) Corlett PR, Murray GK, Honey GD, et al (2007) Disrupted prediction-error signal in psychosis : evidence for an associative account of delusions. *Brain*, 130 (9) : 2387-2400.
- 6) Friston KJ (2005) Hallucinations and perceptual inference. *Behav Brain Sci*, 28 (6) : 764-766.
- 7) Friston K, Mattout J and Kilner J (2011) Action understanding and active inference. *Biol Cybern*, 104 (1-2) : 137-160.
- 8) Frith CD, Blakemore SJ and Wolpert DM (2000) Explaining the symptoms of schizophrenia : abnormalities in the awareness of action. *Brain Res Rev*, 31 (2-3) : 357-363.
- 9) Idei H, Ohata W, Yamashita Y, et al (2022) Emergence of sensory attenuation based upon the free-energy principle. *Sci Rep*, 12 (1) : 14542.
- 10) Idei H and Yamashita Y (2024) Elucidating multifinial and equifinal pathways to developmental disorders by constructing real-world neurorobotic models. *Neural Netw*, 169 : 57-74.
- 11) 国里愛彦, 片平健太郎, 沖村幸, 他 (2019) 計算論的精神医学—情報処理過程から読み解く精神障害. 勁草書房, 東京.
- 12) Maeda T, Kato M, Muramatsu T, et al (2012) Aberrant sense of agency in patients with schizophrenia : forward and backward over-attribution of temporal causality during intentional action. *Psychiatry Res*, 198 (1) : 1-6.
- 13) Maeda T, Takahata K, Muramatsu T, et al (2013) Re-

- duced sense of agency in chronic schizophrenia with predominant negative symptoms. *Psychiatry Res*, 209 (3) : 386-392.
- 14) Okimura T, Maeda T, Mimura M, et al (2023) Aberrant sense of agency induced by delayed prediction signals in schizophrenia : a computational modeling study. *Schizophrenia (Heidelb)*, 9 (1) : 72.
- 15) Powers AR, Mathys C and Corlett PR (2017) Pavlovian conditioning-induced hallucinations result from overweighting of perceptual priors. *Science*, 357 : 596-600.
- 16) Rao RP and Ballard DH (1999) Predictive coding in the visual cortex : a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci*, 2 (1) : 79-87.
- 17) Redish AD and Gorden JA Eds (2016) *Computational psychiatry : new perspectives on mental illness*. MIT Press, Cambridge, MA.
- 18) Sterzer P, Adams RA, Fletcher P, et al (2018) The predictive coding account of psychosis. *Biol Psychiatry*, 84 (9) : 634-643.
- 19) Tani J (2003) Learning to generate articulated behavior through the bottom-up and the top-down interaction processes. *Neural Netw*, 16 (1) : 11-23.
- 20) Whitford TJ, Mathalon DH, Shenton ME, et al (2011) Electrophysiological and diffusion tensor imaging evidence of delayed corollary discharges in patients with schizophrenia. *Psychol Med*, 41 (5) : 959-969.
- 21) 山下祐一, 松岡洋夫, 谷淳 (2013) 計算論的精神医学の可能性 : 適応行動の代償としての統合失調症. *精神医学*, 55 (9) : 885-895.
- 22) Yamashita Y and Tani J (2012) Spontaneous prediction error generation in schizophrenia. *PLoS One*, 7 (5) : e37843.

■ ABSTRACT

Computational frameworks for dissecting the psychopathology of psychosis

Yuichi Yamashita

Department of Information Medicine, National Center of Neurology and Psychiatry, National Institute of Neuroscience

Computational psychiatry offers a novel perspective on understanding the pathophysiology of psychosis. This review focuses on the predictive processing framework, one of the most influential computational theories of the brain, to explore the abnormalities in prediction, prediction error, and prediction precision that underlie psychotic symptoms. Hallucinations and delusions, key features of psychosis, are hypothesized to result from hierarchical interactions among these factors. Using a neuro-robotics experimental framework, studies have empirically examined how disruptions in prediction, prediction error, and precision can contribute to the formation of psychotic symptoms. These experiments replicated behaviors and experiences analogous to schizophrenia, providing insights into the mechanisms driving the disorder. Furthermore, this review highlights the importance of biological non-specificity (multifinality) and heterogeneity (equifinality) in psychosis pathophysiology. Non-specificity refers to the phenomenon where the same biological factor leads to distinct symptoms, while heterogeneity describes how different biological factors produce similar symptoms. These concepts underscore the necessity of moving beyond a single causal pathway to consider multiple interacting mechanisms. Understanding these dynamics may pave the way for integrative approaches to the diagnosis and treatment of psychosis, contributing to a more unified and comprehensive understanding of its pathophysiological basis.

(Japanese Journal of Biological Psychiatry 36 (3) : 105-108, 2025)
