	領域代表者	理化学研究所・脳神経科学研究センター・ユニットリーダー 磯村 拓哉（いそむら たくや） 研究者番号:50802972
	研究領域情報	領域番号：23A402 キーワード：神経科学、人工知能、生成モデル、自由エネルギー原理、リバースエンジニアリング 研究期間：2023年度～2027年度

なぜこの研究を行おうと思ったのか（研究の背景・目的）

●研究の全体像

脳の計算原理を解明し、生物のように考える人工知能を構築することは、自然科学と情報工学における最大のフロンティアです。人工知能はこれまで神経科学からヒントを得ることで発展し、特に特徴抽出と強化学習において大きな成功を収めてきました。しかしそれでも、人工知能と人間の知能の間には未だ大きなギャップが存在します。

生物は外界のダイナミクスを表す「生成モデル」を脳内に構築することで外界の状態を能動的に推論・予測し、将来のリスクを最小化するように行動を最適化しています。このような生物の予測と行動の基盤となる脳の情報理論として、ベイズ脳仮説や自由エネルギー原理等の理論が提案されてきました。最近では自由エネルギー原理が脳の統一理論の候補に挙げられていますが、理論の抽象度が高く、生命現象との対応付けはこれまで困難だと考えられていました。

しかし、最近の脳科学における実験技術の発展によって細胞種の特定や複数の層、領野を対象とした高精度大規模データの取得が可能となった事に加え、理論面でも神経回路ダイナミクスと神経活動モデルから導かれる生成モデルを一つ一つに対応付ける生成モデルのリバースエンジニアリング手法が開発された事で、脳と心の理解に極めて重要な「生物が持つ生成モデルを実験データから同定すること」が現実的になってきました。

そこで本研究領域では、脳の神経活動を高精度・大規模に取得し、データから脳が持つ生成モデルをリバースエンジニアリングすることで様々な脳の情報理論を検証し、予測と行動の統一理論の確立を目指します（図1）。そのためにサカナ・ネズミ・サル・ヒトを含む様々な生物種を対象として外界の予測や行動に関連する神経活動を計測し、実験データから構築した生成モデルが動物の脳活動や行動、学習に伴う脳活動や行動の変動を予測できるかテストすることで理論を検証し、改良・拡張します。理論と実験の双方性の連携を中核として研究を進めることで、知覚的な予測と行動の計画・生成を統一的に説明可能な脳の統一理論を確立し、ヒトのように考える人工知能や精神疾患の早期診断手法の開発への道筋を開拓していきます。

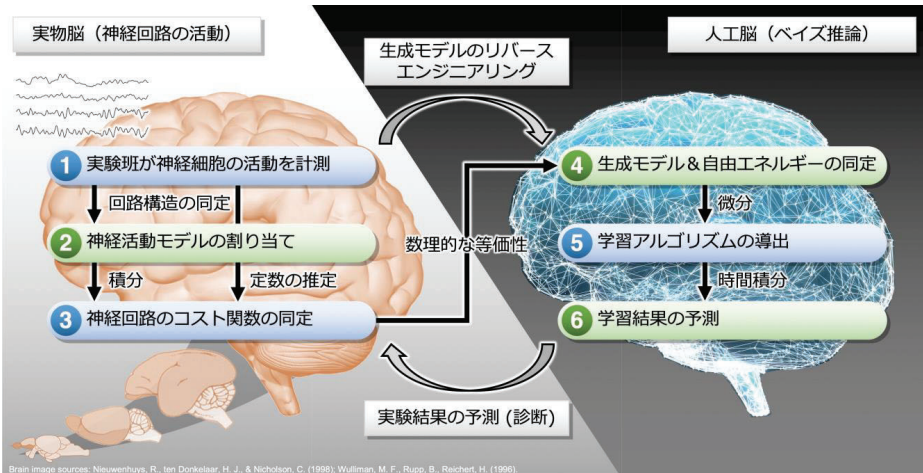


図1 生成モデルのリバースエンジニアリングを用いた統一理論の検証

この研究によって何をどこまで明らかにしようとしているのか

●実験と理論の融合による統一理論の開拓・検証

A00理論班とB00実験班という2つの計画研究グループを設定し、グループ内での研究フレームワーク共有に加えてA00とB00の連携を通じて脳の統一理論の開拓と検証という領域の目標に貢献する体制をとりま（図2左）。階層的にサカナ（岡本）・マウス（銅谷）・マーモセット（蝦名・小松）・ヒト（健常者・精神疾患患者）（高橋）の生体計測を配し、最先端の計測を行える布陣を整えております。神経細胞の活動データにリバースエンジニアリングを適用することで、生成モデルを推定することができ、さらに脳型AIを構築して比較・検証することが可能になります。様々な動物種と計測手法を用いて、リバースエンジニアリングがどこまで普遍的に動物の脳活動や行動を予測可能かを明らかにしていきます。

領域の目標達成にとりわけ重要な課題として、3つの共通課題を設定し、共通の実験・解析パラダイムを用いて普遍的な原理の解明に取り組める連携体制を整えました（図2右）。課題①では、ヒトとマーモセットの脳活動データと深層生成モデル（鈴木）の比較により、高次な視覚・聴覚入力の予測を担う深層予測符号化の検証を行います。課題②では、サカナやマーモセットが迷路などの課題を解く時の脳活動を調べることで、動物が報酬を得るために数手先まで計画し行動できる意思決定の仕組みを解明していきます。課題③では、マウスとマーモセットのレバー押し課題を用いて、動物が現実の連続状態空間内で適切に行動計画し身体を制御する神経メカニズムを解明していきます。これら課題①②③は全て自由エネルギー原理やベイズ脳仮説の検証になっており、このようにして予測と行動の統一理論の確立という最終目標の達成を目指します。

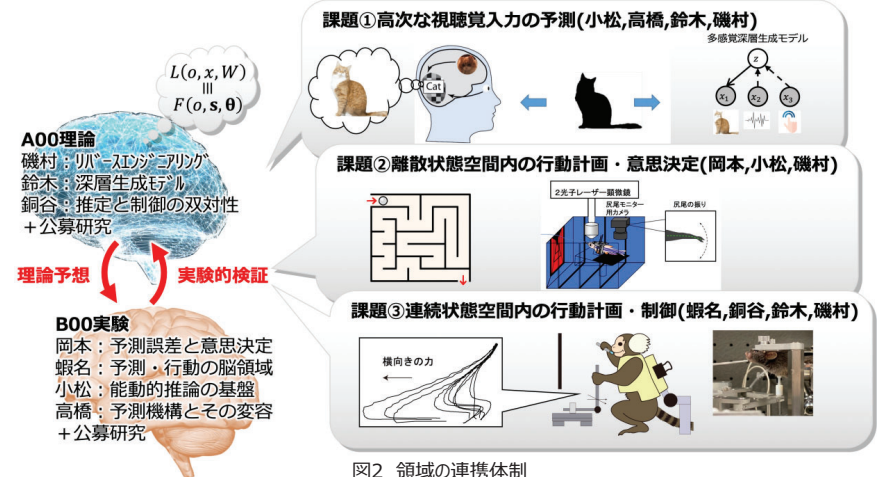


図2 領域の連携体制

●期待される波及効果

予測と行動の統一理論を確立することで、神経回路の計算原理、つまり脳や心の仕組みへの理解が深まり、神経科学全体が進歩するとともに、脳を予測することが可能になります。本領域の成果は、精神疾患の機序解明や早期診断・治療方法の開発、新たなニューロテックの創出、ヒトのように考える脳型人工知能の開発への応用が期待できます（図3）。特にAI応用に関しては、脳の効率的な学習の仕組みを解明することで、既存の手法よりも効率の良い学習アルゴリズムの発見が期待できます。



図3 期待される波及効果

ホームページ等	ホームページ	https://unifiedtheory.jp
	YouTube	https://www.youtube.com/@unifiedtheoryjp
	X(旧Twitter)	https://twitter.com/unifiedtheoryjp
	お問い合わせ	https://unifiedtheory.jp/contact/